Conflict Avoidance in Social Navigation—a Survey

REUTH MIRSKY, Computer Science Department, Bar-Ilan University, Ramat Gan, Israel
XUESU XIAO, Computer Science Department, George Mason University, Fairfax, USA
JUSTIN HART, Computer Science Department, The University of Texas at Austin, Austin, USA
PETER STONE, Computer Science Department, The University of Texas at Austin, Austin, USA

A major goal in robotics is to enable intelligent mobile robots to operate smoothly in shared human-robot environments. One of the most fundamental capabilities in service of this goal is competent navigation in this "social" context. As a result, there has been a recent surge of research on social navigation; and especially as it relates to the handling of conflicts between agents during social navigation. These developments introduce a variety of models and algorithms, however as this research area is inherently interdisciplinary, many of the relevant papers are not comparable and there is no shared standard vocabulary. This survey aims at bridging this gap by introducing such a common language, using it to survey existing work, and highlighting open problems. It starts by defining the boundaries of this survey to a limited, yet highly common type of social navigation—conflict avoidance. Within this proposed scope, this survey introduces a detailed taxonomy of the conflict avoidance components. This survey then maps existing work into this taxonomy, while discussing papers using its framing. Finally, this article proposes some future research directions and open problems that are currently on the frontier of social navigation to aid ongoing and future research.

CCS Concepts: • **Human-centered computing** → **Interaction design theory, concepts and paradigms**; *HCI theory, concepts and models*; • **General and reference** → **Surveys and overviews**; • **Computing methodologies** → **Mobile agents**; **Cooperation and coordination**; **Multi-agent systems;**

Additional Key Words and Phrases: Social navigation, mobile robots, human-robot interactions

ACM Reference Format:
Reuth Mirsky, Xuesu Xiao, Justin Hart, and Peter Stone. 2024. Conflict Avoidance in Social Navigation—a Survey. *ACM Trans. Hum.-Robot Interact.* 13, 1, Article 13 (March 2024), 36 pages. https://doi.org/10.1145/3647983

1 INTRODUCTION

Enabling autonomous robots to navigate in the presence of people and/or other robots has been studied for the past 70 years. One of the first examples of social navigation is Grey Walter's work, who built robotic "turtles" that could navigate on their own [169]. These robots, named Elmer and Elsie, were an exercise in minimalism and demonstrated that a small number of brain cells could give rise to complex behaviors. They each consisted of "two miniature radio tubes, two sense organs, one for light and the other for touch, and two effectors or motors, one for crawling and the other for steering". Their power supply was a hearing aid battery. Nevertheless, these robots

Authors' addresses: R. Mirsky, Computer Science Department, Bar-Ilan University, Ramat Gan 5290002, Israel; e-mail: reuthde@gmail.com; X. Xiao, Computer Science Department, George Mason University, Fairfax, Virginia 22030, USA; e-mail: xiao@gmu.edu; J. Hart and P. Stone, Computer Science Department, The University of Texas at Austin, Austin, Texas 78712, USA; e-mails: justinhart@gmail.com, stonep@cs.utexas.edu.

ACM 2573-9522/2024/03-ART13
https://doi.org/10.1145/3647983

could navigate freely in an enclosed space and change their trajectory in response to light and touch.

Modern mobile robots are much more sophisticated and complex. Most feature a variety of sensors, intricate steering systems, and several layers of hardware and software to control their movement. Despite these improvements, mobile robots are still not prevalent in our homes and offices. One of the main reasons for this deficit is that comprehensive autonomy is still achievable only in controlled environments and is usually induced by hard-coded rules or learned from a relatively clean dataset [18, 66, 142]. The problem of navigation in the presence of other robots and humans is complex and cross-disciplinary in nature. Solutions draw from robotics, artificial intelligence, engineering, psychology, biology, and other areas of study. As such, each of these communities has defined social navigation differently. In the multi-robot community [167]^{R1}, social navigation usually refers to robot navigation in the presence of additional robots. In **human-robot interaction** (**HRI**), social navigation refers strictly to the task of navigating in a shared space with people. Rios-Martinez et al. [132] gave a compact description of socially-aware navigation: *Socially-aware navigation is the strategy exhibited by a social robot that identifies and follows social conventions (in terms of management of space) to preserve a comfortable interaction with humans. The resulting behavior is predictable, adaptable, and easily understood by humans. This definition implies, from the robot's point of view, that humans are no longer perceived only as dynamic obstacles but also as social entities.*

In the general social navigation setting, a social agent is an agent (either human or robot) that is aware of the objectives of others (human or robot) and considers them in its behavior, either by adjusting its policy or by indicating why it chose a potentially "anti-social" behavior. This general definition is quite broad, encompassing a wide variety of multi-agent navigation scenarios, including those that involve only robots. In practice, the term "social navigation" usually refers to a more human-centric perspective. Thus, this survey focuses on three requirements, beyond the collision avoidance itself, that separate human-centric social navigation from more general social navigation. These requirements are:

(1) There exists an autonomously navigating agent. The agent has a specific, reachable navigational goal.
(2) There exists one (or more) humans or animals in the environment.
(3) The interaction takes place in the real world (either a controlled or natural environment), not in simulation.

Many papers have discussed challenges that occur when only one or two of these requirements are met. Teleoperation of robots is widely investigated within HRI, but it is not consistent with (1). The **multi-agent systems** (**MAS**) and distributed planning communities focus on constructing algorithms for multi-robot navigation, which do not meet requirement (2). Even within the HRI community, many works describe progress in social navigation in simulations rather than in real-world environments, so requirement (3) does not hold. Significant work has been done in the graphics community to model crowds and swarms, but these works also do not meet requirement (3). Our main focus is on papers that meet all three requirements. This survey also cites some papers for which not all of the above requirements hold, due to their contributions to our understanding of social navigation. In cases where the underlying scope of a paper is not fully aligned with this survey, we indicate the requirements that do hold on the first occasion that the paper is referenced. For example: Walter [169]^{R1,R3} is a work in which there is an autonomous agent (R1)—a mechanical "turtle" that navigates in the wild (R3)—but no human pedestrians are present (R2).

Even within the context of the three requirements discussed above, many behaviors could be considered "social": following, giving navigational instructions, waiting in line, and others; as discussed later in this section. To limit the scope of this survey, we focus on one specific type of social interaction with people which requires the robot to reason about an encounter, specifically *conflicts*. A conflict is a short-term interaction between a robot and a human in which there is a chance that the robot and the human will collide. Note that this potential event can be objective, meaning that if no party changes its course they will collide; or it could be that the passing of the robot is perceived as being on a collision course by the human. Additionally, not all interactions in social navigation are conflict avoidance. For example, when a robot is designed to carry a person's luggage and follow them, the task is a social navigation task in which the robot needs to detect the person, reason about the proper distance from them, and drive at a safe and comfortable speed. These challenges, however, are orthogonal to the challenge of avoiding conflicts with other pedestrians. Understanding conflicts in social navigation requires a definition of what a conflict is in this context:

Definition 1. A **conflict** between a robot and other mobile robots or pedestrians is a situation in which if there is no change of direction or a change in speed by at least one of the parties, they will collide.

By this definition, not all conflicts end in a physical collision, but every collision is preceded by a conflict. Moreover, as the interacting parties can falsely predict an upcoming collision (e.g., a human feels that the robot will come too close and is risking a collision), the presence of a conflict is a subjective matter that depends on the interpretation of the interacting parties. This survey is not the first to identify navigational conflicts as being separate from collisions. A footnote from Van Den Berg et al. [164] implies a difference between reasoning about conflicts in motion planning and avoiding collisions:

> Note that the problem of (local) collision avoidance differs from motion planning, where the global environment of the robot is considered to be known and a complete path towards a goal configuration is planned at once, and collision detection, which simply determines if two geometric objects intersect or not.

