Social Momentum: Design and Evaluation of a Framework for Socially Competent Robot Navigation

CHRISTOFOROS MAVROGIANNIS, Sibley School of Mechanical & Aerospace Engineering, Cornell University, USA
PATRÍCIA ALVES-OLIVEIRA, ISCTE-University Institute of Lisbon, Portugal
WIL THOMASON and ROSS A. KNEPPER, Department of Computer Science, Cornell University, USA

Mobile robots struggle to integrate seamlessly in crowded environments such as pedestrian scenes, often disrupting human activity. One obstacle preventing their smooth integration is our limited understanding of how humans may perceive and react to robot motion. Motivated by recent studies highlighting the benefits of intent-expressive motion for robots operating close to humans, we describe Social Momentum (SM), a planning framework for legible robot motion generation in multiagent domains. We investigate the properties of motion generated by SM via two large-scale user studies: an online, video-based study (N = 180) focusing on the legibility of motion produced by SM and a lab study (N = 105) focusing on the perceptions of users navigating next to a robot running SM in a crowded space. Through statistical and thematic analyses of collected data, we present evidence suggesting that (a) motion generated by SM enables quick inference of the robot's navigation strategy; (b) humans navigating close to a robot running SM follow comfortable, low-acceleration paths; and (c) robot motion generated by SM is positively perceived and indistinguishable from a teleoperated baseline. Through the discussion of experimental insights and lessons learned, this article aspires to inform future algorithmic and experimental design for social robot navigation.

CCS Concepts: • **Computer systems organization** → **Robotics**; **Robotic control**; • **Computing methodologies** → **Robotic planning**; • **Human-centered computing** → *User studies;*

Additional Key Words and Phrases: Social navigation, multiagent systems, social robotics, benchmarking

ACM Reference format:
Christoforos Mavrogiannis, Patrícia Alves-Oliveira, Wil Thomason, and Ross A. Knepper. 2022. Social Momentum: Design and Evaluation of a Framework for Socially Competent Robot Navigation. *ACM Trans. Hum.-Robot Interact.* 11, 2, Article 14 (February 2022), 37 pages.
https://doi.org/10.1145/3495244

C. Mavrogiannis and R. Knepper acknowledge support from the National Science Foundation, under Grants IIS-1526035 and IIS-1563705. P. Alves-Oliveira acknowledges support from the Fundação para a Ciência e a Tecnologia, under grant SFRH/BD/110223/2015. W. Thomason acknowledges support from the Office of Naval Research, under the NDSEG fellowship program.

Authors' addresses: C. Mavrogiannis, Room 259, Bill & Melinda Gates Center for Computer Science & Engineering, University of Washington, 3800 E Stevens Way NE, Seattle, WA, 98195-2355; email: cmavro@cs.washington.edu. P. Alves-Oliveira, Room 262, Bill & Melinda Gates Center for Computer Science & Engineering, University of Washington, 3800 E Stevens Way NE, Seattle, WA, 98195-2355; email: patri@cs.washington.edu. W. Thomason, Room 350, Bill & Melinda Gates Hall, Cornell University, 107 Hoy Rd, Ithaca, NY 14853; email: wbthomason@cs.cornell.edul. R. Knepper, Room 571, Upson Hall, Cornell University, 107 Hoy Rd, Ithaca, NY 14853; email: ross.knepper@gmail.com.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.
2573-9522/2022/02-ART14 $15.00
https://doi.org/10.1145/3495244

1 INTRODUCTION

Imagine a crowded city street. It is easier to picture humans smoothly flowing with the crowd as they navigate toward their destination. However, this skill of social navigation, which is so natural and intuitive for humans, is challenging forrobots to master. Indeed, the problem of enabling robots to traverse crowded pedestrian environments in a socially competent fashion has received considerable attention over the past three decades [44, 47, 61, 63, 69]. Deploying a mobile robot in a human environment requires paying significant attention to topics related to human safety and comfort. To this end, researchers drawn insights and frameworks from diverse fields ranging from motion planning and prediction to design research and social sciences.

Earlier work looked at mechanisms of reproducing collision-avoidance maneuvers. Often, humans are approximated as dynamic obstacles [24, 57], and the robot reacts to their motion or projected future path to avoid collisions. This is known to result in problematic interactions [22]; humans are intelligent agents with sophisticated mechanisms of inference and decision making that are sensitive to the motion of others. In fact, human navigation in crowds is known to be a highly cooperative activity—humans tend to expect others to adjust their motion and share the responsibility for collision avoidance [80]. Failing to account for the existence of such mechanisms in humans is known to result in suboptimal and often oscillatory robot behavior that appears to be uncomfortable [22, 72]. To account for human expectations, a body of work has focused on mathematically formalizing and reproducing social norms such as passing from the right [37], respecting humans' personal space [38], or civil inattention [41]. In practice, the way humans navigate can be fluid and dependent on the context (e.g., culture, type of environment, density of crowd) or variable across individuals. Sometimes, physical constraints arise or social norms are violated but humans quickly and safely adapt. Humans tend to follow subconscious insights and instincts but also social rules [28, 29, 80] that are often hard to understand and quantitatively model in robots. Thus, enumerating and reproducing a fixed set of social rules may be impractical for handling the complexity of real-world situations.

Motivated by these observations, recent work has focused on incorporating models of multiagent interaction into the robot's decision making. Some approaches [25, 65, 73] employ crowd motion simulations [32] as prediction mechanisms for human motion. The behavior generated by such models is dependent on a set of parameters defining considerations such as preferred speed, proximity to others, and so on. Tuning these parameters for navigation among humans is not trivial. This has motivated a large body of work to learn models of pedestrian dynamics from data generated using simulation engines [32, 76], inverse reinforcement learning [42], Gaussian process regression [72], sequence-to-sequence models [50], or deep reinforcement learning [21]. While some of these approaches are capable of producing safe robot motion in human spaces, they tend to suffer from the high inherent complexity [15] (sample or computational) induced by reasoning about the motion of multiple agents over a future time window at planning time. Further, their emphasis is chiefly on collision avoidance, ignoring important aspects that govern interaction, such as human perception and inference mechanisms.

Recently, a few studies have highlighted that legible robot motion tends to reduce the planning effort [12, 43] and increase comfort and productivity [17] for nearby humans. Despite the recognized value of legible robot motion for human–robot interaction applications, existing work has focused on scenarios in structured domains involving a single human agent. However, navigation in real-world pedestrian spaces is inherently multiagent, motivating the development of new frameworks for modeling and generating legible robot motion in the presence of multiple dynamic agents.

In this article, we take an interaction-first approach to social robot navigation. We present the design principles, algorithmic formulation and extensive evaluation of a planning framework for

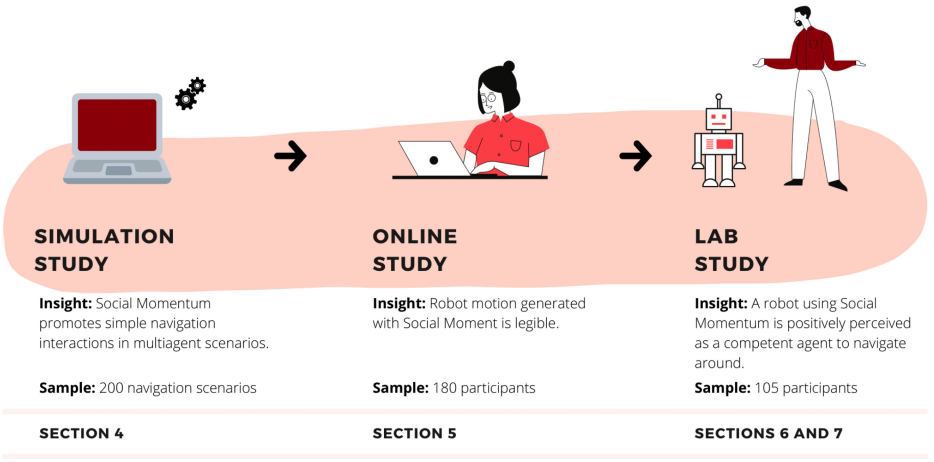


Fig. 1. Roadmap of evaluation followed in this article.

multiagent humans spaces. First, we introduce a representation inspired by the physical quantity of angular momentum that factors the unfolding multiagent dynamics in a crowded scene into a set of likely pairwise collision avoidance maneuvers between the robot and other agents. Based on this representation, we design an objective function that quantifies the extent of the agreement between a robot action and the perceived preferences of other agents over passing sides. By maximizing this objective, the robot may signal its intention of complying with these perceived preferences (Figure 2). Based on this insight, we design a decision-making policy—called **Social Momentum (SM)**—that chooses actions balancing between this objective and progress toward the robot's destination.

Through an online video-based user study ($N = 180$), we show that this type of decision making can be perceived as *legible* in the sense that it enables an observer to quickly and confidently infer the robot's intention over a passing side. We then deploy this mechanism on a real robot and evaluate its performance via a lab-study with groups of human subjects ($N = 105$). Key findings of this study include that (a) humans are more likely to follow low-acceleration paths next to a robot running our framework compared to a set of baselines and (b) our framework is perceived just as well as a Wizard-of-Oz baseline in which a human operator teleoperates the robot. Our overall findings suggest that even a simple interaction-aware policy can be sufficient for navigation in close proximity with humans in selected settings. While our policy may struggle in more complex domains, it does not impose strong requirements on datasets or on-board computation like existing state-of-the-art approaches. A visual roadmap of the studies described is presented in Figure 1.

1.1 Contributions

In summary, with this article, we make the following contributions that expand upon our prior work [51, 53]:

- We conduct an extensive review of related work on algorithmic approaches to social robot navigation, clearly placing our framework with respect to the rich literature in the field (Section 2).
- We present a unified discussion of our work on the design and evaluation of our Social Momentum framework [51, 53]. We clearly lay out the principles from psychology and sociology literature that inspired our approach (Section 3) and build upon them to frame our algorithm design (Section 4).

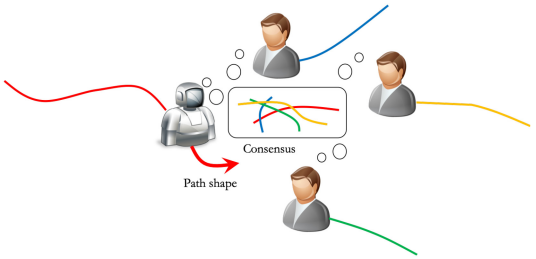


Fig. 2. The key idea of our framework: achieving consensus over a joint strategy of collision avoidance using the robot's path shape as a communicative modality.

- We conduct a new, extended statistical analysis of the Likert-scale questionnaire dataset collected in our lab study [51] to better understand the perceptions of humans about a robot navigates close to them (Section 7.1). Through a factor analysis methodology, we show that human perceptions of robot behavior are driven by impressions along five key categories, namely *social competence*, *intelligence*, *human comfort*, *predictability*, and *discretion*.
- We conduct a qualitative thematic analysis of the dataset extracted from responses of participants to open-form questions collected in our lab study [51] (Section 7.2). We elaborate on participants' attitudes toward the robot, which revolve along key themes such as robot *navigation intention*, *proximity*, *robot performance*, and *human emotions*.
- We conclude with a comprehensive, unified discussion of all findings. We include experimental insights, lessons learned and discuss their implications for the future of social robot navigation research (Section 8).

2 RELATED WORK

Robot navigation in crowded human environments requires a series of algorithmic components including motion prediction, motion planning, and control, among others. As such, research in the field interfaces with a number of neighboring communities including crowd simulation, path tracking, trajectory prediction, multirobot systems, autonomous driving, and more. Further, the inherently close interaction between the robot and human subjects motivates the incorporation of features related to social awareness and comfort. Thus, social robot navigation also interfaces significantly with the field of design and often draws insights from psychology, sociology, and human–robot interaction. Taking into account these interfaces, in this literature review, we specifically focus on frameworks that were developed to be deployed on real robots in crowded environments. We review relevant works, highlighting their algorithmic foundations and evaluation methodologies.

2.1 Initial Efforts: Humans as Obstacles

First efforts in the field of social robot navigation were motivated by specific real-world applications. For instance, Rhino [10] and Minerva [71] were some of the first known robotic systems to be deployed as tour guides in museums in Bonn, Germany, and Washington DC, USA, in 1997 and 1998, respectively. These robots successfully served thousands of tour guide requests and navigated alongside museum visitors over a span of weeks. Their performance set the stage for real-world challenges, such as the AAAI Mobile Robot Challenge, organized by the American Association for Artificial Intelligence, which attracted wide interest and participation [54, 56]. Although these deployments featured remarkable system development and integration efforts, their underlying navigation frameworks [24, 57] treated humans as non-reactive obstacles without explicitly mod-

eling interaction phenomena or humans' decision-making strategies. This assumption has generally sufficed to provide a practical solution to the crowd navigation problem, and it has inspired a significant amount of research over the years [7, 23, 60]. However, it has been empirically observed that this assumption can create practical problems including the "reciprocal dance" [22]: The failure of the robot to infer the human's intended motion results in the robot choosing a velocity that emerges as surprising to the human, who in turn reacts unpredictably to the robot, contributing to a short oscillatory interaction. Such issues motivated the introduction of models for reasoning about uncertainty in crowd navigation.

2.2 Reasoning about Uncertainty

Motivated by the practical issues caused by the lack of predictive models, a second wave of efforts on the crowd navigation problem focused on methodologies for reasoning about uncertainty. For instance, Du Toit and Burdick [18] present a receding-horizon control framework that incorporates predictive uncertainty in the robot's decision making. Thompson et al. [70] introduce a probabilistic model of human motion based on individual human intent inference, designed to assist in motion planning problems. Joseph et al. [34] propose a Bayesian framework reasoning about individual human motion patterns to inform a motion prediction pipeline. Similarly, Bennewitz et al. [8] extract patterns of human motion in a crowded environment and derive a hidden Markov model to perform online human motion prediction. Unhelkar et al. [75] introduce a motion prediction framework that makes use of biomechanical features to anticipate human turning actions.

Despite the introduction of principled models for reasoning about uncertainty, these works treat human agents as individual non-interactive entities. The lack of explicit coupling over the possible motion of human agents often results in an *uncertainty explosion* as observed by Du Toit and Burdick [18]. In practice, this may cause the robot to overestimate the uncertainty over the unfolding crowd motion and thus falsely infer that no collision-free paths exist. This phenomenon may effectively yield the *freezing robot problem* as described by Trautman et al. [72], who pointed out that a practical way to resolve it is to enable the robot to explicitly expect human cooperation.

2.3 Interaction in Social Robot Navigation

Motivated by phenomena such as the reciprocal dance [22] and the freezing robot problem [72], a relatively recent wave of work focuses on explicitly modeling *interaction*. Unlike conventional instances of the robot navigation problem that take place in isolated, structured environments, social robot navigation takes place in domains that are inherently dynamic, multiagent, and uncertain. Modeling the interaction dynamics among a set of multiple navigating agents is a computationally hard problem: Related instances of the problem such as inference over dynamic Bayesian networks [15] have been shown to be NP-hard. However, humans are remarkably effective in avoiding conflicts with others in crowded domains. Human effectiveness is largely attributed to sophisticated mechanisms of cooperation [28, 35, 80], often realized implicitly via channels of implicit, nonverbal communication [4, 16, 79].