However, in Van Den Berg et al. [164] they do not elaborate on this idea. Based on this scope, the contributions of this survey are as follows:

(1) It surveys work in which the authors include conflict avoidance in their models.
(2) It introduces a taxonomy of the attributes that vary between models and algorithms for conflict avoidance.
(3) Based on this taxonomy, it identifies the attributes of existing works and categorizes these works into tables.
(4) It summarizes the current state-of-the-art in conflict avoidance in social navigation, including a practical checklist to follow when introducing a new contribution to the body of literature.

Previous works have presented ideas that overlap those in this survey but from different perspectives. There are surveys on topics relating to social robotics [41]^{R1,R2}; and to numerous related navigation topics such as path planning [14]^{R1}, vision for navigation [10, 31]^{R1}, perception and semantics [45] and localization and mapping [28, 43, 139]^{R1,R3}. Many surveys on social navigation focus on elements such: as joint or group navigation [72, 104, 127, 174], giving navigational instructions [156, 175]^{R1}, detecting dynamic objects [35, 76]^{R1}, social contexts such as waiting in line [109]^{R1,R2} or distributing flyers [138]^{R1,R2}, and other factors which are not discussed in this survey [125]. None of these surveys, however, focus specifically on assisting in detecting or avoiding conflicts. Here we provide details on the major related surveys, both to provide a reference for readers who are interested in those different points of view and to define the scope of this survey.

Kruse et al. [83]^{R1,R2} highlight a rising interest in the topic of social navigation since 2000, and identify specific tasks and challenges that social navigation encompasses. Interest is still on the rise, meaning that there are many new works on this topic; requiring this survey to narrow its focus somewhat as we update their coverage of the topic. Our focus is on the narrower topic of conflicts that arise between robots and pedestrians. Hoogendoorn and Bovy [61]^{R2,R3} introduced a three-tiered model of navigation utility, decomposing it into strategic (high-level decision making), tactical (global navigation), and operational (local navigation and event handling) levels. This survey focuses mostly on the operational level: setting local goals and re-planning as needed. Recently, Gao and Huang [44] provided a review of scenarios, datasets, and methods used in social navigation. They described the main use cases as: passing, crossing, overtaking, approaching, following, leading, accompanying, and combinations thereof. Our survey's perspective is different in that it does not categorize papers according to the aim of the navigating parties, but rather according to situations in which these parties are (or will be) in conflict. In this sense, Gao and Huang [44] review a wider set of social navigation tasks, though they do not propose a taxonomy of conflicts as introduced in this survey.

Charalampous et al. [22] present a survey in which they aim *"to systemize the recent literature by describing the required levels of robot perception, focusing on methods related to a robot's social awareness, the availability of datasets these methods can be compared with, as well as issues that remain open and need to be confronted when robots operate in close proximity with humans."* This survey extends their initial discussion on robot design for operation in close proximity to humans; or as we refer to it, robots in conflict situations. Specifically, we aim at providing basic definitions to be used to standardize future works on the problem of robots that navigate in close proximity to people. López et al. [88]^{R1,R3} provide a survey on turn prediction and how upper body kinematics can signal upcoming turns. In their survey, they identified that Gaze Yaw is the earliest predictor of walking turns; but that existing data do not support quantifying how much—or how reliably—timing and distance can be anticipated. They found, however, that Head Yaw was the most reliable kinematic variable for predicting walking turns about 200ms from commencing to turn. Their survey can inform the design of conflict resolution by enabling the robot to predict upcoming turns using these signals. Another recent survey focuses on algorithmic requirements and methodologies for robot navigation [102]. Their survey revolves mostly around perception and trajectory modeling rather than actuation. While the authors mention collision avoidance as an important robot navigation task; they do not focus their survey on collision avoidance, as presented here.

The survey by Xiao et al. [172] reviews methods that use machine learning techniques for the general problem of mobile robot navigation. Their survey focuses on the comparison between machine learning and classical approaches in terms of their scope and performance on real-world navigation problems. In contrast, this survey is on social navigation, focusing specifically on conflict avoidance, and the papers may use any (learning or non-learning) method in approaching the problem. For a more general perspective on the current state of social navigation, Mavrogiannis et al. [96] identified three broad themes that are being investigated: planning, behavior design, and evaluation. These themes impact all social navigation tasks rather than being specific to conflict avoidance, and thus their discussion does not focus on this aspect. This survey is more specific to the context of collision avoidance in social navigation, and it drills down to provide an elaborate taxonomy of models and algorithms for such scenarios.

The remainder of this survey is organized as follows: Section 2 proposes a taxonomy for social navigation, identifying important factors of the social navigation problem. Sections 3 and 4 present a selection of relevant works that have contributed models and algorithms, respectively. Section 5

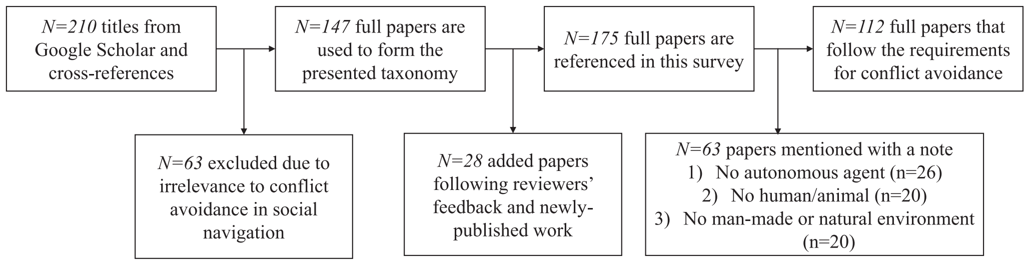

Fig. 1. Study flow diagram showing the inclusion methodology used in this survey.

focuses on the evaluation metrics used in social navigation and refers to some existing benchmarks. Finally, Section 6 highlights open problems in social navigation concerning the proposed taxonomy and provides a checklist for researchers to consult when investigating a new social navigation problem.

2 TAXONOMY

This section systematically describes a taxonomy used in this survey to categorize social navigation models (Section 3) and algorithms (Section 4). Here we describe the process used to collect the papers used in this survey. We started with existing surveys on social navigation [21, 83] and we collected all of their references, as well as papers that cite these works using Google Scholar. In selecting which papers to include, we used the criteria specified in Section 1 to guide the process. Overall, this survey contains 63 (out of 175) citations that do not meet all three criteria outlined in Section 1, but which nonetheless provide fundamental contributions to our understanding of the social navigation problem; or which are surveys on topics relevant to social navigation. We iterated through the process of collecting papers that cite, and are cited by, our current bibliography, until doing so yielded no new papers meeting all of the outlined requirements. The only exception to this process is when several papers have been published by the same group. Research groups often publish multiple papers on the same project. In these cases, we include more than one paper if they are categorized differently by our taxonomy. Otherwise, we include only the most recent paper. Figure 1 summarizes the paper selection process for this survey.

For each of the resulting 112 papers on conflict avoidance in social navigation, we identify seven attributes, listed in Table 1. Below we discuss this list of attributes (in bold) and the values (in italics) they can take. (Abbreviations for many values are used in tables in Sections 3 and 4. These abbreviations appear in parentheses next to their corresponding value.). We acknowledge that not all papers can be situated precisely within this taxonomy. In these cases, or if the value is not stated in the relevant paper, we label the corresponding attribute with the value "None" or "Neither" (e.g., some of the papers do not provide any empirical analysis, and thus the experiment type attribute is "None"). This taxonomy is constructed with the goal of encompassing as much work as possible, such that any new contribution can be easily placed in a clear context.

2.1 Taxonomy Attributes and Values

Some of the attributes and the values presented here are not intuitive. Here, we explain their rationale.