2.3.1 The Human Paradigm. Several sociology studies on pedestrian navigation have pointed out the cooperative character of collision avoidance in crowds. From a broad perspective, Karp et al. [35], in their definition of the *Mini-Max Hypothesis of Urban Life*, specify that "urbanites seek to minimize involvement and to maximize social order." Goffman [28] tailors this idea to crowd navigation in his concept of "civil inattention": upon acknowledging the presence of a fellow pedestrian via eye gaze, a pedestrian often looks away "so as to express that he does not constitute a target of special curiosity or design." This process hints toward the existence of some form of

trust across pedestrians: trust that the responsibility for collision avoidance will be shared by both agents. Building on this concept, Wolfinger [80] introduces the concept of the *pedestrian bargain* to describe the mechanism of trust that sustains social order in crowd navigation. This mechanism is based on two simple rules: (1) "people must behave like competent pedestrians" and (2) "people must trust co-present others to behave like competent pedestrians." Trust in the rules of the bargain constitutes the basis of smooth co-navigation in human environments, as it enables pedestrians to plan with the expectation that others will also behave competently and thus *cooperate* to resolve potential conflicts. On top of these ideas, the theory of *Proxemics* by Hall [31] specifies a series of considerations related to social spaces and comfort in public places.

2.3.2 Reproducing Social Conventions. Some works have focused on generating motion that reproduces selected social conventions and norms observed in human navigation. For instance, Kirby et al. [39] introduce a constrained optimization-based algorithm that incorporates considerations of passing side and *personal space* [31] into the robot's decision making. Sisbot et al. [66] present a cost-based planner that employs a set of social costs to generate motion that is *visible* and *safe* around humans. Knepper and Rus [41] distill the concept of *civil inattention* [28] into a multirobot path planner.

2.3.3 Modeling Interaction. Some other works focus on capturing the dynamics of interaction to inform prediction and planning. For example, Warren [78] introduces a dynamical systems approach describing organization across tasks including human navigation. Luber et al. [46] learn a set of dynamic navigation prototypes and use them to design dynamic costmaps that capture objective and subjective human navigation objectives. Moussaïd et al. [58] employ behavioral heuristics to model pedestrian behavior in public spaces and show how humans adjust their speed and directionality guided by distance-based visual features. G. Ferrer [25] integrate human motion predictions generated with a Social Force model [32] variant into a sampling-based planning framework. The works of Ziebart et al. [81], Vasquez et al. [77], Henry et al. [33], Kim and Pineau [37], and Kretzschmar et al. [42] employ inverse reinforcement learning as a technique to recover features of human navigation objectives and use them to design motion planners for humanlike robot navigation in a variety of scenarios. Trautman et al. [72] introduce a Gaussian Process-based model for motion prediction that accounts for interdependencies arising as a result of human cooperative collision avoidance and show how it can be used to enable safe robot navigation in dense crowds. Chen et al. [14] and Everett et al. [21] present deep reinforcement learning models that implicitly capture elements of cooperative collision avoidance from observations of multiagent interactions. Che et al. [13] integrate modalities of explicit and implicit communication into a motion planner to generate plans that are easier for humans to read. Finally, focusing on autonomous driving domains, Sadigh et al. [64] and Roh et al. [62] show that a motion planner with a model of multiagent dynamics may leverage communicative signals encoded in robot motion to influence the behavior of other vehicles. We note that there has been relevant work in this domain from the field of social motion prediction [63], but the focus there is not on robot navigation but rather on object tracking.

2.3.4 Our Perspective: Leveraging the Mathematical Structure of Interaction. Our past work [49, 50, 52] introduces a unique perspective in social robot navigation, proposing frameworks that *explicitly* model cooperative collision avoidance in multiagent navigation through the use of mathematical representations inspired by work in low-dimensional topology. Topological representations elegantly capture salient features of multiagent interactions into objects of dual algebraic and geometric nature, enabling the use of both symbolic reasoning and learning techniques. For instance, we employed topological braids [3, 9] to develop motion planners that explicitly reason about the unfolding topological patterns of multiagent interactions in crowded domains [48, 50, 52].

The high computational cost of these methods motivated us to look for more tractable alternatives. To this end, we employed alternative formalisms such as winding numbers [49] and angular momentum [51, 53] as approximations of multiagent interaction dynamics that retain the benefits of topological reasoning at a low computational cost. In this article, we expand and build upon our analysis of the performance of our Social Momentum motion planner [53], presenting additional quantitative and qualitative evidence.

2.4 Experimental Evaluation in Algorithmic Social Robot Navigation

Deployment and experimental testing in the presence of humans is an important component of evaluating social robot navigation frameworks. In this section, we summarize the methodologies applied by published studies over the past three decades and argue for the uniqueness of the experimental procedure followed by our work. Broadly, we observe three main trends in existing validation methodologies: (a) works illustrating proof-of-concept demonstrations in lab environments or public settings, (b) studies presenting controlled experimental validation in the lab, and (c) field studies conducted in public environments.

2.4.1 Experimental Demonstrations. A number of works validate their proposed approaches via proof-of-concept demonstrations in the presence of humans. For instance, Bennewitz et al. [8] report a series of 10 experiments involving interactions between a mobile robot and navigating humans under semi-controlled settings in a hallway. Sisbot et al. [66] document a series of navigation interactions between a robot running their framework and a human in a lab environment. Park et al. [60] test their control framework on a robotic wheelchair inside a corridor of an academic building and report a set of successful collision-avoidance encounters. Kretzschmar et al. [42] also deploy their model on a robotic wheelchair documenting a set of experiments in a narrow hallway under controlled settings. Chen et al. [14] document an experimental demo involving a robot navigation experiment in a crowded area of an academic building.

2.4.2 Lab Studies. Pacchierotti et al. [59] test their control framework with a study involving an autonomous robot navigating next to human subjects at a corridor under controlled settings. They present their findings from the interactions of 10 participants with a robot exhibiting different navigation strategies corresponding to different passing distances. Kirby et al. [39] and Kirby [38] present a user study involving 27 human subjects navigating alongside a robot in an academic hallway. Kruse et al. [43] document a series of interactions between a mobile robot and a human subject in a lab study involving 10 participants. Truong and Ngo [73] document a series of interactions between a robot running their planning algorithm and human participants in a lab environment. Our past work [51] featured an experimental validation of our planning framework [53] in a lab experiment involving interactions between a navigating robot and three human subjects at a time in dense navigation settings, yielding a total sample of 105 participants. Lo et al. [45] evaluate a series of robot collision-avoidance strategies on a self-balancing mobile robot in a lab study with 98 human subjects.

2.4.3 Field Studies. Another common approach involves deploying robots in the wild in public environments. Burgard et al. [10] deployed the Rhino guide robot in the Museum of Bonn in Germany in 1997 and documented thousands of in-person and virtual interactions between the robot and visitors over 47 hours of runtime spanning 6 days. Thrun et al. [71] deployed Minerva, a second-generation robot tour guide, at a Smithsonian Museum in Washington, DC, in 1998 and also documented thousands of interactions with visitors for two weeks. Both studies documented statistics related to performance, collision avoidance, and visitors' impressions of the robot. Foka and Trahanias [23] report logs and performance aspects upon running their robot for 70 hours in

an indoor academic building. Shiomi et al. [65] test their navigation framework through a 4-hour field study in a shopping mall. Trautman et al. [72] test their navigation framework on a robot in a field study comprising 488 robot runs in a crowded cafeteria. Kato et al. [36] test their approach on a humanlike robot employee in a crowded mall and record interactions with 130 people. Kim and Pineau [37] evaluate their framework on a robotic wheelchair in a crowded hallway over 10 field runs.

2.4.4 Our Approach: Extensive Subjects Research. This article documents our efforts in experimentally validating our *Social Momentum* planning framework [53]. Our experimental evaluation is split across two main studies: (a) an online, video-based user study that involved 180 human users and (b) a lab study featuring 105 human subjects navigating in close proximity with a telepresence robot [1]. To the best of our knowledge, our evaluation is unique in terms of sample size, experimental settings considered (close but controlled interactions in a dense environment), goals (evaluation of group interactions with a navigating robot), and thoroughness. We employ both quantitative and qualitative measures of interaction, aiming at documenting an in-depth, holistic insight over the features of our approach. Our empirical findings and remarks may help inform the design of future studies for evaluating social robot navigation frameworks.

3 FOUNDATIONS

Consider a robot r navigating in an obstacle-free workspace $\mathcal{Q} \subset \mathbb{R}^2$ where n human pedestrians are also navigating. Denote by $q \in \mathcal{Q}$ the state of the robot and by h_i the state of a human pedestrian $i \in N = \{1, \ldots, n\}$. The robot starts from a state $s_r \in \mathcal{Q}$ and navigates toward a destination $d_r \in \mathcal{Q}$. Similarly, human i starts from state $s_{h_i} \in \mathcal{Q}$ and navigates toward a destination $d_{h_i} \in \mathcal{Q}$. Each agent $a \in \{r, h_1, \ldots, h_n\}$ is following a control policy of the form $\pi_a : \mathcal{Q} \rightarrow \mathcal{V}_a$ mapping their current state q_a to a control input (velocity) $v_a \in \mathcal{V}_a$, where V_a is a space of velocities. This policy is assumed to account for (a) progress to destination, (b) collision avoidance, and (c) social compliance. Agents are not aware of each other's policies, destinations or trajectories *a priori* but are assumed to be able to perfectly observe each others' state. Our goal is to design a policy π_r that enables the robot r to reach d_r while avoiding hindering co-navigating humans. In particular, we are interested in enabling the robot to avoid collisions but also navigate in a way that is perceived as socially competent by nearby humans.

3.1 Socially Competent Navigation as Cooperative Intent Expressiveness

How can we engineer robot motion that is perceived as socially competent in a dynamic multiagent environment? Although there is no clear consensus among researchers about an exact and unifying definition of what constitutes *socially competent* behavior in a navigation domain, we build our framework around foundations extracted by studies from the social sciences and human–robot interaction research. In particular, our perspective is driven by three key observations: (a) social order in pedestrian navigation in crowds appears to rely heavily on cooperation [80], (b) from an early age, humans develop mental models that enable them to interpret observed actions by assigning them context-specific goals [16], and (c) intent-expressive robot motion in joint human–robot collaboration tasks improves performance [12]. Based on these insights, a central thesis in our approach is that under the assumption of human cooperation (rationality) in navigation, intent-expressive robot motion with respect to an appropriately defined notion of a *goal* may enable more efficient, minimally disruptive human–robot interactions.

3.1.1 Cooperation. We follow the insights of the sociology studies highlighting the cooperative nature of human navigation as described in the *Mini-Max Hypothesis of Urban Life* [35] or the *Pedestrian Bargain* [80]. We view these insights as guiding principles to our design. By explicitly

modeling cooperation, we can enable a robot to cooperate with and expect cooperative behavior from nearby humans, much like humans themselves do. We hypothesize that by doing so, we could enable robots to blend in more naturally in pedestrian spaces.

For our purposes, cooperation corresponds to the shared responsibility between the robot and humans for avoiding collisions with each other.

3.1.2 Inference. Cooperation in crowd navigation relies heavily on the extraction of inferences about the motion of others. For humans, these motion inferences may be quite sophisticated, conducted by mechanisms shaped by evolution, personal experience, context, the understanding of biomechanical constraints, and so on. However, at their core, these inferences seem to follow a specific blueprint: Humans, from an early age tend to innately interpret observed actions as goal-directed [16] in a given context.

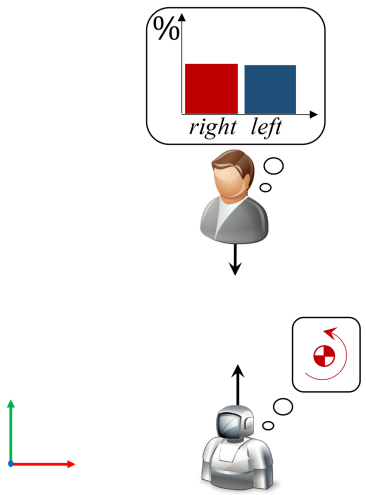
Despite their typically limited past experiences with robots and understanding of robot kinematics, dynamics, and computational capabilities, we hypothesize that non-expert humans may still be able to connect simple robot actions to robot intent. Conversely, in our setup, we expect the human motion to be indicative of the human intent.

3.1.3 Legibility. Legibility [17], or *Readability* [12], is an important property of motion in tasks conducted in the presence of a human observer. Dragan and Srinivasa [17] defined Legibility as the property of motion that enables an *observer* to infer quickly and confidently the correct *goal* of an *actor*, given observation of the actor's past *actions*. Humans employ a variety of modalities to express intent: Body posture, eye gaze, gestures, and verbal communication are only a subset of the information streams that humans may leverage to broadcast intention signals. Robots are limited in the set of modalities available for communication of their intentions by their design. For this reason, many human–robot interaction applications feature implicit communication mechanisms [40] encoding robot intent into otherwise purely functional motion. In these applications, implicitly communicative robot motion has been shown to enable effective human–robot collaboration [12] and reduced planning effort for humans [11].

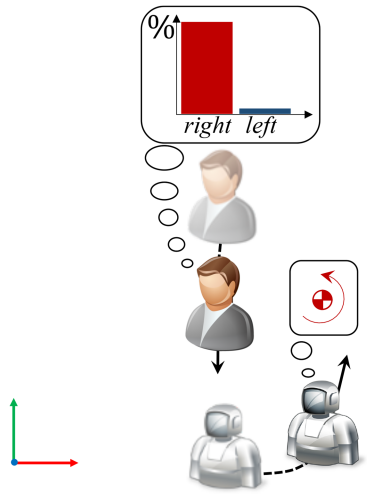
In our application, we assume that the robot may only use the modality of navigation path shape to communicate, due to its universality in mobile robot platforms.

3.1.4 What Is the Proper Notion of a Goal? A central notion in the emergence of inferences [16] and the generation of legible motion [17] is the notion of a *goal* or *intent*. Existing works on legible motion generation tend to associate the notion of a goal or intention with a point in a configuration space (e.g., References [17, 43]). In a static and structured environment, where the dynamics of interaction among agents is predictable or known *a priori*, this is a well-motivated modeling decision, as the observers' belief could be assumed to be an isolated relationship between an observed motion and a potential destination. However, in dynamic and unstructured environments, such as typical pedestrian navigation domains, where the dynamics of interaction among multiple agents is rich, knowledge only of an agent's destination may be insufficient to inform others of the agent's immediate behaviors. This highlights the need for a new consideration of legibility that captures interactions with neighboring agents.

Past work of ours has employed topological representations such as braids [48, 50, 52] and topological invariants [49] to encode intentions in multiagent navigation settings. However, the high computational cost associated with running those algorithms motivated us to pursue tractable approximations that would retain the salience of topological features. To this end, we developed the SM planning framework [53], which employs a physics-inspired approach to encoding the topology of passing side in navigation as the angular momentum between a pair of agents.



(a) Despite the ambiguity, the robot detects a slight inclination towards the right hand side.



(b) The robot's *legible* maneuver resolves the ambiguity and reinforces a right hand side passing.

Fig. 3. Social Momentum. A human and a robot navigate toward opposing sides of a workspace. The initial configurations of the agents make it hard for the human to predict the emerging avoidance strategy ("right" or "left"). The robot detects a slight inclination toward the "right" strategy and acts legibly to facilitate human inference.

4 SOCIAL MOMENTUM: A PLANNER FOR LEGIBLE ROBOT NAVIGATION

In this section, we recap the SM planning framework [53]. The planner is based on modeling pairwise passing-side intentions as angular momenta. Under this model, our planner generates legible motion that attempts to reconcile the (potentially incompatible) human–robot passing-side intentions.

4.1 Angular Momentum for Collision Avoidance

Consider the scene of Figure 3, where a robot r and a human h navigate in an obstacle-free workspace. The robot lies at q_r, moving with velocity v_r, whereas the human is positioned at q_h and moves with velocity v_h. The two agents form a dynamical system; representing them as unit-mass particles, we may derive the angular momentum of the system as

$$L^{rh} = p_r^C \times v_r + p_h^C \times v_h, \tag{1}$$

where

$$p_r^C = q_r - p_C, \qquad p_h^C = q_h - p_C, \tag{2}$$

are agents' positions, defined with respect to their center of mass,

$$p_C = (q_r + q_h)/2. \tag{3}$$

For a system of two agents on the horizontal plane, the angular momentum is a 3-vector normal to the plane, pointing along the positive direction of the z-axis for counterclockwise agent rotations (see Figure 3) and along the negative direction of the z-axis for clockwise rotations. These correspond to the agents passing on the right- and left-hand sides of each other, respectively. The magnitude of the momentum depends jointly on the distance between the agents and also on the relative angle between agents' velocities. Inspecting Equation (1), we can see that larger

distances and antiparallel velocities increase $|L^{rh}|$. Thus, the angular momentum is indicative of the tendency of two agents for picking a passing side—the larger the magnitude $||L^{rh}||$, the higher the certainty over the passing side given by the sign of L^{rh}_z, the z component of L^{rh}.

4.2 The Social Momentum Algorithm

This intuition represents the key insight underlying the design of our planner. By contributing toward increasing the magnitude of the angular momentum along its current directionality, an agent reinforces the currently established passing side. Generalizing to a multiagent environment, by selecting an action $v_r \in \mathcal{V}_r$ that increases the magnitude of the pairwise momenta with all other agents h_i, $i \in N$, the robot reinforces the currently established sides of passing. To enable this decision-making strategy, we construct the *Social Momentum* objective $\mathcal{L} : \mathcal{V}_r \to \mathbb{R}$, defined as:

$$\mathcal{L}(v_r) = \begin{cases} \sum_i w_i ||\hat{L}^{rh_i}(v_r)||, & \text{if } (L^{rh_i})^\top \hat{L}^{rh_i}(v_r) > 0, \ \ \forall i \in N \\ 0, & \text{otherwise} \end{cases}, \tag{4}$$

where $\hat{L}^{rh_i}(v_r)$ denotes the expected pairwise momentum between agents r and h_i, upon the robot taking an action in consideration, v_r, and the human h_i moving with its current velocity v_{h_i}. L^{rh_i} is the current pairwise momentum between r and h_i, and $w_i \in \mathbb{R}$ is an importance weight prioritizing reacting to agents that are closer. The sign of the quantity $(L^{rh_i})^\top \hat{L}^{rh_i}(v_r)$ determines whether the projected updated momentum $\hat{L}^{rh_i}(v_r)$ is pointing toward the direction of the current momentum L^{rh_i}. A positive sign corresponds to an action that preserves the current momentum sign and thus the currently preferred pairwise avoidance protocol. A negative sign indicates inversion of the established pairwise avoidance protocol, which is undesired. For this reason, an action that results to inversion of at least one pairwise momentum is assigned a score of zero.

The *Social Momentum* objective is the core of the *Social Momentum* planning algorithm (SM). The planner balances between intent-expressiveness (represented by $\mathcal{L}$) and efficiency, represented by an efficiency function $\mathcal{E}$. The algorithm relies on frequent replanning: At every planning cycle, it selects an action that corresponds to the optimal compromise between progress to the agent's destination and legible avoidance of others as follows:

$$v_r^* = \underset{v_r \in \mathcal{V}_r}{\operatorname{argmax}} \, \mathcal{E}(v_r) + \lambda \mathcal{L}(v_r), \tag{5}$$

where $\lambda \in \mathbb{R}$ is a temperature parameter accounting for scaling and weighting of the two quantities. As shown in Equation (4), if an action results in inversion of momentum between the robot and at least one other agent, then it will be scored minimally (zero). In case all actions are momentum-inverting, the action selected for execution will be purely based on the efficiency objective for that timestep. Under the assumption of cooperation [80], humans will share the responsibility of avoiding an imminent collision with the robot, and they will react with actions that contribute toward collision avoidance. Gradually, the system of agents will converge to an equilibrium satisfying the updated preferences of all agents.

We define the progress function $\mathcal{E} : \mathcal{V}_r \to \mathbb{R}$ to be the inverse of the length of the unobstructed line to the agent's destination. The action space $\mathcal{V}_r$ comprises a pre-sampled discrete set of actions of finite duration that are executable by the agent. Finally, the weight w^{rh_i} is chosen as the inverse of the distance between the robot r and agent h_i.

Algorithm 1 describes the SM algorithm in pseudocode. At the start of execution, the algorithm initializes a list of reactive agents $\mathcal{R}$ to contain all perceived human agents. While the robot is further than a distance threshold δ from its destination, a replanning cycle runs. The function `check_collision` tests the action space $\mathcal{V}$ for collisions with human agents and returns a set

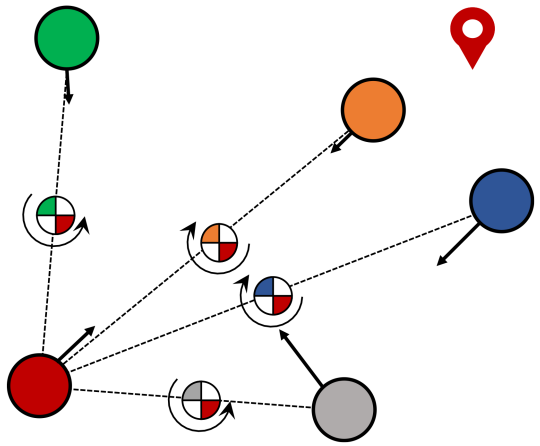


Fig. 4. Social Momentum: The robot (depicted in red color) is moving toward its destination (represented as a red landmark). while complying with its pairwise momenta with all other agents.

ALGORITHM 1: SM($q_r, Q, \mathcal{V}_r, d_r, AtGoal, v_r^*$)

Input: q_r – robot state; $H = (h_1, \ldots, h_n)$ – list of human agents; $Q = (q_{h_1}, \ldots, q_{h_n})$ – list of human states; $V_h = (v_{h_1}, \ldots, v_{h_n})$ – list of human velocities; $\mathcal{V}_r$ – robot action set; d_r – robot destination; λ – optimization weight; δ – destination distance threshold.

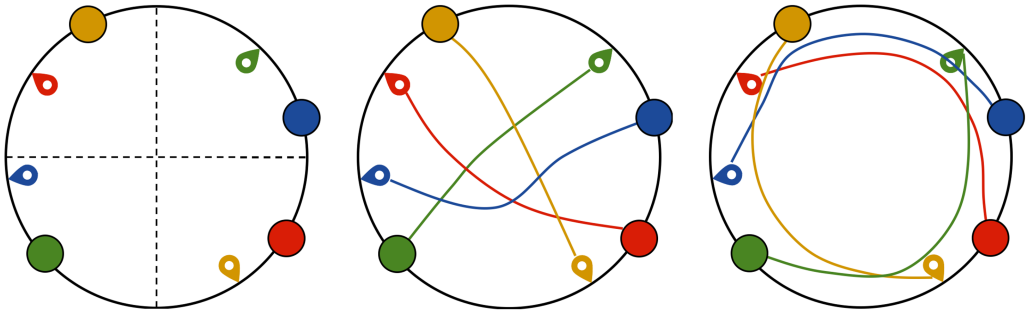
Output: v_r^* – velocity selected

$\mathcal{R} \leftarrow H$ ▷ List of reactive agents
while $||q_r - d_r|| > \delta$ **do**
  $\mathcal{V}_{cf} \leftarrow$ `check_collision`$(\mathcal{R}, Q, \mathcal{V})$
  $\mathcal{R} \leftarrow$ `update_reactive`$(Q, V_h, \mathcal{R})$
  if $\mathcal{R} \neq \varnothing$ **then**
    $v_r^* \leftarrow$ `optimize_momentum`$(Q, \mathcal{V}_{cf}, \lambda, d_r)$
  else
    $v_r^* \rightarrow$ `optimize_efficiency`$(\mathcal{V}_{cf}, d_r)$
return v_r^*

$\mathcal{V}_{cf} \subseteq \mathcal{V}$ of collision-free actions under the assumption that agents maintain their velocities for the next timestep. Then, function `update_reactive` determines the subset of agents $\mathcal{R}$ to which the planning agent should be reacting: Only agents that lie in front of the planning agent (i.e., agents lying from $-90°$ to $+90°$ from the robot's direction of motion; see Figure 4) are considered. In case $\mathcal{R} \neq \varnothing$, the planning agent determines a legible action v_r^* by compromising between *Progress* to destination and *Social Momentum* (function `optimize_momentum`); otherwise, the algorithm switches to progress maximization mode (function `optimize_efficiency`). Termination occurs once the agent comes closer than δ to its destination.

4.3 Simulation Study

We conduct a simulation study in which we examine the properties of motion generated by SM in scenarios involving multiple agents navigating in close proximity. Our goal is to understand features of collective behavior that could affect the perceptions of human observers over agents' intentions.



(a) Agents' initial states (circles) and destinations (landmarks).

(b) High topological complexity, low path irregularity.

(c) Low topological complexity, high path irregularity.

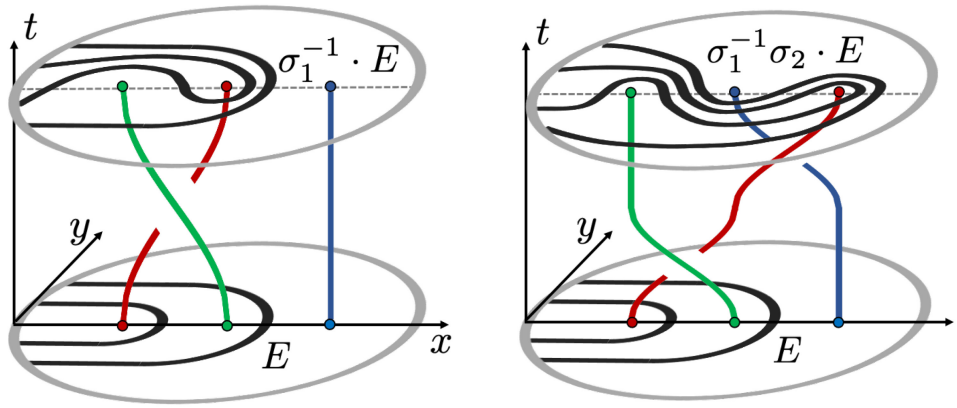
Fig. 5. Simulation setup. Four agents, represented as colored disks, are placed along the circumference of the circular workspace. Each agent is headed toward the landmark of the same color. Figure 5(a) illustrates the initial scenario state. Figure 5(b) depicts an execution with **high** Topological Complexity index but **low** Path Irregularity. Figure 5(c) depicts an execution with **low** Topological Complexity index but **high** Path Irregularity.

4.3.1 Experimental Setup. We consider a setup in which groups of holonomic homogeneous agents (agents running the same navigation algorithm) navigate in a shared circular workspace Figure 5(a). We specifically design navigation scenarios giving rise to complex multiagent encounters. Each scenario is generated through the following steps: (1) the workspace circumference is partitioned into n arcs of equal length; (2) each arc is assigned to an agent; (3) each agent is placed at a random, collision-free starting position on their arc; and (4) each agent is assigned a destination that is antipodal to their starting location and lies on the workspace circumference. The workspace boundary is a circle with a diameter of 5 m, whereas the agents are discs of diameter 0.6 m. We consider four different classes of scenarios, each corresponding to a different number of agents, ranging from three to six. For each class, we generate 200 scenarios at random.

4.3.2 Conditions. We consider three experimental conditions: (1) *Social Momentum*, (2) the **Social Force (SF)** model [32], and the (3) **Optimal Reciprocal Collision Avoidance (ORCA)** framework [76]. Each condition corresponds to a different algorithm executed by all agents. The three algorithms selected represent distinct algorithmic designs resulting in behaviors with qualitatively distinct properties: ORCA is theoretically and empirically shown to prioritize efficient behaviors; SF is purely reactive and often follows highly suboptimal paths to avoid collisions; and SM was designed to exaggerate motions to increase intent expressiveness over passing sides. Note that the SF and ORCA baselines are algorithms designed to produce realistically looking crowd simulations and not to produce legible motion in multiagent scenarios. However, to the best of our knowledge, our planner is the first framework designed to automatically produce legible motion in multiagent settings of close interaction. In the absence of a directly relevant baseline, SF and ORCA allow us to directly relate to the literature as they are commonly employed for benchmarking algorithms in the areas of multirobot navigation and social navigation [21, 42, 46, 73, 77].

We empirically tuned the parameters for all algorithms to ensure a fair comparison. To tune the parameters, we payed special attention at the emerging average clearance and speed. Note that for different algorithms, the notion of preferred or maximum speed affects performance to different degrees. We note the parameters here for completeness:

- SF: $v^0 = 1.5$ m/s, $\sigma = 0.5$ m, $Vab = 21$ m^{2}s^{-2}, $tau = 0.4$ s, $v_max = 2.5$ m/s, $UaB^0 = 5$ m, $R = 0.2$ m, $\phi = 180°$, $c = 1$, $\Delta t = 0.1$ s.



(a) Curve diagram for σ_1^{-1}. (b) Curve diagram for $\sigma_1^{-1}\sigma_2$.

Fig. 6. Curve diagrams for braids of different complexity (top). The braid $\sigma_1^{-1}\sigma_2$ depicted in (b) is more complex ($TC = 2$) than the braid σ_1^{-1} ($TC = 1.585$) shown in (a). This is reflected in the increased number of intersections between the curve diagram $\sigma_1^{-1}\sigma_2 \cdot E$ and the x-axis (dotted line).

- ORCA: `neighborDist` = 3, `maxNeighbors` = 10, `timeHorizon` = 2 s, `timeHorizonObst` = 2 s, `radius` = 0.35 m, `maxSpeed` = 1 m/s.
- SM: We used an action set of 50 actions of fixed speed $v = 1.2$ m/s, $\lambda = 0.11$. The weight w_i was defined to be the inverse of the distance between the robot and agent i. We normalized these weights across agents so that $\sum_i^n w_i = 1$. We also normalized the metrics of Equation (5) to control their range of values.

4.3.3 Metrics. We consider two metrics, each targeting a different aspect of trajectory quality: (a) the Topological Complexity index [19] and (b) the Path Irregularity index [30]. These metrics complement each other, capturing respectively topological and geometric properties of multiagent interaction that may drive human perceptions with respect to agents' intentions.