> **Number of Agents** *Absolute Number (Abs)/Density (D).* Some papers deal with a one-on-one interaction whereas others deal with multiple agents in a shared space. We mention when known, how crowded the environment is. Most works report either an *Absolute Number* of participants or a *Density* (measured as $\#people/m^2$). When presenting an absolute number

Table 1. The Social Navigation Taxonomy

Attributes	Values
Robot Role	*Reactor (R)/Initiator (I)/Both (B)/Neither (N)*
Number of Agents	*Absolute Number (Abs = # of agents)/Density (D = #/*m^2*)*
Observability	*Full/Partial/Depth/RGB*
Motion Control	*SFM/ORCA/ROS/Human/Other*
Communication	*None (N)/Indirect (I)/Direct (D)*
Experiment Type	*Simulation (Sim)/In the Lab (Lab)/In the Wild (ItW)/Survey (Sur)*
Agent Type	*Human-Robot (H-R)/Human-Agent (H-A)/Human-Human (H-H)/ Robot-Robot (R-R)/Homogeneous Agents (Hom)/Heterogeneous Agents (Het)*

of pedestrians, we include the navigating robot in the count. This allows comparison with multi-robot research where the number of agents includes multiple robots that are running the same algorithm.

Observability *Full/Partial/Depth/RGB*. If the work is set up in simulation, the robot can have either full or partial observability. Work that involves experiments or evaluations with real robots usually reports specific type(s) of sensors that were used, such as depth sensors (e.g., LIDAR), or cameras (e.g., RGB, or RGBD). If more than one type of sensor is used, we mention all of them.

Motion Control *SFM/ORCA/ROS/Human/Other*. Most robots in these papers rely on an existing motion controller, and the robot is augmented with a new component for social navigation. This survey classifies the main types of motion control used in these papers: the **Social Force Model** (***SFM***), **Optimal Reciprocal Collision Avoidance** (***ORCA***), the ROS `move_base` navigation stack (*ROS*),[1] evaluation of human behavior without any existing robot (*Human*), and *Other*. The "other" category includes both papers in which the motion control is not significant (such as research projects that use cellular automata, point-based navigation, Dijkstra's algorithm, or other types of search for motion planning), and in which the motion control is novel and is a major part of the paper's contribution (such as Social Momentum [98] or LM-SARL [23]). We mention the specific motion control that is used when possible.

Communication *None (N)/Indirect (I)/Direct (D)*. This attribute refers to communication that is conveyed by the robot, and not to communication that is conveyed by the other agents. *None* means that the robot is not doing anything specifically to convey its navigational goal. *Indirect* communication refers to situations where the robot uses whatever mechanisms it already possesses to signal its intentions, such as legibility [34]^{R1} and stigmergy [9]^{R2}. *Direct* communication means that there is some mechanism that is added to the robot to allow communication. See Figure 2 for examples.

Experiment Type *Simulation (Sim)/Laboratory (Lab)/In the Wild (ItW)/Survey (Sur)*. Many researchers run experiments in *Simulation* as part of their evaluation, either as the only type of evaluation or in addition to real-world experiments. *Laboratory* experiments are defined as experiments in the real world in a controlled environment such as a laboratory or using a scripted scenario. *In the Wild* are real-world experiments in an unstructured environment or with no predefined script for the pedestrians. All of these types of experiments can be accompanied by post-interaction *Surveys*. When a paper reports on more than one type of experiment, we include the details of one experiment, ordered in this prioritized order: In

[1] https://www.ros.org/

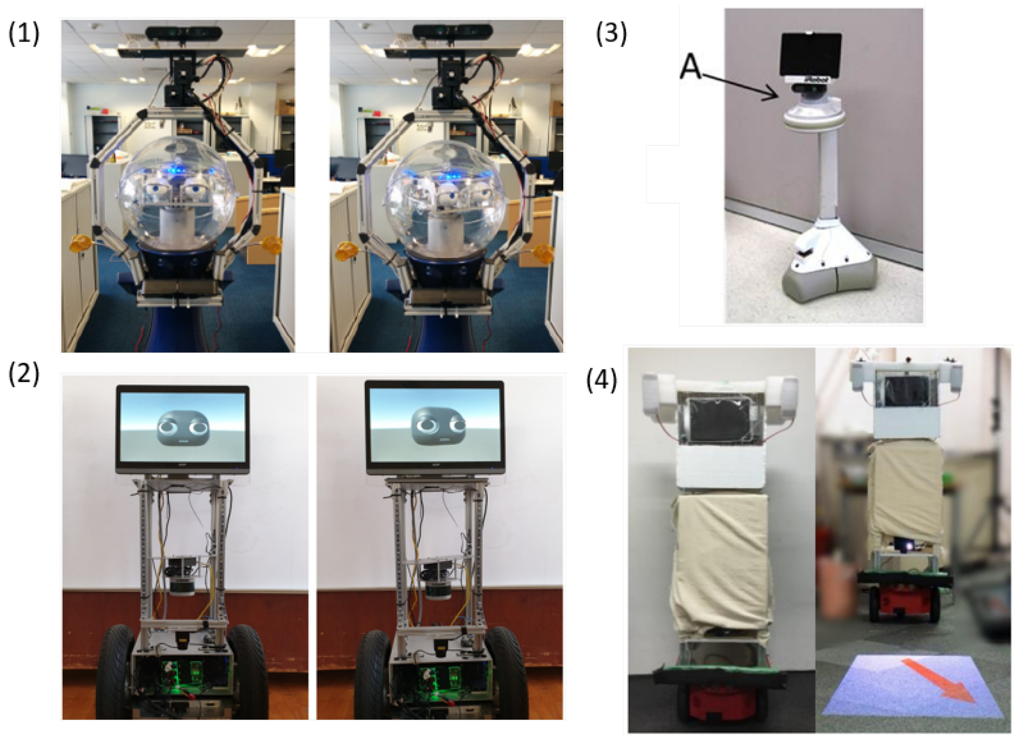

Fig. 2. Various direct communication behaviors: (1) mechanical gaze [99], (2) virtual gaze [55], (3) sensor rotation [39], (4) arrow signaling [140].

the Wild, Laboratory, Simulation, and Survey (when more than one methodology is used). There are two exceptions to this policy: the first exception regards surveys, which are often used as an additional metric for an experiment in the wild or the laboratory. Thus, if an experiment is accompanied by a survey, the survey is also mentioned. The second exception regards papers that report two or more experiment types, where one of them is a small-scale in the wild experiment that does not report significant results. In such cases, we report the paper according to the experiment with reported results, but add a superscript + symbol next to it to indicate that the paper also includes an in the wild experiment (e.g., Lab^{+}).

Agent Type *Human-Robot (H-R), Human-Agent (H-A), Human-Human (H-H), Robot-Robot (R-R), Homogeneous Agents (Hom), Heterogeneous Agents (Het).* This survey focuses on social navigation involving a person and a robot (*Human-Robot*). Due to the difficulty of evaluating such interactions, many models and algorithms are evaluated on a different set of agents. The most common approach is running a simulation in which the human counterparts are controlled by a real human (*Human-Agent*) or by some other set of predefined or learned behaviors (either *Homogeneous Agents* or *Heterogeneous Agents*). Several included papers provide a fundamental understanding of human navigation and present evaluations that do not involve robots at all (*Human-Human*) or that do not involve humans (*Robot-Robot*). These papers are cited using the notation presented earlier (e.g., citation^{R1}), highlighting that they do not satisfy one or more of the inclusion criteria.

Observability is important to consider, especially when discussing simulations. Simulations explicitly model the observations that can be made by agents acting in a scene. Many simulations assume that a robot (or pedestrians around it) has full (ground truth) observability. Other simulations restrict observability in artificial ways, attempting to emulate realistic sensing capabilities (partial observability). In the discussion of these papers, it is important to note that some observation modalities may be unrealistic to implement on real robots. The conclusions of such papers may not translate to the context of real-world embodied social navigation.

With respect to the **Communication** attribute, we make a distinction between communication that is *indirect* or *direct* and communication that is implicit or explicit. Implicit communication is often used to describe any non-verbal communication that is conveyed by people (e.g., the interpretation of eye gaze is implicit), and explicit communication is performed specifically with the intention of communicating with others (e.g., speech is explicit) [29]^{R2,R3}. Robots do not generally naturally communicate implicitly (for example, not all robots have "eyes" and those that do not necessarily need to turn them to "look" at something, or reflexively turn them to where they are about to navigate). As such, we make the distinction between *direct* and *indirect* communication as defined above and keep the implicit/explicit distinction as one reflective of mimicking human behavior. Using these definitions, the possible combinations for robot communication are implicit-indirect (e.g., velocity change [164]^{R1}), implicit-direct (e.g., gaze change on a virtual head [1]), and explicit-direct (e.g., arrow projections on the floor [170]).

Some papers present more than one set of experiments. An example would be presenting both a *Human-Robot* laboratory study and a simulation of *Heterogeneous agents*. We choose to highlight *Human-Robot* experiments; a particularly relevant format for studies in social navigation. In general, for papers that present more than one set of experiments we categorize them by the values most relevant to HRI on the attributes of **Experiment Type** and **Agent Type**: *In the Wild* > *In the Lab* > *Simulation* > *Survey*.

We also highlight that some of the taxonomy attributes are very concrete and define low-level components used in the interaction (e.g., the motion control used), while other attributes are more abstract (e.g., robot role). Usually, the abstract attributes and their values depend on the concrete attributes. Figure 3 presents the hierarchical structure of these attributes, in work that is consistent with the three requirements outlined in Section 1. The bottom part represents the attributes that are independent of other attributes. The values assigned to the attribute at the end of an edge affect the values that can be assigned at its origin. For example, the values of the **Communication** attribute will be directly affected by the **Number of Agents** in the environment and the robot's **Observability**. In turn, the choice of value for the **Communication** attribute directly affects the **Agent Type**s that can perceive the chosen communication channel.

2.2 Additional Concepts

There are some additional concepts that are worth mentioning, but which we decided to exclude from our taxonomy. Here we list these concepts and explain why they are not included in the taxonomy. As research and discussion on social navigation progresses, this taxonomy could be extended to include these attributes.

One seemingly-important factor to consider in the taxonomy is **collision type**. When referring to collisions, most articles describe head-on collisions or side-on collisions, with rear-end collisions as the least commonly investigated type. Among the articles in this survey, none explicitly discuss only one type of collision. There are several articles that propose ways to categorize collisions according to the required response from pedestrians and/or the robot. Reynolds [130] defines two types of collisions: unaligned collision avoidance and separation. Unaligned collision avoidance is a behavior that *"tries to keep characters which are moving in arbitrary directions from*

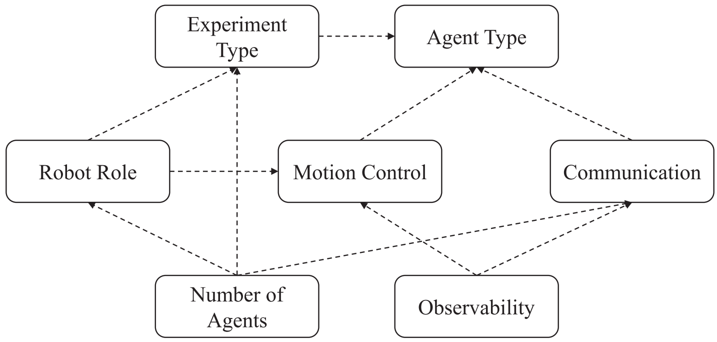

Fig. 3. The hierarchical structure of the taxonomy's attributes.

running into each other [130]." Separation is similar to a rear-end collision and refers to a simpler form of movement: *"Separation steering behavior gives a character the ability to maintain a certain separation distance from others nearby[130].*" Mavrogiannis et al. [98] discuss the point in space and time where agents collide, calling this point "entanglement." This concept raises an additional question about the concrete implementation of this collision point—what is considered close enough to be an entanglement in a social context? For example, Mavrogiannis et al. [97] utilized a minimum distance of $d \leq 1$ meter between the robot and the human. While it is simple to classify the direction of a collision, it is more challenging to define properly the minimal requirements of an encounter to be considered a collision. Is entering a person's personal space a collision? Is brushing against their leg? Overall, the definition of collision varies between researchers and may be a parameter that can be adjusted.

Another common discussion point is **context awareness and semantic mapping**. Many articles discuss the need for mobile social robots to be aware of their context [22]. A leading approach to enable this is semantic mapping, where the robot constructs maps that represent not only a metric occupancy grid but also other properties of the environment [79]^{R1,R3}. This survey does not focus on the mental model of the navigating robot (or of the other agents) in the environment, so this is not included in the taxonomy. It is, however, an important factor to consider when designing a robot for social navigation, as context awareness could greatly influence a robot's behavior.

Another thing to consider when designing interactions between mobile robots and pedestrians is how people **react to humans vs. robotic counterparts**. Will humans interacting with other people produce a similar or different response when interacting with robots? The assumption that people will behave in the same way when encountering a robot as they would another human is common in HRI and other research communities, though it is not unanimously agreed upon. In their survey on proxemics for social navigation, Rios-Martinez et al. [132] stated that *"This article starts from the idea that people will keep the same conventions of social space management when they interact with robots than when they interact with humans. Researchers in social robotics that believe in that hypothesis can rely on the rich sociological literature to propose innovative models of social robots."* As a counter opinion, Butler and Agah [13] indicate that people are most comfortable when a robot moves at speeds that are between $0.254 m/s$ and $0.381 m/s$, while the normal walking speed for young humans is about $1 m/s$. This difference suggests that people prefer a robot that moves more slowly than people do. Until there is a clear theory regarding the reactions of people to other people vs. robots in social navigation—and until that theory is tested—it is reasonable to exclude assumptions regarding whether people react to robots similarly or differently from how they react to other people from this taxonomy.

The distinction between **social cues and social signals** [168][R2,R3] is used in this survey, but they are not included as attributes in the taxonomy. Cues are the low-level inputs that the robot can receive or send, such as gaze, position, language, and so on. Signals, on the other hand, are emotions, personality, and other traits that are more high-level. Signals discussed in the context of social navigation usually serve a purpose in conflict avoidance, and the way to implement them in a robot (or detect them in a human) is through social cues. How a robot can best communicate with humans is a rich and versatile research area; and is taken into consideration through the attributes **observability** and **communication** in the taxonomy.

One attribute that is relevant in a broader context than social navigation is **focused vs. unfocused interaction**. Goffman [49][R2,R3] defines these terms to categorize scenarios in which the robot and the human share their focus (shared attention) vs. scenarios in which the robot and the human share an environment, but not attention. Rios-Martinez et al. [132] use this attribute to identify different types of navigational behaviors in robots: minimizing the probability of encounter, avoiding collisions, passing people, staying in line, approaching humans, following people, and walking side-by-side. Because the articles in this survey revolve around conflicts, the robot and the human do not share focus, and hence all included articles involve strictly unfocused interactions. Focused vs. unfocused interactions are not considered as part of the taxonomy.

Additionally, the topic of differences in navigation between **independent pedestrians, groups**, and **crowds** has enjoyed recent popularity [51, 106, 174]. Most social navigation articles either consider interactions with a single individual or with a crowd of individuals (as defined as **Number of Agents** in our taxonomy). An early sociological study showed that people tend to move in small groups rather than alone, but that group size distribution depends greatly on context (a casual Saturday afternoon stroll vs. a workday morning commute) [27][R2,R3]. Recent research has demonstrated that in many contexts, more than 50% of pedestrians are traveling in groups [104][R2,R3]. Thus, the context in which navigation takes place determines whether it is necessary to consider the surrounding crowd.