Topological Complexity Index. In past work [52], we have showed that the motion of multiple navigating agents can be abstracted into a topological braid [9] by projecting their trajectories (represented as sequences of (x, y, t) tuples) onto a selected spatiotemporal plane and tracing any crossings that emerge among them. The emerging object is a set of strings entangled with each other in a way that reflects the way agents' trajectories entangled over time as agents navigated toward their destinations. The pattern that a braid describes can be simple when the strings do not entangle significantly with each other or complex when they do. This can also be true for the trajectories of agents—when agents mix with each other in close proximity, their trajectories become more strongly entangled over time (see Figure 5). The Topological Complexity index, proposed by Dynnikov and Wiest [19], quantifies the complexity of the entanglement of a topological braid. In this work, we use it to quantify the complexity of the entanglement of agents' trajectories in a navigation domain: First, we map a set of agents' trajectories onto a topological braid β by projecting their trajectories onto the x-t plane[1] and scaling them to the time window from $t = 0$ to $t = 1$ as described in our earlier work [52]. It might be helpful to think of this projection as the the motion of a set of particles from time $t = 0$ to time $t = 1$ (Figure 6 depicts example braids). Then, denote by D^2 a closed disk surrounding the initial positions of the particles and by E a set of

[1]As long as we are consistent throughout experiments, the choice of projection plane does not make a difference [67].

$n-1$ disjoint arcs, anchored on the disk, that clearly separate the particles for time $t = 0$, defining n distinct regions in the disk (see Figure 6). Assume that these regions are rigidly attached to the particles. As the particles follow the motion described by the braid β from $t = 0$ to $t = 1$, the regions dynamically deform.

A *curve diagram* of a braid is the image $D = \beta \cdot E$ of E corresponding to the regions obtained by applying the motion described by β on E (see Figure 6). The *norm* of a curve diagram D is defined as the number of intersections of D with the x-axis. The **Topological Complexity (TC)** index of a braid $\beta \in B_n$ is defined as

$$TC(\beta) = \log_2(||\beta \cdot E||) - \log_2(||E||). \tag{6}$$

This expression is equivalent to the logarithm of the *increase* in the number of intersections between the x-axis and the arcs E upon the application of a braid β.

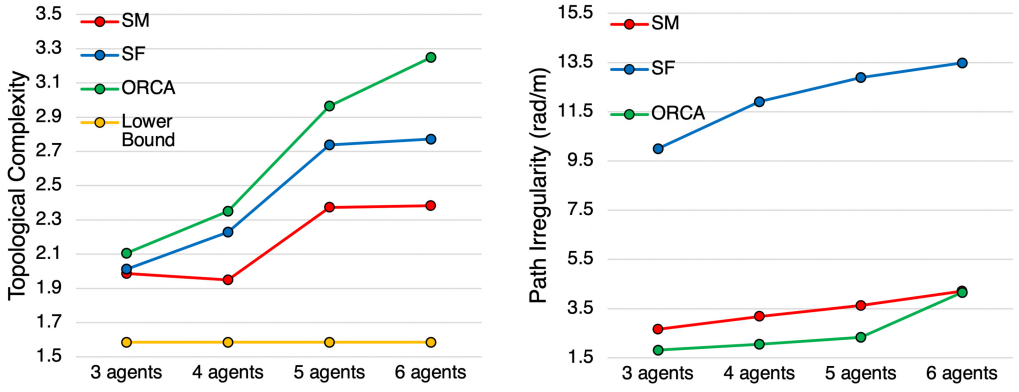
Intuitively, TC measures how much the motion corresponding to a braid twists or mixes the regions around each initial point, describing (in our context) how much agents move past each other. Figure 6 depicts the curve diagrams induced by applying motion described by two different braids on the canonical curve diagram E. Figure 5(b) and Figure 5(c) illustrate examples of high and low Topological Complexity, respectively. We see that high TC values describe scenarios in which agents directly *mix* with each other as they navigate toward their destinations, whereas low TC values describe scenarios in which agents avoid interacting with each other. Qualitatively, we may observe that high-TC behaviors make it hard to predict agents' passing sides—agents navigate through the center of the workspace without showing inclination toward a collision-avoiding passing side. In contrast, low-TC agents appear easier to infer—agents show early and consistent inclination toward a passing side. These properties make TC relevant to studying the Legibility of robot motion in multiagent scenarios. Section 5 investigates this connection using data from human subjects.

Path Irregularity Index. The path irregularity index, proposed by Guzzi et al. [30], characterizes the geometric inefficiency of agents' trajectories. Specifically, **Path Irregularity (PI)** is computed as the average amount of angular divergence between an agent's heading and its direction to its destination per unit path length, averaged per agent:

$$PI(\Xi) = \frac{1}{n}\sum_{i}^{n}\frac{\sum_1^T \theta_t^i}{L_i}, \tag{7}$$

where $\Xi = (\xi_1, \ldots, \xi_n)$ is a tuple containing agents' trajectories ξ_i, $i \in \mathcal{N}$, θ_t^i is the angle difference between the velocity of agent i and the direction to its goal at time t, L_i is the total length of the path followed by agent i, and T is the total time of the experiment. Measured in rad/m, the higher PI gets, the higher the geometric inefficiency of an execution. Figure 5(b) and Figure 5(c) illustrate examples of low and high Path Irregularity.

4.3.4 Analysis. Figure 7(a) depicts the average *Topological Complexity* for each planner and class of scenarios considered. The *Topological Complexity* of SF and ORCA appear to consistently rise with the number of agents. In contrast, SM exhibits a slower rise: The transitions between three and four agents and between five and six agents have almost constant complexity, with the only rise taking place in the transition between three and four agents. Overall, SM achieves consistently lower topological entanglement with statistical significance, except from the case of three agents, where the scenarios are not geometrically challenging enough to yield significantly diverse behaviors. Detailed statistics of paired t-tests conducted for the SM-SF and SM-ORCA pairs are reported in Table 1.



(a) *Topological Complexity* for each scenario class. (b) *Path Irregularity* per agent for each scenario class.

Fig. 7. Comparative evaluation of the selected simulation frameworks (SM, SF, and ORCA) over four different classes of scenarios involving three, four, five, and six agents, respectively. Within each class, 200 randomly selected scenarios are executed and evaluated with respect to *Topological Complexity* (a) (a theoretical lower-bound baseline is included for reference) and *Path Irregularity* (b). Each point on the graph represents the average value over all 200 experiments for the class.

Table 1. Statistics of Paired t-Tests between Policies for Different Agent Numbers

t-Tests		*Topological Complexity*		*Path Irregularity*	
Num. of agents	*Pair*	t value	p value	t value	p value
3	SM-SF	−2.497	0.013	−26.397	<0.001
	SM-ORCA	−0.593	0.553	9.197	<0.001
4	SM-SF	−7.963	<0.001	−34.514	<0.001
	SM-ORCA	−5.740	<0.001	17.336	<0.001
5	SM-SF	−9.424	<0.001	−41.400	<0.001
	SM-ORCA	−5.395	<0.001	7.934	<0.001
6	SM-SF	−11.561	<0.001	−51.430	<0.001
	SM-ORCA	−5.250	<0.001	0.152	0.879

For reference, we also include a lower bound on Topological Complexity corresponding to a centralized planning baseline from our past work [52]. This baseline generates a path between agents' initial and final states involving the minimum number of swaps in the order of agents along the x-axis (the same axis used to generate the braids used for the computation of TC along all experiments). The lower bound value ($TC = 1.5850$) is an artifact of the selected scenarios (traversal between antipodal locations) and corresponds to the ideal case in which agents reach their destinations in a coordinated way that avoids trajectory entanglements (see Figure 5(c) for an example). In practice, we see that all planners are suboptimal with respect to the lower bound baseline, which reflects the price of no explicit communication across agents.

Figure 7(b) depicts the average Path Irregularity per agent, for each planner and class of scenarios considered. Although for all planners the path irregularity rises with the number of agents, each planner performs differently. The different performance of each planner is indicative of the distinct philosophies with which they have been designed. SF, lacking predictive mechanisms, yields significantly more irregular paths than SM and ORCA. ORCA achieves consistently the lowest path

irregularity, as a result of its geometrically optimal behavior, which in practice results in minimal divergence from the unobstructed line connecting an agent with its destination at any time. SM performs slightly worse than ORCA, as a result of its consideration of collision avoidance as a rotation; SM agents diverge from their shortest paths more often to convey intent. For the case of six agents, the geometric complexity of the scenarios is too intense even for ORCA, which performs almost equally to SM.

The findings from our simulation study suggest that SM produces generally less topologically complex trajectories in multiagent scenarios than two other baselines from the area of multiagent navigation.

5 SOCIAL MOMENTUM GENERATES LEGIBLE MOTION

In a multiagent domain, the actions of navigating agents contribute to the formation of inferences on others. For instance, from the motion of agents in Figure 5(b), it is unclear—until later in time—how they intend to avoid each other, as they all appear to be travelling through the center of the workspace. In contrast, the motion of agents in Figure 5(c) demonstrates an emergent organization that facilitates inference of how they intend to avoid collisions. These examples illustrate a possible connection between the notion of legibility in multiagent navigation and the measure of Topological Complexity. Here, following the discussion of Section 3.1.4, by *legible*, we refer to motion that clearly conveys the robot's intention over a passing side (i.e., *right*, or *left*). Qualitatively, from Figure 5, we see that executions of low topological complexity seem to be more legible; conversely, executions with high topological complexity appear to be less legible. Thus, given the findings of Section 4, we expect the behaviors generated by SM to be more legible to human observers.

To evaluate the validity of this expectation, we formally investigate the following hypothesis **H1** by examining its constituent sub-hypotheses **H1a** and **H1b**:

H1: *The SM framework generates multiagent navigation behaviors that human observers perceive as legible.*

- **H1a:** *Social Momentum produces multiagent trajectories of comparatively low topological complexity.*
- **H1b:** *Trajectories of low topological complexity are perceived as more legible by human observers.*

Section 4.3.4 has already demonstrated empirically that SM generates executions of low topological complexity compared to two other baselines. Thus, **H1a** is confirmed. To contextualize and validate this finding, we separately investigated sub-hypothesis **H1b** by conducting an online user study in which we asked human subjects to predict the evolution of simulated multiagent scenarios from partial observation.

5.1 Study Design

We designed an online user interface for a simple game around predicting agent motion. The interface played videos of simulated multiagent navigation scenarios and asked the user to predict which side a specific pair of agents would pass each other on (right or left) by clicking on a button corresponding to each side (see Figure 8). The interface incentivized fast, accurate responses through a scoring system that awarded points for accurate answers given early in the video (before the agents passed each other) and deducted points for incorrect or slow responses. Users were shown information about their performance relative to the median performance of previous users to further motivate healthy competition toward giving correct answers as quickly as possible. We employed a within-subjects design in which all users were shown the same set of

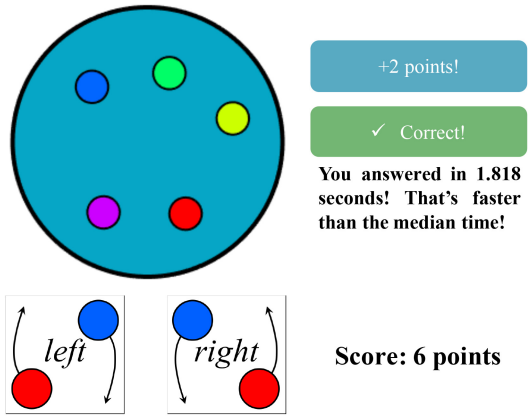


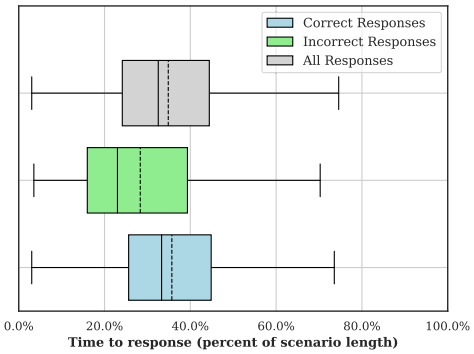
Fig. 8. Study interface: A video of a simulated scenario (large teal circle at top) plays and users predict how the red agent will avoid the blue agent (passing on its left or right) by pressing the corresponding button at the bottom. The user's score and relative performance statistics are displayed to motivate fast, accurate responses.

15 videos, presented in random order. Each video showed a distinct simulated multiagent scenario. All scenarios involved agents travelling between antipodal points on a circular workspace, incentivizing agents to plan nontrivial collision avoidance maneuvers. The scenarios were selected to span a broad range of Topological Complexity index values [19], ranging from 1.585 to 4.250, and ranged in duration from 6.3 to 15.7 s.

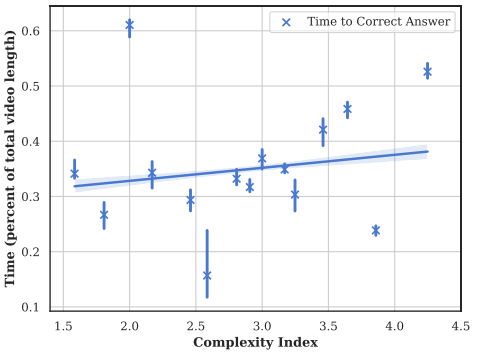
5.2 Analysis

We recruited 180 unique users via open recruitment calls on the social media platforms Reddit and Facebook without special selection criteria. Considering incomplete entries (entries in which users did not watch all 15 videos), we recorded a total of 2,704 video views. For each view, we recorded (a) the user's prediction (right or left) and (b) the time of that prediction (relative to the start of the corresponding video). We removed responses that were given either before the start of a video or after the end of a video from the data, resulting in a dataset of 2,647 responses. Figure 9(a) describes the distribution of prediction times across all videos, itemized with respect to correctness, for this dataset. The timing values are normalized to total video duration to allow for timing comparison between videos of different length. Overall, only 300 of 2,647 responses were incorrect, but we found that correct responses required significantly more time (paired Student's t-test, $t = 2.065, p = 0.039$).

Figure 9(b) demonstrates the effect of topological complexity on the time participants take to give a *correct* answer. We fit a linear model to the data using iteratively reweighted least squares, shown in Figure 9(b) as a blue line with a 95% confidence interval. The effect of topological complexity on the median time to correct answer is *positive* (slope 0.0236) and significant according to a Student's t-test ($t = 5.60$, $p < 0.001$). This finding suggests that as the topological complexity of an execution increases, users take more time to accurately predict the side of passing. Per the definition of legibility given in Section 3.1.3 and the original paper by Dragan and Srinivasa [17], legibility requires human observers to be able to infer an agents' intent (here, the agents' passing side) *quickly* and *correctly*. In this study, we see that correct predictions of the agents' passing side require more time(normalized to video length) for executions with higher topological complexity. Thus, executions with high topological complexity are *less legible*, as they require more time to predict correctly. This finding confirms sub-hypothesis **H1b**, which, in conjunction with **H1a** confirm



(a) Distribution of response times. Solid vertical lines indicate quartiles of response times; dashed lines indicate mean response times.



(b) Relation between topological complexity and the median time to correct prediction (normalized to scenario length).

Fig. 9. Distribution of response times (left) and relation between topological complexity and normalized time to correct answer (right). We see a statistically significant positive correlation, indicating that scenarios with greater topological complexity require more time to accurately predict.

H1. Overall, this confirmation suggests that within the multiagent navigation setting of this study, navigation behaviors generated by *Social Momentum* are perceived as more legible compared to the considered set of baselines.