Lastly, we address a distinction that is relatively straightforward to understand intuitively but is challenging to formalize: Conflict **Prevention vs. Resolution**. Consider a person walking in a crowded environment who is looking at their phone. Without watching the surrounding crowd, two people might collide—which means they have reached a conflict. If the person looks up early enough, they might side-step abruptly without a change of speed—which means that a conflict was resolved. If the person decides to step away to a less crowded area, this behavior is prevention. On one hand, it is clear that prevention and resolution are different tasks that can direct the robot's behavior: prevention is the task of designing the robot's motion to steer away from potential conflicts, while resolution is the task of altering the robot's motion and behavior when a conflict is already imminent. On the other hand, formalizing this distinction is challenging, as it is non-trivial to define what is an "imminent" conflict. Whether a robot is designed to prevent or resolve a conflict, the premise of all of the covered work in this article is that the robot is always attempting to avoid conflicts in social navigation. This requirement provides a crisp way to identify relevant articles that fit into this survey, without the need to explicitly cluster the interaction into prevention or resolution.

3 MODELS

This section details various models used for social navigation. The discussion is grouped according to three main underlying models: MAS, human-inspired models, and physics-based models (specifically, the SFM and other force models). Each of these categories represents a different set of assumptions—as well as a different research community—that each model stems from. Navigation in MAS is usually designed with the premise that agents navigating in an environment are homogeneous. These articles include multi-robot navigation models and crowd modeling. A multiagent social navigation model generally reasons about agents with different—sometimes unknown—behaviors. Other models are inspired by insights about human navigation. These articles provide measurements and rules that explain how people navigate among each other. Such a social navigation model translates these rules into robot motion and perception. We taxonomize articles using the social force model in their own category; inspired by physical force modeling. Many models have been proposed which build upon the seminal work by Helbing and Molnar [58]^{R2,R3}, using additional forces. Finally, some papers sit at an intersection between two categories. In such cases, the paper is grouped with work that uses similar motion control.

3.1 Multiagent Systems

Two communities that have contributed significantly to the study of social navigation are the multi-robot navigation and graphics communities. Both of these communities have proposed different approaches to model the behavior of a crowd. The multi-robot community focuses more on safety and feasibility in the real world, while the graphics community focuses on robustness. Multi-robot work usually is based on a few interactions under realistic constraints. On the other hand, the challenge of crowd modeling taken on by the graphics community is to model interactions between hundreds and thousands of agents simultaneously. However, because the graphics community does not need to implement these systems on real robots, the perception and movement restrictions on those agents tend not to be grounded in the physical constraints that both robots and real people must contend with.

Many researchers have approached the challenge of multi-robot navigation [173]^{R1,R3}. This is a fertile and active research area that deserves its own survey. We discuss only a few selected publications that have had a significant influence on social navigation. Van Den Berg et al. [164] present the principle of ORCA, which provides a sufficient condition for multiple robots to avoid collisions among one another, and guarantees collision-free navigation. Chen et al. [23] model human-robot and human-human interactions, then infer the relative importance of learned features through a pooling module via a self-attention mechanism, and finally planning motions.

Another branch of multi-robot research focuses on planning under uncertainty and leverages **Markov Decision Processes** (**MDPs**). Foka and Trahanias [40] model a probabilistic prediction of people's destinations. They use a **Partially Observable MDP** (**POMDP**) solved online at each time step to determine which actions the robot actually performs. Gupta et al. [51] recently presented an additional POMDP model for intention-aware navigation in crowds, where the model can address decisions related both to the robot's speed and its heading. Bandyopadhyay et al. [3] model human intention with a **Mixed Observability MDP** (**MOMDP**), and then plan the motion of a robot leveraging this model.

The graphics community has contributed several important models to social navigation, as well as simulation environments that can be utilized to evaluate other models and algorithms (see more about these simulation environments in Section 5). Musse and Thalmann [108] propose a model of crowd behavior, where agent behavior is determined using a predefined set of rules. Strassner and Langer [149] use behavioral rules for modeling each person's behavior in a crowd. Such behaviors include perceiving, storing, and forgetting knowledge. Bonneaud and Warren [11] model pedestrian behavior using an empirically-grounded emergent approach, where the local control laws for locomotor behavior are derived experimentally, and the global crowd behavior is emergent. Okal and Arras [117] present a model for crowd behavior in which groups are formed. Their representation gives each individual an internal state, where under a set of predefined conditions pedestrians can choose to walk together.

Table 2 summarizes the taxonomy values for models inspired by MAS research.

Table 2. An Overview of the Different Multiagent-based Models Used in Social Navigation

Year	Paper	Role	# Agents	Obs.	Motion Control	Com.	Exp. Type	Agent Type
1997	Musse and Thalmann [108]	R	Abs = 10	Full	Other (Hand Coded)	N	Sim	Hom
2005	Strassner and Langer [149]	N	Abs = 2	Partial	Other (Hand Coded)	N	Sim	Hom
2010	Foka and Trahanias [40]	R	Abs = 6	Depth	Other (POMDP)	I	ItW	H-R
2011	Van Den Berg et al. [164]	R	Abs = 1000	Full	ORCA	I	Sim	Hom
2013	Bandyopadhyay et al. [3]	R	Abs = 4	RGB + Depth	SFM	N	Lab	H-R
2014	Bonneaud and Warren [11]	R	Abs = 20	Full	Other	N	Sim	Hom
2014	Okal and Arras [117]	R	Abs = 176	Partial	SFM	N	Sim	Hom
2016	Godoy et al. [48]	B	Abs = 100	Full	Other	N	Lab	R-R
2019	Chen et al. [23]	R	Abs = 6	Full	Other (LM-SARL)	N	Sim$^+$	Hom
2022	Gupta et al. [51]	R	Abs = 401	Partial	Other (POMDP)	N	Sim	Hom

Role refers to the robot role, **Obs.** is observability, **Com.** refers to communication, and **Exp. type** is the experiment type.

3.2 Psychology and Human-inspired Models

The contributions discussed so far have focused on multiagent or multi-robot navigation systems that have been adapted to accommodate human pedestrians. A different approach starts with the modeling of human behavior, which then leverages these models to improve robot navigation. Cutting et al. [30]^{R2,R3} empirically evaluates human behavior in situations of obstacle avoidance. Their work investigates the relationship between object avoidance and finding one's aimpoint in a series of human studies. Their results are summarized as a decision tree to facilitate reasoning about collision detection with other objects (static or moving) and **Gaze-Movement Angle** (**GMA**); the angle between one's gaze and one's direction of movement. Their model can be used to estimate where a collision might occur. As a different way to estimate an expected collision point, Schewe and Vollrath [15] defined τ to be the time to bypass a dynamic obstacle (human or not). Moussaïd et al. [103] use τ to heuristically plan how to navigate in a way that avoids collisions. Park et al. [122] claim that GMA-based collision prediction has several advantages over the time-to-contact (τ) approach. It is more robust to variations in the speed and the path of the other pedestrian. It also does not assume either constant speed or a linear path, so the accuracy of the prediction is not affected by these variations. Kitazawa and Fujiyama [78] investigate the **Information Process Space** (**IPS**) of a navigating person when walking in a hallway in the presence of static objects and other pedestrians. In this work, they identify the area that the observing pedestrian considers as the one in which a collision with another pedestrian could occur in a short time (see Figure 4). In an extension of this work, Park et al. [122] propose a collision avoidance behavior model that is based on their empirical results about IPS to generate more human-like collision avoidance behaviors.

Another concept from psychology that has had a significant impact on social navigation is that of personal space [46, 53, 62]. While the original formulation of personal space is depicted by Hall [53]^{R2,R3} as a concentric circle, later work extends that to an egg [56]^{R2,R3}, ellipse [58], or as asymmetrical (smaller on the dominant side) [46]^{R2,R3} shape. Closely related to personal space is the concept of density in crowds. The average density of people in a non-crowded environment has been evaluated to be 0.03 pedestrians per m^2, whereas, in a moderately crowded environment, there are 0.25 pedestrians per m^2 [104]. Rios-Martinez et al. [131] incorporate both personal space

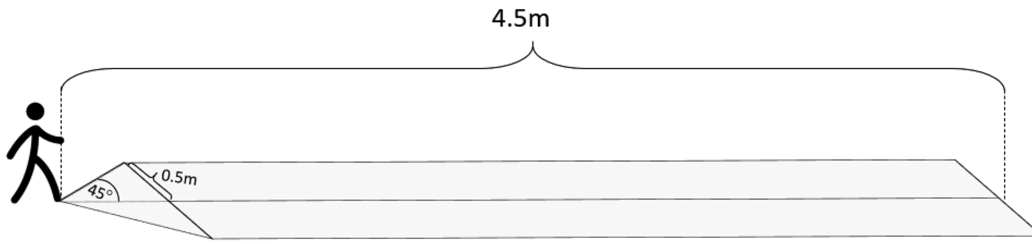

Fig. 4. IPS—the visual processing coverage of pedestrians, as measured by [78] and depicted by Rios-Martinez et al. [132].

and IPS-based constraints into an adaptive optimization algorithm to enable more human-like navigation. Truong and Ngo [161] propose a comprehensive framework that reasons about pedestrians' extended personal space and social interaction space to identify a **Dynamic Social Zone** (**DSZ**); a concept which is incorporated into their motion planner.

Others have analyzed how gait and posture are affected by sudden trajectory changes, as one might expect to see in conflict avoidance. Patla et al. [123]^{R2,R3} analyzed head yaw, trunk yaw, and foot position when turning due to an expected obstacle vs. turning abruptly due to an unexpected obstacle. To analyze the relationship between head pose and predicted walking trajectory, Unhelkar et al. [163]^{R2,R3} discretized walking trajectories as a decision problem regarding which target a person would walk toward. They incorporated this information into an anytime path planner [110]^{R1} and evaluated this enhanced planner in simulation. Holman et al. [60]^{R2,3} extend this predictive model to incorporate gaze. Senft et al. [137] identify and implement a navigational pattern for making space in a hallway. Their model involves controlling the robot's rotation and sliding motion and consists of three steps: step, slide, and rotate.

All of the contributions above leverage insights from empirical studies on humans and robots to manually construct models for social navigation. However, together with the improving abilities of machine learning, different learning techniques have been used to learn models of navigation in social contexts. Lu et al. [90] propose a planning model that can be tuned to match different social navigation contexts. Bennewitz et al. [7] learn motion patterns of people that can be used for trajectory prediction in social robots. Henry et al. [59] extend this approach by modeling partial trajectories. More recently, Vasquez et al. [165] used **Inverse Reinforcement Learning** (**IRL**) to infer a reward function for social navigation. They introduce a new software framework to systematically investigate the effect of features and learning algorithms used in the literature. They investigate the task of socially compliant robot navigation in crowds, evaluating two different IRL approaches and several feature sets in large-scale simulations. Karnan et al. [71] collected a large-scale human demonstration dataset, containing socially compliant data of navigation behaviors in natural indoor and outdoor spaces on a university campus. They used behavior cloning to learn a global and local planner to mimic human navigation behaviors.

Table 3 summarizes the taxonomy values for models inspired by human behavior, physiology, and psychology research.

3.3 Physics-based Models

Researchers have also used models inspired by physics to represent dynamics and interactions between different moving agents. Helbing and Molnar [58] were the first to propose the SFM, a model inspired by fluid dynamics that describes an agent's motion using a set of repelling and attracting forces. They evaluate this model in a simulation of homogeneous SFM-based agents. Many contributions extend SFM models to handle additional forces: Karamouzas et al. [69] add an evasive

Table 3. An Overview of the Different Human-inspired and Psychology-based Models Used in Social Navigation

Year	Paper	Role	# Agents	Obs.	Motion Control	Com.	Exp. Type	Agent Type
1995	Cutting et al. [30]	R	Abs = 2	Partial	Other (Hand Coded)	I	Sim + Sur	H-H
1998	Jeffrey and Mark [62]	R	Abs = 4+	Partial	Human	N	Sim	H-H
1999	Patla et al. [123]	R	Abs = 2	Partial	Human	N	Lab	H-H
1999	Reynolds [130]	R	Abs = 2	Full	None	N	None	Hom
2002	Bennewitz et al. [7]	R	Abs = 2	Depth	Other (Learned)	N	Lab	H-R
2008	Gérin-Lajoie et al. [46]	R	Abs = 2	None	Human	N	Lab	H-H
2010	Henry et al. [59]	R	None	Depth	Other (A*)	N	Sim	Hom
2010	Kitazawa and Fujiyama [78]	R	Abs = 4	Partial	Human	N	Lab	H-H
2011	Moussaïd et al. [103]	R	Abs = 96	Partial	Other (Hand Coded)	N	Lab	Hom
2011	O'Callaghan et al. [115]	R	Abs = 2	Depth	Other (Planner)	N	ItW	H-R
2012	Rios-Martinez et al. [131]	R	Abs = 6	RGB+Depth	ROS	N	Sim	Hom
2013	Lu et al. [90]	R	Abs = 2	Partial	ROS	N	Sim	Het
2013	Park et al. [122]	R	D = 0.1–1	Partial	Other (Hand Coded)	I	Sim	Hom
2014	Charalampous et al. [21]	R	Abs = 2	RGB + Depth	Other	N	ItW	H-R
2014	Papadakis et al. [121]	I	Abs = 2	RGB + Depth	None	I	Lab	H-R
2014	Vasquez et al. [165]	R	Abs = 6+	Full	Other (Dijkstra)	I	Sim	H-A
2015	Unhelkar et al. [163]	R	Abs = 2	Full	Other (SIPP)	N	Lab	H-A
2016	Mead and Matarić [100]	I	Abs = 2	RGB	None	D	Lab	H-R
2016	Truong and Ngo [161]	R	Abs = 4	RGB+Depth	Other (D*)	N	Lab	H-R
2020	Senft et al. [137]	R	Abs = 2	Depth	Other (Hand Coded)	I	Lab	H-R
2022	Karnan et al. [71]	I	Abs = 2	RGB + Depth	Other (Teleoperation)	I	ItW	H-R

Role refers to the robot role, **Obs.** is observability, **Com.** refers to communication, and **Exp. type** is the experiment type.

force that uses collision prediction and avoidance, which makes agents more proactive and anticipatory than the classical SFM. Moussaïd et al. [104] propose several group-related forces that help model pedestrians that walk in a group. Swofford et al. [151]^{R2,R3} use a **Deep Affinity Network** (**DANTE**) to predict the likelihood that two individuals in a scene are part of the same conversational group. They take into consideration the social context in which these interactions take place. A different type of force-inspired work uses potential fields attached to moving pedestrians [150]. This model has been leveraged in a modified **Rapidly-exploring Random Tree** (**RRT**) for navigation in human environments, though it assumes access to full state information.

Table 4 summarizes the taxonomy values of models inspired by physics and mechanical engineering research.