6 EXPERIMENTAL EVALUATION OF MOTION PRODUCED BY SOCIAL MOMENTUM IN THE PRESENCE OF HUMANS

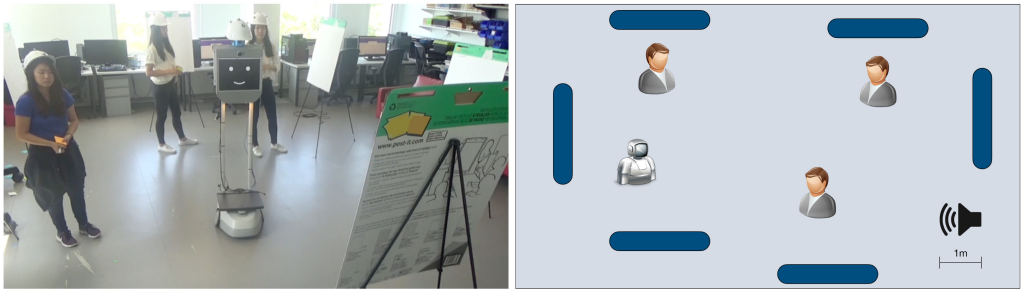
We conducted a lab study to understand the effects of the embodiment of our proposed navigation framework on human behavior and human perception. A video excerpt from our study can be found at this link: https://youtu.be/8aO4P6_OzW4.

6.1 Study Design

The lab environment afforded us with the ability to control human–robot interactions across space and time. In particular, we employed a design that enabled us to

- Enforce a setting of implicit, nonverbal social engagement among agents, similar to the type of interaction among walking pedestrians;
- Construct a moderately crowded scene that balanced close interactions with space for the robot to showcase its distinct navigation strategies (see Figure 10(a));
- Promote the emergence of nontrivial interactions, involving challenging collision-avoidance maneuvers between participants and the robot through the definition of rules;
- Motivate natural walking behaviors by not disclosing the real purpose of the study until the debriefing process and by manipulating participants' cognitive load through a background scenario and task;
- Bound the total duration to facilitate recruiting and minimize potential effects resulting from participants' fatigue.

6.1.1 General Procedure. Our study is organized into a set of experiment sessions. In each session, three different human subjects participate in a set of three experiment trials. Before the first trial, participants are asked to give written consent to confirm their participation and optionally to be video recorded. A member of our research team delivers the instructions and answers questions.



(a) Footage from a study session. (b) Schematic of the workspace.

Fig. 10. Illustration of the experiments conducted at our study sessions. Figure 10(a) shows footage from a study session. The setup comprises a telepresence robot [1] and a set of six easels representing machines in a fictional factory workspace. Three participants, wearing tracking helmets navigate between stations to perform fictional maintenance tasks on the machines. Figure 10(b) illustrates the experimental process: once a sound signal is broadcast, the humans and the robot start moving between different machines (shown in blue color). As they move, the perform collision-avoidance maneuvers.

During each trial, participants repeatedly visit a set of stations inside a rectangular workspace of area 16 m^2 (see Figure 10(b)), driven by a fictional scenario (see Section 6.1.2). In parallel, a mobile robot (a Suitable Technologies Beam Pro, equipped with a quad core i7 processor laptop from 2017), shown in Figure 10(a), also moves between the stations as part of the scenario. During each trial, we track the human and robot trajectories using an overhead motion capture system of six high-accuracy (<1 mm), high-fidelity (frequency 180 Hz) cameras and record the experiment if participants gave consent. Real-time tracking is enabled through the use of construction helmets (see Figure 10(a)) with reflective markers. After each trial, participants are asked to fill in a questionnaire, containing questions about their impressions from their interactions with the robot. At the end, participants are asked to provide basic demographic data and information regarding their prior experience with user studies and robotics technology. Participants are then debriefed, compensated, and dismissed.

6.1.2 Background Scenario and Task. Participants are asked to imagine that they are workers in a factory (the factory setting helps justify the tracking helmets) where the robot is a supervisor. The factory environment (lab workspace) contains six machines, represented as easels, spread around the workspace, as shown in Figure 10(b). Each worker is given a distinctly colored marker and a contrasting, distinctly colored set of sticky notes. The duty of a worker is to *perform* maintenance tasks to machines and *assign* tasks for other workers to perform. Assigning a task is done by drawing a square on the pad of an easel, whereas performing a task is done by posting a sticky note inside a square drawn on an easel pad. Participants are asked to perform only tasks represented with squares of color that matches the color of their sticky notes (see Figure 10(a)).

6.1.3 Trial Description. Before the start of each trial, participants are randomly positioned next to different machines, and the robot is placed in the middle of the workspace, as shown in Figure 10(a). A trial is organized into a set of maintenance cycles, initiated by a gong sound, played by the robot. Each time the gong is played, participants are instructed to leave their machines toward a non-adjacent machine of their choosing that is not occupied by another participant. Each time participants reach a new machine, they are instructed to perform up to one pre-assigned task (if one exists) and assign a new task. At the same time, the robot is navigating in the workspace by following the same rules of transitioning between stations, i.e., it only moves to a randomly picked,

non-adjacent machine when the gong sound is played (note that it is possible that the robot might pick a machine occupied by a human). For synchronization purposes, the gong sound is played when the robot is ready to move toward its next machine. Each trial lasts exactly three minutes, during which an ambient factory sound track is played.

6.1.4 Conditions. All participants were exposed to the same three conditions (within-subjects design), each corresponding to a different navigation strategy, executed by the robot. To account for potential ordering effects (i.e., due to fatigue, frustration, learning), the condition order was methodically varied and approximately equally spread across all sessions. We compared the performance of SM with ORCA [76] and **teleoperation (TE)**. These strategies were mainly selected due to the diversity of decision-making principles that they represent. SM represents an intention-aware motion planning paradigm, whereas ORCA (called OR for brevity from now on) is designed to be efficient, and TE is designed to be easy for an operator to use and to appear humanlike. Additional influences on our selection of navigation strategies included (1) the fact that OR constitutes a common benchmark and work of reference for multiagent simulations (e.g., References [14, 27, 42, 53]); (2) the existence of an open source, optimized C++ implementation of OR; and (3) the widespread use of telepresence robot platforms through teleoperation via their navigation interfaces. The complexity of a real-world pedestrian environment would pose a significant challenge to any of these navigation planners. However, we believe that an extensive and comparative evaluation of planners with distinct philosophies could provide us with significant insights and experience for the design of the next generation of social navigation planning algorithms.

6.1.5 Implementation of Teleoperation. The *teleoperation* strategy (TE) was implemented through the official navigation interface provided by the manufacturer [1], using the arrow keys on a standard laptop keyboard. This interface contains two live streams of video, providing the teleoperator with real-time video streams of a forward, wide-angle field of view (top) and a floor view (bottom). Navigation commands may be issued via the laptop keyboard's arrow keys or with a mouse. Commands are demonstrated as projected future trajectories on the video streams, providing visual feedback to the operator (but not to the study participants). The teleoperation condition was executed by the same member of our research team across all sessions, from a remote location (outside of the lab). The teleoperator had significant prior experience with the navigation interface. Before collecting data for our final dataset, we completed a total of seven rounds of pilot sessions under different variants of the final study setup. Thus, by the time we officially started the study, the teleoperator had reached a skill level that qualitatively appeared to be appropriate for the needs of the condition. Although it is hard to precisely quantify the operator's skill level, his experience was on the order of several hours of operation prior to the start of the study, and thus we do not believe that his performance evolved over the course of the study as a result of learning.

6.2 Hypotheses

Upon experimenting with the three navigation strategies considered (simulations conducted with SM and OR, and personal teleoperated teleconference sessions with the Beam), we observed very different patterns of decision making. We interpreted these patterns as the result of the different design principles and objectives underlying the operation of each framework: OR was developed to produce efficient, realistic simulations of virtual multiagent scenarios; SM was designed to generate legible robot motion in dynamic multiagent environments; TE was based on a navigation interface [1], specifically designed to allow non-expert users to control a robot intuitively. To the best of our knowledge, these strategies have never been tested against each other under challenging, multiagent, experimental settings. It was unclear how close interaction between the robot and different human participants would affect the motion generated by the

different strategies. Furthermore, it was uncertain how humans would react to different behaviors exhibited by the robot and how this interaction would affect overall performance for both humans and the robot. Using the dataset generated by our study, we explore these questions by examining the validity of the following hypotheses:

H2 - Robot Performance: In close interactions with humans:

- **H2a:** OR generates the most geometrically efficient paths. This is motivated by the documented efficiency of ORCA [76] compared to other baselines in multiagent simulation.
- **H2b:** SM generates high-acceleration paths. This is motivated by the tendency of SM to produce suboptimal, exaggerated motion in an effort to convey intent over a passing side.
- **H2c:** TE generates the most energy-efficient paths. This is motivated by the superior long-term planning capabilities of humans and by the robot teleoperation interface that enables comfortable and efficient navigation.

H3 - Human performance: Humans navigating in close proximity with the robot:

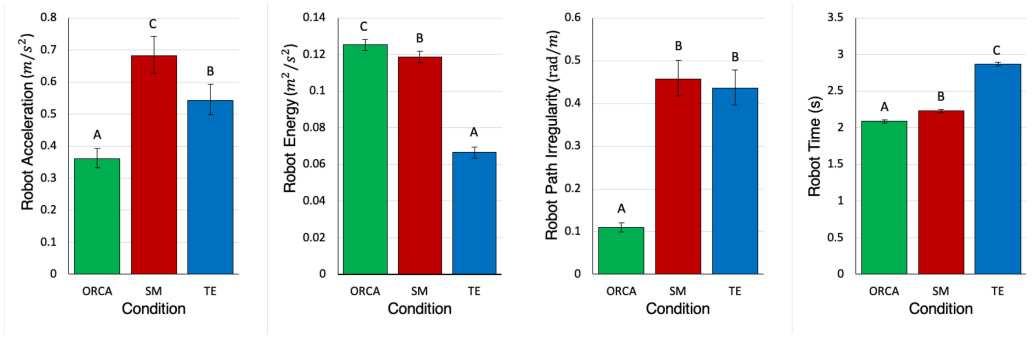
- **H3a:** follow the lowest-acceleration paths when the robot runs SM. Our insight is that by being legible, SM could enable humans to anticipate its motion more accurately. This would mean that participants would have to adjust their speed levels fewer times, thus following lower-acceleration paths.
- **H3b:** spend the least energy when the robot runs TE. Our insight is that a human-operated robot could exhibit behaviors with anthropomorphic traits that could be observable by participants. This could enable them to understand and trust the robot behaviors more than the baselines, which we would expect to see reflected in the robot's lower energy.
- **H3c:** spend the most energy when the robot runs OR. Our insight is that by striving for efficiency, a robot running ORCA would end up expecting humans to spend more effort into avoiding collisions with it.

H4 - Group performance: Global group (human and robot) behavior under SM: results in trajectories of lower Topological Complexity than the other two conditions. This is motivated by the finding of the online study suggesting that SM generates motion of lower TC than its baselines.

6.3 Datasets

We conducted 35 experiment sessions, in which a total of 105 human subjects were exposed to all three conditions. Subjects were recruited from a university population (Cornell University), through a centralized, university-run subject-recruitment website as well as through fliers posted across campus. The subjects (59 female, 45 male, 1 unidentified) were 21.45 years old on average ($SD = 3.19$ years) with their age ranging from 18 to 33 years. About half of them (57) had prior experience of user study participation and they rated their familiarity with robotics technology with an average of 2.47 ($SD = 1.27$) on a 5-point Likert scale.

We collected a dataset comprising the trajectories of all 105 participants and the robot across all trials. Focusing on dynamic interactions of close proximity, we split this raw dataset into two datasets of trajectory segments: (a) a dataset comprising 1,033 robot trajectory segments of close interaction with humans (minimum distance $d < 1$ m) and (b) a dataset comprising 1,566 human trajectory segments of close interaction with the robot (also, of minimum distance $d < 1$ m). We also collected a dataset comprising the responses of all 105 participants to a questionnaire, containing Likert-scale style questions, based on the instrument of Bartneck et al. [5] and short response questions.



(a) Acceleration (b) Energy (c) Path Irregularity (d) Time

Fig. 11. Expected means and confidence intervals for robot trajectory acceleration (a), robot trajectory energy (b), robot path irregularity (c), and robot time (d), averaged per trajectory segment. Quantities labeled with different letters (A, B, C) come from significantly different distributions (Tukey's HSD test, $p < 0.05$).

Table 2. Effect of Navigation Strategy on Robot Behavior

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
a	56.09	28.05	2	65.12	58.94	<0.0001
E	0.7083	0.3541	2	1015	440.1	<0.0001
$\mathscr{E}$	0.05796	0.02898	2	999.1	4.825	0.008213
PI	454.4	227.2	2	1012	355.3	<0.0001
τ	116.5	58.27	2	1016	1056	<0.0001

6.4 Trajectory Analysis

We analyze the trajectory dataset using a set of trajectory quality measures from relevant literature [30, 42, 53], computed over fixed timestep intervals (100 timesteps, totaling 0.2 s). In particular, we computed (1) the average *Acceleration* per segment, a; (2) the average *Energy* per segment, E, where energy is defined as the integral of the squared velocity of an agent throughout its trajectory; (3) the minimum *Distance* between the robot and any humans per segment, d; (4) PI, measuring the total amount of unnecessary rotation (angle between an agent's heading and direction to goal) that an agent exhibits per unit path length [30]; (5) *Path Efficiency*, $\mathscr{E}$, defined as the ratio of the distance between the endpoints of a segment over the length of the path that the agent actually followed; (6) time spent per unit path length over a segment, τ; and (7) TC [19, 53], defined as the amount of entanglement among agents' trajectories throughout a trial (the Braidlab software package [68] was used for these computations).

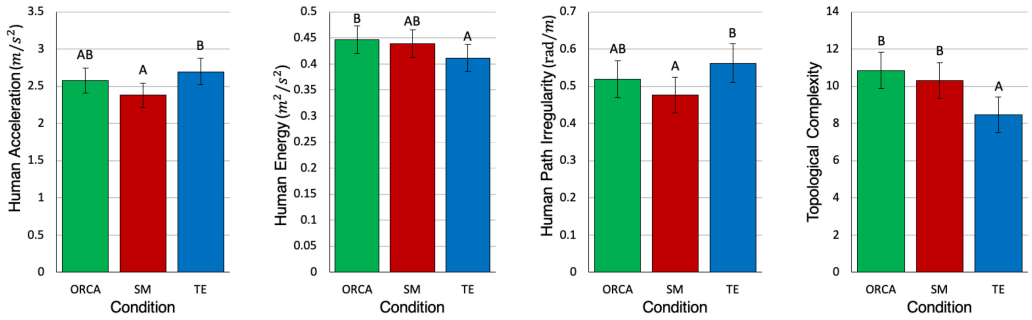
6.5 Effect of Navigation Strategy on Robot Behavior

We model the effect of condition (OR, SM, and TE) on each one of the trajectory quality measures considered. We use linear mixed-effects regression models to account for both fixed effects resulting from the conditions but also for random effects resulting from the session and the trial (expected means with confidence intervals are depicted in Figure 11).

One-way ANOVA performed on the models demonstrates a significant effect of the condition on all robot trajectory quality criteria at the $p < 0.05$ level (see Table 2 for the test statistics and Figure 11 for the expected means and confidence intervals for all criteria), and thus we find that H2 is confirmed. More specifically, it can be observed that OR generates the smoothest motion among

Table 3. Effect of Navigation Strategy on Human Behavior

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
a	1.415	0.7073	2	250.4	3.888	0.02173
d	0.1075	0.05377	2	231.5	0.5872	0.5567
E	0.112	0.05599	2	253.3	3.449	0.03326
$\mathscr{E}$	0.02977	0.01489	2	68.46	1.959	0.1488
PI	0.5394	0.2697	2	249.4	3.286	0.03904
τ	0.08277	0.04139	2	252.7	2.145	0.1192



(a) Acceleration (b) Energy (c) Path Irregularity (d) Top. Complexity

Fig. 12. Expected means and confidence intervals for human acceleration (a), human energy (b), and human path irregularity (c), averaged per trajectory segment. Figure 12(d) depicts the mean and confidence interval for the topological complexity, averaged across trials considering both the humans and the robot. Pairs labeled with different letters (e.g., A and B) are significantly different to each other (Tukey's HSD test, $p < 0.05$).

all strategies (lowest acceleration, lowest path irregularity, lowest time), which confirms H2a. This trend was expected as OR selects actions that minimize divergence from an agent's direction to goal and desired speed to ensure collision avoidance for a desired time window. This results in a smoother speed profile than other conditions. SM, however, prioritizes intent-expressiveness by exaggerating its motion to indicate an intended passing-side intention; this results in higher acceleration (due to rotation) and path irregularity, which confirms H2b. Finally, TE is the most energy efficient—which confirms H2c—but also the least time efficient of all strategies. These findings could mainly be attributed to the defensive driving style of the teleoperator and the navigation through arrow keys.

6.6 Effect of Navigation Strategy on Human Behavior

Similarly to robot trajectory, we model the dependency of the human trajectory quality measures to the condition with linear mixed-effects models, accounting also for random effects of session, trial, and helmet per trial. Figure 12 depicts the expected means and confidence intervals for the human trajectory quality measures, whereas Table 3 contains statistics extracted upon performing ANOVA on the models at the $p < 0.05$ significance level.

Overall, we find that H3 is confirmed. In particular, we see that humans in close proximity with the robot followed smoother trajectories of lower acceleration and path irregularity when exposed to SM than humans exposed to either OR or TE, which confirms H3a. This was in line with our expectations: SM's intention-aware navigation strategy adapts the robot's behavior to the preferences of humans, thus facilitating human inference and decision making. Further, it

Table 4. Effect of Navigation Strategy on Group Behavior

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
TC	107.3	53.66	2	67.71	8.075	0.000716

was observed that humans spend the least energy when exposed to TE, which confirms H3b. We attribute this finding to the perceived humanlike nature of the motion generated by a teleoperated robot: The embodiment of human decision making on a robot platform features humanlike traits that potentially enable a higher level of human comfort. Finally, humans spend the most energy around OR, which confirms H3c. This could be perceived as an result of OR's more predictable motion (minimal divergence from desired direction). Higher predictability potentially results in higher confidence for participants, which allows them to move faster and thus spend more energy.

6.7 Effect of Navigation Strategy on Group Behavior

We compute the TC of the complete motion of all agents (robot and humans) throughout each trial. We model the effect of condition on the TC of the group trajectory (the set of all agents' trajectories) over a trial, using a linear mixed-effects model (accounting for random effects of session, trial, and helmet per trial). Overall, we find that H4 is rejected. ANOVA performed on the model uncovered a significant variance among conditions ($F(2, 67.71) = 8.075, p = 0.000716$, see Table 4, Figure 12). While we see that the TC of SM is lower than the TC of OR, the result is not statistically significant. Further, we see that the TC of trajectories generated by groups exposed to TE was significantly lower than both SM and OR.

While the lab experiment resembles in many ways the experiment featured in the online study, several variables introduce important differences. Importantly, the embodiment on a differentially constrained robot results in robot motion that differs from the motion generated in the simulated experiments. Further, the online study only compared SM against simulation baselines (ORCA and SF). We see that SM still exhibits lower TC on average than OR, albeit not to a statistically significant extent. The relation to a human-operated robot was a new baseline for which we had less clarity, especially under the embodiment settings considered. Finally, this finding could be attributed to the mechanisms underlying human navigation, as the decision-making computations under TE were done by the human teleoperator. Lower *TC* represents trajectory entanglement that intuitively corresponds to behaviors of passing *around* as opposed to passing *through* others (see Figure 5). Thus, this trend could be attributed to the tendency of the human teleoperator to avoid collisions more *globally*, by avoiding any type of encounter with other participants, whereas the robot was employing a more local collision avoidance mechanism by sequentially responding to any challenging encounters. This finding is perhaps unsurprising since both autonomous algorithms considered explicitly favor the avoidance of closer collisions over further ones.

7 HUMAN PERCEPTIONS OF SOCIAL ROBOT NAVIGATION BEHAVIORS

In this section, we investigate how the social robot navigation frameworks considered in our lab study shaped human perceptions. We focus on the dataset of human responses to the questionnaire distributed after each condition. We first analyze the responses to a Likert-scale questionnaire, which focused on participants' impressions about the robot's intelligence, social competence, and emotional impact. We then perform a thematic analysis on participants' responses to an open-form question that extracted qualitative feedback from participants.

Table 5. Ratings of Robot Behavior and Human State

Ratings posed as adjective pairs and sentences		Mean	SD
Incompetent	Competent	3.16	1.02
Irresponsible	Responsible	3.22	1.02
Unpredictable	Predictable	2.64	1.11
Incompliant	Compliant	3.14	0.95
Foolish	Sensible	3.09	1.06
Unfriendly	Friendly	3.13	1.02
Unsafe	Safe	3.27	1.22
Unpleasant	Pleasant	3.09	0.96
Rude	Polite	3.02	0.92
Clumsy	Coordinated	3.26	1.09
Unintelligent	Intelligent	3.12	0.98
Untrustworthy	Trustworthy	3.15	0.87
Socially unaware	Socially aware	2.61	1.15
Indiscreet	Discreet	2.67	1.10
Anxious	Relaxed	3.28	1.13
Agitated	Calm	3.39	1.11
Surprised	Tranquil	3.17	0.99
The robot's presence was not noticeable		2.04	1.11
I will be able to tell where the robot is going in the future		2.25	1.12
The robot will not bump into me in the future		2.97	1.24

7.1 Exploratory Factor Analysis: Effect of Navigation Strategy on Human Ratings

Table 5 contains descriptive statistics of the contributed ratings. The left column lists the adjective pairs and questions shown to participants in a 1–5 Likert-scale format, where 5 corresponds to a positive answer. Ratings from 3 subjects were omitted as they were incomplete, resulting in an effective dataset of 102 human subject ratings for each condition. The right column lists the means and standard deviations of the ratings from all 102 subjects, computed over all three conditions.

To understand the ratings better, we perform an exploratory factor analysis to group them into thematically consistent categories. Factor analysis expresses an observation $X \in \mathbb{R}^p$, where p is the number of observed variables, as a linear relationship,

$$X = \Lambda F + E, \tag{8}$$

where Λ is a matrix of factor loadings, $F \in \mathbb{R}^m$ is a factor space of size $m \leq p$, and E is a vector of specific factors. The factor loadings in Λ represent correlations between observed variables X and latent factors F.

In our case, we had $p = 20$ questions, answered 306 times (3 times from each of the 102 subjects). We verified the sampling adequacy by computing the Kaiser-Meyer-Olkin measure (KMO) for each question independently ($KMO > 0.74$), and for all questions together ($KMO = 0.89$). We verified that the observed variables were not pathologically correlated by examining the correlation matrix R (Bartlett's test, $\chi^2(190) = 2787.774$, $p < 0.001$, $det(R) > 0.00001$). Following Kaiser's criterion, we decided to extract $m = 5$ principal factors corresponding to eigenvalues >1. Based on these 5 factors, we were able to explain 53.3% of the total variance. Table 6 depicts the retrieved factor matrix Λ, rotated by the Varimax method. The factors are ranked in order of decreasing variability explained.

Table 6. Factor Matrix (Varimax Rotated)

Rating	Factor I	Factor V	Factor II	Factor III	Factor IV	Communality
incompetent_competent	0.423	0.481	0.048	0.496	0.127	0.673
irresponsible_responsible	0.496	0.380	0.061	0.479	0.144	0.643
unpredictable_predictable	0.027	0.114	0.162	**0.599**	0.055	0.401
incompliant_compliant	0.413	0.304	0.098	0.413	0.135	0.461
foolish_sensible	0.426	**0.529**	0.016	0.284	0.074	0.547
unfriendly_friendly	**0.698**	0.133	0.253	−0.022	0.096	0.578
unsafe_safe	**0.584**	0.225	0.324	0.179	0.217	0.576
unpleasant_pleasant	**0.664**	0.284	0.260	0.043	0.159	0.615
clumsy_coordinated	0.216	**0.549**	0.125	0.170	0.200	0.432
unintelligent_intelligent	0.167	**0.845**	0.007	0.140	−0.083	0.767
untrustworthy_trustworthy	0.463	0.462	0.167	0.175	0.019	0.487
sociallyUnaware_aware	0.219	**0.586**	0.190	0.178	0.220	0.506
anxious_relaxed	0.229	0.020	**0.771**	0.149	0.093	0.677
agitated_calm	0.287	0.211	**0.789**	0.186	0.111	0.796
predict_where_future	−0.078	0.148	0.190	**0.602**	−0.038	0.427
rude_polite	**0.547**	0.229	0.133	−0.076	0.161	0.401
indiscreet_discreet	0.216	0.022	0.088	0.119	0.464	0.284
surprised_tranquil	0.149	0.068	**0.570**	0.218	0.252	0.462
presence_not_noticeable	0.052	0.114	0.157	−0.051	**0.735**	0.582
wont_bump_future	0.231	0.335	0.271	0.052	0.327	0.348
SS loadings	2.947	2.729	2.060	1.721	1.215	
Proportion Variance	0.147	0.136	0.103	0.086	0.060	
Cumulative Variance	0.147	0.283	0.386	0.472	0.533	
Proportion Explained	0.276	0.255	0.193	0.161	0.113	
Cumulative Proportion	0.276	0.531	0.724	0.886	1.000	

We interpret the factors by focusing on the questions with loadings greater than 0.5 in Table 6. Based on this assumption, we extract the following labels. In particular, we label factor I as *social competence*, since "friendly," "safe," "pleasant," and "polite" scored high; factor V as *intelligence*, since the "intelligent," "coordinated," "socially aware," and "sensible" adjectives scored the highest; factor II as *human comfort*, since the "relaxed," "calm," and "tranquil" adjectives scored highest; factor III as *predictability*, since the "predictable" and "predict where the robot is going" scored highest; and factor IV as *discretion*, since the "presence not noticeable" question scored highest.

Based on the extracted linear model of Equation (8), we can now compute the factor scores for each observation following a standard regression method. We then model the effect of condition on each of the extracted factors using a Linear Mixed Effects Regression Model. Table 7 contains the statistics of one-way ANOVA tests performed for each factor. Overall, we observe no significant variance across conditions for all factors.

7.2 Thematic Analysis: Effect of Navigation Strategy on Short Responses

To gain a deeper understanding of the impressions, attitudes, and views of the participants over the three conditions of robot navigation strategies employed in the lab study (ORCA, SM, TE; see Section 6), participants were invited to answer one open-question after experiencing each condition: *"Please include anything else you would like to share about your experience."* This question was asked to extract users' insights that cannot be captured using traditional Likert-scale questionnaires [2].

Table 7. Effect of Strategy on Factor Scores (ANOVA)

Rating	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Social competence	0.70078	0.35039	2	269.33	0.6025	0.5482
Intelligence	3.4518	1.7259	2	269.32	2.8332	0.06058
Human comfort	1.4287	0.71433	2	269.16	1.0388	0.3553
Predictability	0.083392	0.041696	2	270.57	0.0655	0.9367
Discretion	2.6502	1.3251	2	269.57	2.2132	0.1113

In this section, we detail the thematic analysis conducted, the coding scheme used, the inter-judges agreement, and the results from this qualitative analysis. This thematic analysis helped us understand the participants' experience during the study.

7.2.1 Inter-Coder Agreement. Two independent coders were involved in the coding of the data. First, each coder reads through the transcribed data independently and without attempting to make a correspondence between the data and the conditions. At this stage, the two coders extracted emerging themes from the data. Second, the coders discussed the themes they had identified and established a preliminary coding scheme. With this initial coding scheme, the two coders coded 50% of the data. The inter-coder agreement for this 50% of data analysis ranged from .276 (minimal agreement) to 1 (perfect agreement). Given the need to improve the coding, the coders discussed all the disagreements and coded the remaining of the data. The inter-coder agreement for 100% of the data showed a moderate to strong agreement, $k = .739$ to .807 [55].

7.2.2 Coding Scheme. Qualitative analysis of the data were performed with Nvivo version 12 [6]. From a total poll of 121 participant responses, we analysed 101 (SM = 33; ORCA = 31; TE = 38). To provide focus to this qualitative analysis, we excluded 20 responses that were not related with participants' impression of the robot. The coding scheme used for the thematic analysis emerged from preliminary analysis of the data and is composed of four main themes named "Navigation," "Robot Behavior," "Appearance and Hardware," and "Human Emotions." The coding scheme is visually represented in Figure 13 and detailed in Table 8.

7.2.3 Results. We highlight some of the most relevant findings from our thematic analysis, adding more depth as to how participants experience their interaction with the robot across the three study conditions. Figure 14 depicts a word cloud containing the most frequent words used by participants.

Navigation Intention and Proximity. There were common aspects between all study conditions regarding how participants experienced the robot navigation that are related to the *intention* of the robot and the *proximity* to the robot while navigating in the same space. Often, participants were unclear about the purpose of the robot, *"I had no idea what the robot was doing"* (TE condition) as the behavior of the robot was perceived as unpredictable to them, *"It seemed like the robot would adjust its speed but unpredictably which made me nervous"* (OR condition).

Independently of the study conditions, there was an overall sense that the robot was too close to the participants: *"There were moments when the robot approached me so close that I felt uncomfortable"* (SM condition) or *"I felt the robot was in my personal space"* (OR condition). In the same line, participants also felt the robot was interrupting their trajectories, *"The robot gets in the way occasionally, sometimes I turn around and its right behind me"* (OR condition). This resulted in some collisions between the participants and the autonomous robot across study conditions, *"I saw some near collisions. I think the robot got in my way maybe once or twice"* (SM condition). Interestingly, participants acknowledged the collision experience with the robot in a similar fashion compared

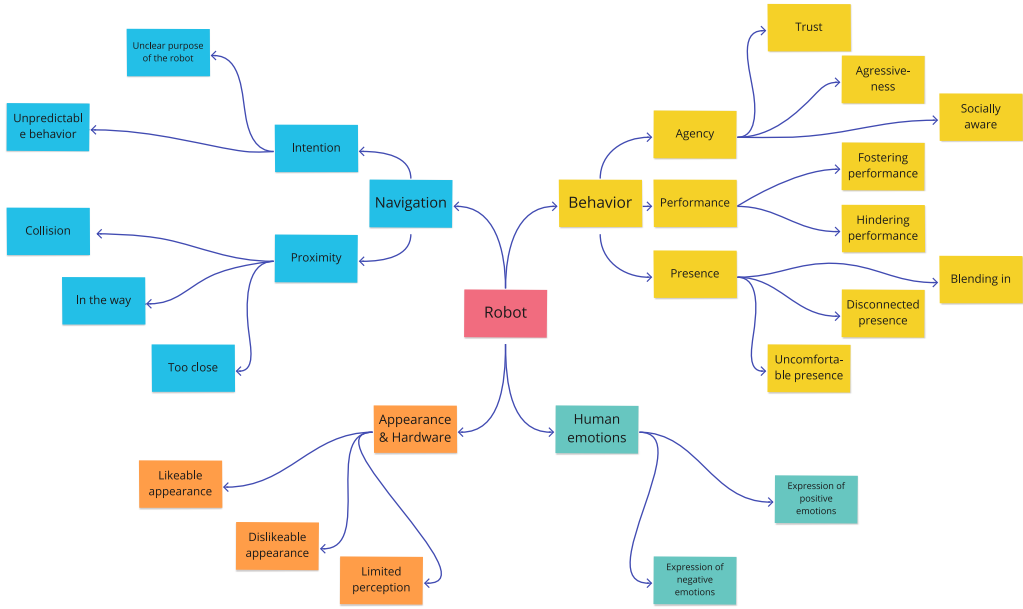


Fig. 13. Coding scheme developed and used in the thematic analysis described in Section 7.2.