4 ALGORITHMS

This section discusses contributions in the form of algorithms and hardware augmentations that enhance social navigation. Most of the work presented here fits our basic definition of social navigation, however, several papers are included that have not been evaluated in the context of navigating around people. These papers are included if their contribution can be applied in the context of social navigation. Broadly speaking, this section is divided into three main approaches: Approaches

Table 4. An Overview of the Different Physics-inspired Models Used in Social Navigation

Year	Paper	Role	# Agents	Obs.	Motion Control	Com.	Exp. Type	Agent Type
1995	Helbing and Molnar [58]	R	D = 0.3	Full	SFM	N	Sim	H-A
2003	Loscos et al. [89]	R	Abs = 6000	Partial	Other (Hand Coded)	N	Sim	Hom
2009	Karamouzas et al. [69]	R	Abs = 1000	Full	SFM	N	Sim	Hom
2010	Moussaïd et al. [104]	N	D = 0.03-0.25	Full	SFM	N	ItW	H-H
2010	Svenstrup et al. [150]	R	Abs = 40	Full	Other (Modified RRT)	I	Sim	Hom
2020	Swofford et al. [151]	I	Abs = 18	RGB	ROS	N	Lab	H-R

Role refers to the robot role, **Obs.** is observability, **Com.** refers to communication, and **Exp. type** is the experiment type.

that **infer** the human's trajectory and adapt to it; Approaches that **convey** the goal or trajectory of the robot to the person it is interacting with before reaching a conflict; and mixed approaches which **mediate** between the inferred trajectory of the human and the desired goal of the robot.

4.1 Inferring Human Trajectories

Many social navigation contributions have been inspired by the way humans navigate in social contexts. The majority of these papers can be split into two categories: online and offline inference. Online inference means that a robot observes the behavior of a person during deployment and incorporates its inference about the person's planned trajectory into its execution. Offline inference happens prior to the execution stage, usually on more than a single trajectory. The robot learns to predict human trajectories or imitate them from a set of observed trajectories.

4.1.1 Online Inference. Cutting et al. [30] offer an early attempt to evaluate the trajectory of a passerby by calculating their GMA and reacting to it. The robot designed by Tamura et al. [155] detects pedestrians by using a laser range finder and tracks them using a Kalman filter. They apply an SFM to the observed trajectory to determine whether the pedestrian intends to avoid a collision with the robot or not and select an appropriate behavior based on the estimation result. Gockley et al. [47] discuss how to avoid rear-end collisions in the context of person following. They propose a laser-based person-tracking method and evaluate two different approaches to person-following: direction-following, where the robot follows the current location of the person; and path-following, where the robot tries to follow the exact path that the person took. They show that while no significant difference was found between the two approaches in terms of the distance or time between tracking errors, participants rated the robot's behavior as significantly more natural and human-like in the direction-following condition. In addition, participants felt that the direction-following robot's behavior was more similar to the participants' expectations.

Others have leveraged the human gaze to infer the trajectory of pedestrians. Gaze is a very strong communicative cue used by humans, in the context of collaborative settings in general [19]^{R2,R3} and for navigation in particular [2]. It has been shown that humans are not the only species that can partially understand gaze cues from a very young age, but also chimpanzees and dogs [126, 147]^{R2,R3}. Gaze and head pose have both been shown to be significant indicators of a person's attention, which can be used to infer navigational goals. Stiefelhagen et al. [148]^{R2,R3} show that the visual focus of a person's attention can be deduced from head pose when the visual resolution is insufficient to determine eye gaze. Smith et al. [146]^{R2,R3} extend their work to a varying number of moving pedestrians. Of course, this gaze behavior extends beyond walking and bicycling. Recent work has studied the use of gaze as a modality for plan recognition in games [144] and as a cue for interacting with copilot systems in cars [63, 64], also to infer the driver's intended trajectory. Gaze is also often fixated on objects being manipulated, which can be leveraged to improve algorithms which learn from human demonstrations [134]^{R1,R2}. Though the use of instrumentation such as head-mounted gaze trackers or static gaze tracking cameras is limiting for mobile robots, recent work in the development of gaze trackers that work without such equipment [133]^{R1,R2} may soon allow us to perform the inverse of the robot experiments presented here, with the robot reacting to human gaze. Ratsamee et al. [128] propose to avoid collisions with humans by considering a social model that takes into consideration body pose and face orientation.

4.1.2 Offline Inference and Learning. While the previous subsection focused on the recognition of human trajectories during execution, some leverage these trajectories to learn and infer how a human would react in a social navigation interaction. Pacchierotti et al. [119] designed a rule-based strategy for people passing that was inspired by spatial behavior studies. This strategy intends to mimic the way people avoid collisions once inside a person's personal space.

One such successful approach uses **IRL** to elicit the explicit cost representation to imitate human's social navigation behavior. Instead of hand-crafted functions, these papers use IRL to leverage data-driven approaches. IRL was extensively used to infer reward (cost) functions from human demonstrations. The most straightforward application of IRL is by Kim and Pineau [74], to learn a cost function that respects social variables over features extracted from an RGB-D sensor. This work used IRL to infer cost functions in a social navigation context: navigational features were first extracted from an RGB-D sensor, then represented as a local cost function learned from a set of demonstration trajectories by an expert using IRL. The system still operated under the classical navigation pipeline, with a global path planned using a shortest-path algorithm, and a local path using the learned cost function to respect social variables. Obstacle avoidance was still handled by a low-level controller. Okal and Arras [118] tackle cost function representation at a global level in a social context: they developed a graph structure and used Bayesian IRL to learn the cost for this representation. With the learned global representation, a traditional global planner (A*) planned a global path over this graph, and the POSQ steer function for differential-drive mobile robots served as a local planner. Henry et al. [59] use IRL to learn motion patterns of humans in simulation that can later be used for planning in social navigation.

An alternative approach to IRL with a similar objective is to model social navigation trajectories using a **Maximum Entropy Probability Distribution**, where cost is also implicitly defined by identifying an underlying model from demonstrated data. Maximum entropy probability distribution has been used by Pfeiffer et al. [124] to model agents' trajectories for planning and by Kretzschmar et al. [80] to infer the parameters of the navigation model that matches the observed behavior in expectation. Kuderer et al. [84] also use human demonstrations, but instead of using an MDP, they elicit features from the human trajectories and then use entropy maximization to determine the robot's behavior. Luber et al. [92] use unsupervised learning from surveillance data to learn motion patterns and augment a motion planner with this knowledge.

Sisbot et al. [145] create a **human-aware motion planner** (**HAMP**) that is explicitly given a cost model for safety and legibility, and the robot reasons about the joint cost of these two properties in its planning process. Costs were also implicitly defined by identifying an underlying model from demonstrated data. Kirby et al. [75] model human social conventions at the global planning stage. This enables it to mediate between different, sometimes conflicting objectives. For example, consider a goal that is down an intersecting hallway to the robot's left. While the social norm in many places is to pass a pedestrian from the right side, the robot may choose to walk across the hallway in front of an oncoming person, effectively passing them to the left of the corridor. This behavior is the result of mediating between two objectives: complying with the right-alignment social norm, and minimizing the time to the goal.

Many algorithms use hand-crafted behaviors to avoid conflicts, i.e., to realize collision avoidance. As a continuation of previous **Collision Avoidance Deep Reinforcement Learning** (**CADRL**) work [25], Chen et al. [24] further propose a hand-crafted reward function to incorporate the social norm of left or right-handed passing in a DRL approach and enabled a physical robot to move at human walking speed in an environment with many pedestrians, called **Socially Aware CADRL** (**SA-CADRL**). Along the same line of research, but to relax the assumption of other agents' dynamics, Everett et al. [36] propose GA3C-CADRL, using an LSTM to allow reasoning about an arbitrary number of nearby agents and GPU to maximize the number of training experiences. Similarly, the reward function by Jin et al. [65] accounts for ego-safety, to measure collision from the robot's perspective, and social-safety, to measure the impact of the robot's actions on surrounding pedestrians. Other options that utilize DRL include using a **Hidden Markov Model** (**HMM**) in a higher hierarchy to learn to choose between target pursuing and collision avoidance using RL [33]. Tai et al. [154] use **Generative Adversarial Imitation Learning** (**GAIL**) to learn continuous actions and desired force toward the target. This improved safety and efficiency over pure BC. Li et al. [85] propose a new problem: **socially concomitant navigation** (**SCN**). In addition to collision avoidance in traditional social navigation, in SCN the robot also needs to consider the motion of its companion to maintain a sense of affinity when they are traveling together towards a certain goal. Taking features extracted from a LiDAR sensor along with the goal as input, a navigation policy is trained by **Trust Region Policy Optimization** (**TRPO**) to output continuous velocity commands for navigation. Bera et al. [8] created SocioSense, a social navigation algorithm that categorizes pedestrians according to psychological traits (e.g., shyness, tense) and adjusts the robot's velocity according to the pedestrians around it. Lu et al. [91] incorporated a dynamic measure into their reward to reason about the density of the crowd when deciding on the distance from other pedestrians. They then extended the deep neural network architecture from SARL [23] to choose the optimal action with the shaped reward that reasons about the "uncomfortable distance" between the robot and a pedestrian.