Table 8. Coding Scheme Used in the Thematic Analysis

Navigation	**Intention**	***Unclear purpose of the robot***	Confusion about the robot's purpose.
		Unpredictable behavior	Impossibility to anticipate the robot's behavior.
	Proximity	***Collision***	Reference to an actual collision or a near-collision experience.
		Too close	The robot approached too close, yielding discomfort.
		In the way	The robot intersected the path.
Behavior of the robot	**Agency**	***Trust***	The robot's behavior promoted trust in it.
		Aggressiveness	The robot's behavior was perceived as aggressive.
		Socially aware	The robot's navigation motion was perceived to respect the norms of socially acceptable behavior.
	Performance	***Fostering performance***	The robot's behavior fostered the task performance.
		Hindering performance	The robot's behavior hindered the task performance.
	Presence	***Blending in***	The robot's behavior felt natural and fit well within the context.
		Disconnected presence	The robot's behavior did not fit within the context of the task.
		Uncomfortable presence	The robot's presence brought up feeling of discomfort.
Appearance and hardware		***Likeable appearance***	Positive expressions toward the robot's physical appearance.
		Dislikeable appearance	Negative expressions toward the robot's physical appearance.
		Limited perception	The robot is perceived to lack sensing abilities over the environment and the participants.
Human emotions		***Positive emotions***	The interaction with the robot resulted in positive emotions.
		Negative emotions	The interaction with the robot resulted in negative emotions.

to collisions with other participants from the study, *"The robot felt about as in the way as a human colleague during most of the study"* (TE condition).

Although the navigation near the robot and the other participants elicited some near-collisions and invasions of personal space, this may partially be due to the experiment design itself, which attempted to emulate crowded navigation settings. Participants compared the study navigation experience to walking in a crowed public space, *"The robot did bump my shoe but it was very slight and not very different from navigating around a crowd"* (TE condition). Participants acknowledged that the robot was blending in during the study, *"I barely noticed the robot when I was performing the tasks"* (SM condition), comparing the presence of the robot with the other human participants, *"I did not notice the robots movements much more than the movements of other people"* (OR

Fig. 14. Word cloud. This image shows the most frequent words used by the participants to describe their experience interacting with the robot across the three study conditions.

condition). These results capture similarities of the participants' experience across the three study conditions.

Robot Performance and Human Emotions. The factory setting scenario enabled us to avoid telling participants about the true purpose of our study in an effort to elicit more natural navigation behaviors from them. The scenario involved participants performing maintenance tasks in a factory. This led to considerations of performance that appeared frequently in their qualitative responses. It is important to note that while participants referred to their performance, this study did not focused on measuring the performance of the participants but rather on their subjective feelings of working closely with a robot, which we report below.

Some reported the robot hindering their performance, *"The robot's movements relatively came in the way of my work"* (OR condition), *"The robot hindered my performance"* (SM condition). Interestingly, on the TE condition, participants reported feelings of fostered performance related to robot navigation, *"I was more focused on the task this time. It was interesting having it move around and feeling its presence watching/supervising us during the task"* (TE condition). When interacting with the robot in the TE condition, participants also reported a higher social awareness of the robot, *"The robot sensed my presence and stopped."* (TE condition) or, *"I think it was interesting to see the robot stop abruptly when another "worker" walked in front of it. I think sometimes it might be tough to always have to worry about not bumping into the robot but it was nice to see that it at least had some sense to stop"* (TE condition).

Although participants did not report feelings of fostering performance in the remaining conditions, crucially, they associated positive feelings on the interaction with the robot, *"The robot seemed to be checking easels the were not currently occupied and staying away from people. This made the robot seem like it was better at its job and safer because it didn't try to get too close to people. When transitioning the robot seemed to wait for the people to move so it was easy to walk around"* (SM condition). Positive emotions toward the robot arose mostly associated with the notion that the robot is socially aware, *"Sometimes the robot would be blocking my field of view of other machines, but it was pretty aware in the fact that it would stop before hitting into other people"* (SM condition) or that it is not directly interfering with the task, *"I did not feel as if the robot was interfering with anything"* (OR condition). Participants also acknowledged that the novelty of interacting with a robot influenced how they performed during the task, *"The initial presence of the robot was disconcerting, but as I focused on my own task, I stopped noticing it altogether"* (OR condition). As the novelty wears off, participants can focus more on the task, *"I think I got used to the robot's presence by now, and hence it was better. Even though this felt more time constrained, the robot didn't make me as agitated"* (SM condition).

8 DISCUSSION

This article represents a holistic documentation of our research on the design and evaluation of the Social Momentum planning framework. In this section, we offer a unifying summary and discussion of our algorithmic design, empirical methods, and findings.

8.1 Algorithmic Design

Our algorithmic design was inspired by research on the cognitive mechanisms underlying human inference [4, 16, 79] and by studies on the mechanisms governing human navigation and social interactions in public spaces [28, 31, 80]. It was further motivated by recent work in human–robot interaction [17, 40], highlighting the value of implicit, nonverbal communication as an effective tool for fluent and effective human–robot collaboration. Finally, it builds on recent studies showing the value of leveraging the cooperative mechanisms of human navigation into the design of algorithmic social robot navigation frameworks [42, 52, 72].

In contrast to the majority of the literature, which implicitly captures features of interaction in multiagent navigation [14, 37, 42, 72], our framework leverages an *explicit* representation of interaction based on the quantity of angular momentum. Our framework monitors the state of consensus between the navigation strategy of the robot and the inferred strategies of other humans, and contributes robot motion aiming at maximally communicating the robot's intention of complying with them while taking the robot closer to its destination. This is implemented via a computationally efficient decision-making rule based on the objective of *Social Momentum*, which offers model interpretability and compactness. Finally, our framework is unique in that it represents one of the first efforts of generating legible motion in the presence of *groups* of multiple human agents. Legibility is a property of robot motion that has been typically explored in single-human/single-robot interaction settings in structured domains [12, 17, 43] or in simulated scenarios in multiagent navigation [48, 50, 52].

Our algorithmic design is deliberately simplified. One could possibly achieve higher performance by considering an MPC-like design involving, e.g., a larger planning horizon, adding components to the objective function, carefully crafting the action space or leveraging a data-driven human motion prediction model. However, our intention was to produce a planning *framework* rather than a specifically engineered solution. Despite its simplicity, our framework was capable of handling the complex interactions arising in our lab experiments. This was achieved with minimal tuning. We find this observation interesting on its own—one would expect that such a simplified

design cannot handle humans. But we did show that from the perspective of participants, SM did just as well in terms of self-reported human perceptions and allowed participants to navigate at more constant speed levels (H3a). We hope that the community can easily build on top of our framework or extract concepts and ideas such as employing topological models of multiagent interaction.

8.2 Empirical Methods

Our work featured a comprehensive evaluation including a characterization of simulated algorithm performance and a collection of human perceptions to simulated and embodied robot motion.

Critically, while there has been much work in the space of social robot navigation in multiagent human environments [14, 37, 42, 60, 72], our lab study is unique in that it—to the best of our knowledge—is the first to combine the following important elements: (1) controlled experimental settings, (2) crowded navigation settings, (3) challenging navigation encounters requiring nontrivial collision-avoidance maneuvers, (4) natural walking-behaviors settings through the use of a scripted scenario, and (5) large-scale data collection (105 subjects; 315 minutes of interaction). These settings enabled us to enforce a situation resembling real-world interactions in a controlled experimental fashion (Section 6.1). We hope that our lab study design could serve as a template for designing a benchmark for evaluating social robot navigation algorithms.

Furthermore, our work establishes the value of topological tools and analysis for multiagent navigation. First, the abstraction of angular momentum encodes a topological structure to the decision-making process of the robot. Further, we proposed the use of *Topological Complexity* [19] as a tool for the analysis of multiagent navigation behaviors. Our online user study showed evidence that topological complexity affects human perceptions of multiagent interactions. In particular, we established a correlation between high *Topological Complexity* and low *Legibility* of multiagent interactions. We expect that *Topological Complexity* but also additional tools from topological data analysis [20] may also prove very relevant to applications in multiagent navigation.

8.3 Insights on Legibility

Social Momentum was designed to generate robot motion intended to be perceived as *legible* by humans. We showed that in multiagent navigation, high *Legibility* is correlated with low *Topological Complexity*, and we established that *Social Momentum* appears to be more legible compared to baselines such as ORCA [76] and Social Force [32] according to the findings of our online user study. In the lab study, this clear result does not directly transfer; group behaviors (including the robot and humans) under the SM condition exhibit lower average topological complexity than ORCA but the trend is not significant (see Figure 12(d)). However, we observe that under *TE*, group behaviors are significantly less topologically complex. Intuitively, low topological complexity corresponds to multiagent-navigation behaviors featuring low *mixing* in the sense described in Figure 5. Thus, it appears that in the presence of a human-teleoperated robot, the group of all agents engaged in collision-avoidance encounters that were less direct, more coordinated, and—under the correlation between high *Legibility* and low *Topological Complexity*—also more *legible*. In other words, the human-teleoperated robot exhibited behaviors that were more easily interpretable by humans (i.e., legible) than the behaviors generated by either ORCA or SM.

We interpret this observation as an artifact in part of the uniquely collaborative character of human navigation [80]: The human operator (the operator was already experienced with the teleoperation setup and was held constant across all trials) would anticipate human reactions to the robot motion and visibly diverge from a straight line path to convey collision avoidance avoid others. This pattern is also reflected in the higher time per path segment that the robot followed under TE (see Figure 11(d)).

8.4 Insights on Comfort

It is generally challenging to measure the actual levels of comfort of participants throughout the lab study. The novelty effect, the close interaction settings and the traits of the algorithms considered have significantly affected human impressions as it can be seen in the Likert-scale questionnaire analysis (Section 7.1) but also in the thematic analysis of participants' short responses (Section 7.2). However, looking at the trajectory dataset, we observe that human subjects navigating in close proximity to the robot under SM follow low-acceleration paths (significantly lower than under TE; lower than ORCA but not significantly so). This implies that humans switched between different speeds less times when navigating next to the robot running SM, compared to the baselines. The lower frequency of speed switching seems to indicate that humans avoided collisions with the robot more easily under SM, which we interpret as a proxy for higher comfort. While this is not confirmed by human responses to the questionnaires, we believe that this might be because the difference is not perceptible by human subjects.

8.5 Insights on Subjects' Impressions

The analysis of human impressions that were collected via Likert-scale questionnaires did not reveal significant differences across conditions. The thematic analysis on participants' short responses to open-form questions shared this trend but did provide a more nuanced, in-depth insight into participants' thoughts and emotions during close interaction with the robot.

Generally, participants described the feeling that the robot was approaching them too closely in all study conditions. They also recognized that this level of proximity was similar to how they felt with other participants involved in the study, and in they way they feel in crowded spaces. This seems to show that our intention to mimic a crowded space in which a robot and humans need to navigate close to each other was successfully accomplished in the study design. Additionally, it showed that the level of proximity the participants felt between them and with the robot was equivalent. This seems to show that although they would have enjoyed an additional level of comfort by having more space between them and the robot, they understood this closeness was part of the task itself and could be compared to real-world scenarios of crowded environments.

Participants also reported positive emotions from the interaction with the robot, which emerged mostly associated with perceiving a higher social awareness of the robot toward them. It is interesting that this was achieved despite the very limited fashion in which the robot perceived the world around it: just human poses without any other visual/force sensing. Thus, it appears that a robot with solely kinematic sensing may be perceived as being capable of generating socially adequate behavior in crowded navigation settings.

8.6 Future Directions

One of our goals is to move our robot outside of the lab. We conduct a field study to investigate if the findings of the lab study transfer to real-world scenarios where humans and robots navigate in crowded environments. Environments such as airports, shopping malls, and crowded streets are of potential interest. This would require several systems upgrades including on-board perception mechanisms and further tuning to adapt to the different context.

Crucially, while conducting the research discussed in this article, it became apparent that the field of social robot navigation lacks solid evaluation standards. To the best of our knowledge there are no validated scales or measures specific to robot navigation in close proximity to humans. This makes it hard to evaluate how well algorithms perform and how performance varies across environments and interaction settings. Given the rich interest in this area over the recent years, we feel that the development of a social robot navigation scale is a crucial missing element and an

important line of research. Relatedly, we observe a lack of benchmarks and concrete performance metrics that would enable systematic testing. The introduction of such tools in conjunction with recent efforts on the design of simulation environments [26, 74] will enable standardized benchmarking and a smoother transition from simulation to real-world deployment.

ACKNOWLEDGMENTS

We thank Alena Hutchinson and John Macdonald for their assistance in conducting our lab study. Finally, we thank Erika Louise Mudrak from the Statistical Consulting Unit at Cornell University for valuable discussions regarding the statistical analyses presented in this article.

REFERENCES

[1] [n. d.]. Suitable Technologies, Inc. Retrieved September 13, 2018 from https://suitabletech.com/.
[2] Lora D. Allison, Morris A. Okun, and Kathy S. Dutridge. 2002. Assessing volunteer motives: A comparison of an open-ended probe and Likert rating scales. *J. Commun. Appl. Soc. Psychol.* 12, 4 (2002), 243–255.
[3] E. Artin. 1947. Theory of Braids. *Ann. Math.* 48, 1 (1947), 101–126.
[4] C. L. Baker, J. Tenenbaum, and R. Saxe. 2009. Action understanding as inverse planning. *Cognition* 113, 3 (2009), 329–349.
[5] Christoph Bartneck, Elizabeth Croft, and Dana Kulic. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *Int. J. Soc. Robot.* 1 (2009), 71–81.
[6] Patricia Bazeley and Kristi Jackson. 2013. *Qualitative Data Analysis with NVivo.* SAGE.
[7] N. Bellotto and H. Hu. 2010. A bank of unscented kalman filters for multimodal human perception with mobile service robots. 2, 2 (2010), 121–136.
[8] Maren Bennewitz, Wolfram Burgard, Grzegorz Cielniak, and Sebastian Thrun. 2005. Learning motion patterns of people for compliant robot motion. *Int. J. Robot. Res.* 24, 1 (2005), 31–48.
[9] Joan S. Birman. 1975. *Braids Links And Mapping Class Groups.* Princeton University Press.
[10] Wolfram Burgard, Armin B. Cremers, Dieter Fox, Dirk Hähnel, Gerhard Lakemeyer, Dirk Schulz, Walter Steiner, and Sebastian Thrun. 1998. The interactive museum tour-guide robot. In *Proceedings of the 15th National/Tenth Conference on Artificial Intelligence/Innovative Applications of Artificial Intelligence (AAAI'98/IAAI'98).* American Association for Artificial Intelligence, 11–18.
[11] Baptiste Busch, Jonathan Grizou, Manuel Lopes, and Freek Stulp. 2017. Learning legible motion from human–robot interactions. *Int. J. Soc. Robot.* (2017).
[12] Daniel Carton, Wiktor Olszowy, and Dirk Wollherr. 2016. Measuring the effectiveness of readability for mobile robot locomotion. *Int. J. Soc. Robot.* 8, 5 (2016), 721–741.
[13] Yuhang Che, Allison M. Okamura, and Dorsa Sadigh. 2020. Efficient and trustworthy social navigation via explicit and implicit robot-human communication. *IEEE Trans. Robot.* 36, 3 (2020), 692–707.
[14] Y. F. Chen, M. Everett, M. Liu, and J. P. How. 2017. Socially aware motion planning with deep reinforcement learning. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'17).* 1343–1350.
[15] Gregory F. Cooper. 1990. The computational complexity of probabilistic inference using bayesian belief networks. *Artif. Intell.* 42, 2 (1990), 393–405.
[16] G. Csibra and G. Gergely. 2007. 'Obsessed with goals': Functions and mechanisms of teleological interpretation of actions in humans. *Acta Psychol.* 124, 1 (2007), 60–78.
[17] Anca D. Dragan and Siddhartha Srinivasa. 2014. Integrating human observer inferences into robot motion planning. *Auton. Robot.* 37, 4 (2014), 351–368.
[18] N. E. Du Toit and J. W. Burdick. 2012. Robot motion planning in dynamic, uncertain environments. *IEEE Trans. Robot.* 28, 1 (2012), 101–115.
[19] Ivan Dynnikov and Bert Wiest. 2007. On the complexity of braids. *J. Eur. Math. Soc.* 009, 4 (2007), 801–840.
[20] H. Edelsbrunner and J. Harer. 2008. Persistent homology—A survey. In *Surveys on Discrete and Computational Teometry*, Vol. 453. Amer Mathematical Society, 257.
[21] M. Everett, Y. F. Chen, and J. P. How. 2018. Motion planning among dynamic, decision-making agents with deep reinforcement learning. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'18).* 3052–3059.
[22] Franck Feurtey. 2000. *Simulating the Collision Avoidance Behavior of pedestrians.* Master's thesis. University of Tokyo, Tokyo, Japan.
[23] Amalia F. Foka and Panos E. Trahanias. 2010. Probabilistic autonomous robot navigation in dynamic environments with human motion prediction. *Int. J. Soc. Robot.* 2, 1 (2010), 79–94.

[24] Dieter Fox, Wolfram Burgard, and Sebastian Thrun. 1998. A hybrid collision avoidance method for mobile robots. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'98)*. vol. 2. 1238–1243.
[25] A. Sanfeliu G. Ferrer. 2019. Anticipative kinodynamic planning: Multi-objective robot navigation in urban and dynamic environments. *Auton. Robot.* 43, 6 (2019), 1473–1488.
[26] Christian Gloor. 2016. Pedsim: Pedestrian Crowd Simulation. Retrieved from http://pedsim.silmaril.org.
[27] Julio Godoy, Ioannis Karamouzas, Stephen J. Guy, and Maria Gini. 2016. Moving in a crowd: Safe and efficient navigation among heterogeneous agents. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI'16)*, 294–300.
[28] Erving Goffman. 1966. *Behavior in Public Places: Notes on the Social Organization of Gatherings*. Free Press.
[29] E. Goffman. 2009. *Relations in Public*. Penguin.
[30] J. Guzzi, A. Giusti, L. M. Gambardella, G. Theraulaz, and G. A. Di Caro. 2013. Human-friendly robot navigation in dynamic environments. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'13)*. 423–430.
[31] E. T. Hall. 1990. *The Hidden Dimension*. Anchor Books.
[32] Dirk Helbing and Péter Molnár. 1995. Social force model for pedestrian dynamics. *Phys. Rev. E* 51 (1995), 4282–4286. Issue 5.
[33] Peter Henry, Christian Vollmer, Brian Ferris, and Dieter Fox. 2010. Learning to navigate through crowded environments. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'10)*. 981–986.
[34] Joshua Mason Joseph, Finale Doshi-Velez, Albert S. Huang, and Nicholas Roy. 2011. A Bayesian nonparametric approach to modeling motion patterns. *Auton. Robot.* 31 (2011), 383–400.
[35] David A. Karp, Gregory Prentice Stone, and William C. Yoels. 1977. *Being Urban: A Social Psychological View of City Life*. Heath.
[36] Yusuke Kato, Takayuki Kanda, and Hiroshi Ishiguro. 2015. May I help you?: Design of human-like polite approaching behavior. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI'15)*. 35–42.
[37] Beomjoon Kim and Joelle Pineau. 2016. Socially adaptive path planning in human environments using inverse reinforcement learning. *Int. J. Soc. Robot.* 8, 1 (2016), 51–66.
[38] Rachel Kirby. 2010. *Social Robot Navigation*. Ph. D. Dissertation. Carnegie Mellon University, Pittsburgh, PA.
[39] R. Kirby, R. Simmons, and J. Forlizzi. 2009. COMPANION: A constraint-optimizing method for person-acceptable navigation. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'09)*. 607–612.
[40] Ross A. Knepper, Christoforos I. Mavrogiannis, Julia Proft, and Claire Liang. 2017. Implicit communication in a joint action. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI'17)*. 283–292.
[41] Ross A. Knepper and Daniela Rus. 2012. Pedestrian-Inspired sampling-based multi-robot collision avoidance. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'12)*. 94–100.
[42] Henrik Kretzschmar, Markus Spies, Christoph Sprunk, and Wolfram Burgard. 2016. Socially compliant mobile robot navigation via inverse reinforcement learning. *Int. J. Robot. Res.* 35, 11 (2016), 1289–1307.
[43] Thibault Kruse, Patrizia Basili, Stefan Glasauer, and Alexandra Kirsch. 2012. Legible robot navigation in the proximity of moving humans. In *Proceedings of the IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO'12)*. 83–88.
[44] Thibault Kruse, Amit Kumar Pandey, Rachid Alami, and Alexandra Kirsch. 2013. Human-aware robot navigation: A survey. *Robot. Auton. Syst.* 61, 12 (2013), 1726–1743.
[45] S. Lo, K. Yamane, and K. Sugiyama. 2019. 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'19)*. 1243–1250.
[46] M. Luber, L. Spinello, J. Silva, and K. O. Arras. 2012. Socially-aware robot navigation: A learning approach. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'12)*. 902–907.
[47] Christoforos Mavrogiannis, Francesca Baldini, Allan Wang, Dapeng Zhao, Pete Trautman, Aaron Steinfeld, and Jean Oh. 2021. Core challenges of social robot navigation: A survey. arXiv:2103.05668 [cs.RO]. Retrieved from https://arxiv.org/abs/2103.05668.
[48] Christoforos Mavrogiannis and Ross A. Knepper. 2020. Decentralized multi-agent navigation planning with braids. In *Algorithmic Foundations of Robotics XII: Proceedings of the 12th Workshop on the Algorithmic Foundations of Robotics*, Ken Goldberg, Pieter Abbeel, Kostas Bekris, and Lauren Miller (Eds.). Springer International Publishing, Cham, 880–895.
[49] Christoforos Mavrogiannis and Ross A. Knepper. 2020. Multi-agent trajectory prediction and generation with topological invariants enforced by hamiltonian dynamics. In *Algorithmic Foundations of Robotics XIII*, Marco Morales, Lydia Tapia, Gildardo Sánchez-Ante, and Seth Hutchinson (Eds.). Springer International Publishing, Cham, 744–761.
[50] Christoforos I. Mavrogiannis, Valts Blukis, and Ross A. Knepper. 2017. Socially competent navigation planning by deep learning of multi-agent path topologies. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'17)*. 6817–6824.

[51] Christoforos I. Mavrogiannis, Alena M. Hutchinson, John Macdonald, Patrícia Alves-Oliveira, and Ross A. Knepper. 2019. Effects of distinct robotic navigation strategies on human behavior in a crowded environment. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI'19)*. 421–430.

[52] Christoforos I. Mavrogiannis and Ross A. Knepper. 2019. Multi-Agent path topology in support of socially competent navigation planning. *Int. J. Robot. Res.* 38, 2–3 (2019), 338–356.

[53] Christoforos I. Mavrogiannis, Wil B. Thomason, and Ross A. Knepper. 2018. 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'18)*. 361–369.

[54] Bruce A. Maxwell. 2007. Building robot systems to interact with people in real environments. *Auton. Robot.* 22, 4 (May 2007), 353–367.

[55] Mary L McHugh. 2012. Interrater reliability: The kappa statistic. *Biochem. Med.* 22, 3 (2012), 276–282.

[56] F. Michaud, C. Côté, D. Létourneau, Y. Brosseau, J. M. Valin, É. Beaudry, C. Raïevsky, A. Ponchon, P. Moisan, P. Lepage, Y. Morin, F. Gagnon, P. Giguère, M. A. Roux, S. Caron, P. Frenette, and F. Kabanza. 2007. Spartacus attending the 2005 AAAI conference. *Autonomous Robots* 22 (2007), 369–383.

[57] M. Montemerlo, N. Roy, and S. Thrun. 2003. Perspectives on standardization in mobile robot programming: The Carnegie Mellon Navigation (CARMEN) Toolkit. In *Proceedings IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'03)*, Vol. 3. 2436–2441 vol.3.

[58] Mehdi Moussaïd, Dirk Helbing, and Guy Theraulaz. 2011. How simple rules determine pedestrian behavior and crowd disasters. *Proc. Natl. Acad. Sci. U.S.A.* 108, 17 (2011), 6884–6888.

[59] E. Pacchierotti, H. I. Christensen, and P. Jensfelt. 2006. Evaluation of passing distance for social robots. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'06)*. 315–320.

[60] Jong Jin Park, Collin Johnson, and Benjamin Kuipers. 2012. Robot navigation with Model Predictive Equilibrium Point Control. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'12)*. 4945–4952.

[61] Jorge Rios-Martinez, Anne Spalanzani, and Christian Laugier. 2015. From proxemics theory to socially-aware navigation: A survey. *Int. J. Soc. Robot.* (2015).

[62] Junha Roh, Christoforos Mavrogiannis, Rishabh Madan, Dieter Fox, and Siddhartha S. Srinivasa. 2020. Multimodal trajectory prediction via topological invariance for navigation at uncontrolled intersections. In *Proceedings of the Conference on Robot Learning*.

[63] Andrey Rudenko, Luigi Palmieri, Michael Herman, Kris M. Kitani, Dariu M. Gavrila, and Kai O. Arras. 2020. Human motion trajectory prediction: A survey. *Int. J. Robot. Res.* 39, 8 (2020), 895–935.

[64] Dorsa Sadigh, Shankar Sastry, Sanjit A. Seshia, and Anca D. Dragan. 2016. Planning for autonomous cars that leverages effects on human actions. In *Proceedings of the Annual Conference on Robotics: Science and Systems (RSS'16)*. 239–248.

[65] Masahiro Shiomi, Francesco Zanlungo, Kotaro Hayashi, and Takayuki Kanda. 2014. Towards a socially acceptable collision avoidance for a mobile robot navigating among pedestrians using a pedestrian model. *Int. J. Soc. Robot.* 6, 3 (2014), 443–455.

[66] Emrah Akin Sisbot, Luis Felipe Marin-Urias, Rachid Alami, and Thierry Siméon. 2007. A human aware mobile robot motion planner. *IEEE Trans. Robot.* 23, 5 (2007), 874–883.

[67] Jean-Luc Thiffeault. 2010. Braids of entangled particle trajectories. *Chaos* 20, 1 (2010), 017516.

[68] Jean-Luc Thiffeault and Marko Budišić. 2013–2019. Braidlab: A software package for braids and loops. arXiv:1410.0849 [math.GT].

[69] Andrea Thomaz, Guy Hoffman, and Maya Cakmak. 2016. Computational human-robot interaction. *Found. Trends Robot.* 4, 2-3 (2016), 105–223.

[70] S. Thompson, T. Horiuchi, and S. Kagami. 2009. A probabilistic model of human motion and navigation intent for mobile robot path planning. In *Proceedings of the International Conference on Autonomous Robots and Agents*. 663–668.

[71] S. Thrun, M. Beetz, M. Bennewitz, W. Burgard, A. B. Cremers, F. Dellaert, D. Fox, D. Hähnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz. 2000. Probabilistic algorithms and the interactive museum tour-guide robot minerva. *Int. J. Robot. Res.* 19, 11 (2000), 972–999.

[72] Peter Trautman, Jeremy Ma, Richard M. Murray, and Andreas Krause. 2015. Robot navigation in dense human crowds: Statistical models and experimental studies of human-robot cooperation. *Int. J. Robot. Res.* 34, 3 (2015), 335–356.

[73] X. T. Truong and T. D. Ngo. 2017. Toward socially aware robot navigation in dynamic and crowded environments: A proactive social motion model. *IEEE Trans. Autom. Sci. Eng.* 14, 4 (2017), 1743–1760.

[74] Nathan Tsoi, Mohamed Hussein, Jeacy Espinoza, Xavier Ruiz, and Marynel Vázquez. 2020. SEAN: Social environment for autonomous navigation. arXiv:2009.04300 [cs.RO]. Retrieved from https://arxiv.org/abs/2009.04300.

[75] Vaibhav V. Unhelkar, Claudia Pérez-D'Arpino, Leia Stirling, and Julie A. Shah. 2015. Human-robot co-navigation using anticipatory indicators of human walking motion. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'15)*. 6183–6190.

[76] Jur van den Berg, Stephen J. Guy, Ming C. Lin, and Dinesh Manocha. 2009. Reciprocal *n*-Body collision avoidance. In *Proceedings of the International Symposium on Robotics Research (ISRR'09)*. 3–19.

[77] D. Vasquez, B. Okal, and K. O. Arras. 2014. Inverse Reinforcement Learning algorithms and features for robot navigation in crowds: An experimental comparison. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'14)*. 1341–1346.

[78] William H. Warren. 2006. The dynamics of perception and action. *Psychol. Rev.* 113, 2 (2006), 358–389.

[79] Eva Wiese, Agnieszka Wykowska, Jan Zwickel, and Hermann J. Müller. 2012. I see what you mean: How attentional selection is shaped by ascribing intentions to others. *PLoS One* 7, 9 (2012), 1–7.

[80] Nicholas H. Wolfinger. 1995. Passing moments: Some social dynamics of pedestrian interaction. *J. Contemp. Ethnogr.* 24, 3 (1995), 323–340.

[81] Brian D. Ziebart, Nathan Ratliff, Garratt Gallagher, Christoph Mertz, Kevin Peterson, J. Andrew Bagnell, Martial Hebert, Anind K. Dey, and Siddhartha Srinivasa. 2009. Planning-based prediction for pedestrians. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'09)*. 3931–3936.

Received September 2020; revised August 2021; accepted September 2021