To observe social rules when navigating in densely populated environments, Yao et al. [174] propose to utilize information about social groups to address the "naturalness" aspect from the perspective of collective formation behaviors in the complex real world. They used a deep neural network, called Group-Navi GAN, to track social groups and navigate the robot to join the flow of a social group by providing a local goal to the local planner. Other components of the existing navigation pipeline, e.g., state estimation, collision avoidance, and so on, remained the same. The classical navigation pipeline, with the assistance of a learned local goal, was capable of navigating safely in a densely populated area following crowd flows to reach the goal. Liang et al. [86] develop CrowdSteer, an RL-based collision-avoidance algorithm that navigates in dense and crowded environments. The algorithm is trained using PPO in simulation with simulated human agents and was deployed in the real world. Martins et al. [95] propose ClusterNav, an algorithm that gets human demonstrations using teleoperation, then uses Expectation Maximization to learn how to navigate in an unsupervised manner. Their approach cannot reason about dynamic obstacles, hence it is unable to reason about interactions with people during navigation, so it does not appear in our tables.

Table 5 summarizes the taxonomy values for the inference algorithms for social navigation discussed in this subsection.

Table 5. An Overview of the Different Inference Algorithms Used in Social Navigation

Year	Paper	Role	# Agents	Obs.	Motion Control	Com.	Exp. Type	Agent Type
2006	Pacchierotti et al. [119]	R	Abs = 3	Depth	Other (Hand Coded)	N	Sim	Hom
2007	Gockley et al. [47]	R	Abs = 2	Depth	Other (CVM)	D	Lab	H-R
2007	Sisbot et al. [145]	R	Abs = 2	RGB + Depth	Other (HAMP)	I	Lab	H-R
2009	Kirby et al. [75]	R	Abs = 4	Depth	Other (A*)	N	Sim	Hom
2010	Ohki et al. [116]	R	Abs = 5	Full	Other (Hand Coded)	N	Sim	Hom
2010	Pandey and Alami [120]	R	Abs = 2	Full	Other (Hand Coded)	N	Lab	H-R
2010	Tamura et al. [155]	R	Abs = 2	Depth	SFM	N	Lab	H-R
2011	Diego and Arras [32]	R	Abs = 5	None	Other (Modified TSP)	N	Sim	Het
2012	Kuderer et al. [84]	R	Abs = 3	Full	Other (learned)	N	Lab	H-R
2012	Luber et al. [92]	R	Abs = 2	Full	Other (RMP)	N	Sim	H-A
2013	Ratsamee et al. [128]	R	Abs = 2	RGB + Depth	SFM	N	Lab	H-R
2014	Gómez et al. [50]	R	Abs = 5	Full	Other (Planning)	N	Sim	R-R
2016	Kim and Pineau [74]	R	Crowd	RGB + Depth	Other (Costmap Search)	I	ItW	H-R
2016	Kretzschmar et al. [80]	R	Abs = 3	Depth	Other (RPROP)	I	Lab	H-R
2016	Okal and Arras [118]	R	Abs = 4	Depth	ROS	I	Sim$^+$	Hom
2016	Pfeiffer et al. [124]	R	Abs = 891	RGB	Other (Max Entropy)	I	ItW	Hom
2017	Bera et al. [8]	R	D< = 2	Full	Other (SocioSense)	N	Sim	Het
2017	Chen et al. [24]	R	Abs = 10+	RGB	Other (Learned)	N	ItW	H-R
2017	Chen et al. [25]	R	Abs = 6	Full	Other (Learned)	N	Sim	R-R
2018	Ding et al. [33]	R	Abs = 20	Depth	None	N	Sim	R-R
2018	Everett et al. [36]	R	Abs = 10+	RGB + Depth	Crowd	N	Sim$^+$	H-A
2018	Jiang et al. [63]	B	Abs = 2	RGB	Other (Hand Coded)	N	Sim	H-A
2018	Li et al. [85]	R	Abs = 3+	Depth	Other (Learned)	N	Lab	H-R
2018	Long et al. [87]	R	Abs = 100	Depth	Other (Learned)	N	Sim	R-R
2018	Tai et al. [154]	R	Abs = 3	Depth	Other (Learned)	N	Sim$^+$	H-A
2019	Jin et al. [65]	R	Abs = 4	Depth	Other (Learned)	N	Lab	H-R
2019	Meng et al. [101]	N	Abs = 1	RGB	None	N	Sim	Hom
2019	Nardi and Stachniss [111]	N	Abs = 1	Full	Other (Hand Coded)	N	Sim	R-R
2020	Liang et al. [86]	R	Abs = 10+	RGB + Depth	Other (Learned)	N	Lab	H-R
2022	Lu et al. [91]	R	Abs = 5	Depth	Other (Learned)	N	Sim	Hom

Role refers to the robot role, **Obs.** is observability, **Com.** refers to communication, and **Exp. type** is the experiment type.

4.2 Conveying the Robot's Goal to the Human

Dragan et al. [34] formally define the concepts of legibility (motion that allows the observer to confidently infer the correct goal) and predictability (motion that conforms with the observer's expectations) in robot navigation. They show that human-robot collaboration is affected by the way the robot plans its motion, and to perform better, the robot design should switch from a focus on predictability to a focus on legibility. This section presents several approaches to increase the robot's legibility and explicability, with an emphasis on interaction points where there is a conflict between the human pedestrian and the robot. More details about the specific mechanisms that are activated in humans when interacting with a robot can be found in the work by Sciutti et al. [136], who survey the concept of "motor resonance" between an acting robot and an observing human. Kitagawa et al. [77] recently presented a motion planning algorithm for omnidirectional robots to resemble human movements in a time-efficient manner.

Many contributions use verbal signals for guidance [156]. Jeffrey and Mark [62] investigate human navigational behavior in the context of two simulated environments. In these simulations, people could communicate using either text messages or audio. Yedidsion et al. [175] investigate how verbal instructions given by more than one robot can assist humans in navigation in a new environment. However, for the social navigation task, verbal communication is considered less useful, as the navigation is expected to take place seamlessly without demanding the high awareness level that verbal communication requires [20]. To deal with this challenge, many contributions take inspiration from the theory of proxemics [53] as a non-verbal way to convey intent or restriction. Rios-Martinez et al. [132] investigate the comfort zone of people when a robot approaches them and Torta et al. [158] identify specific values for this comfort zone (182 cm from a sitting person and 173 cm from a standing person) or imitate them from a set of observed trajectories, and uses the learned model for online planning.

LED and Artificial Signals. Baraka and Veloso [4] use an LED configuration on their CoBot to indicate some robot states—including turning—focusing on the design of LED animations to address legibility. Their study shows that the use of these signals increases participants' willingness to aid the robot. Shrestha et al. [140] augment their robot with projection indicators to signal the robot's intended path. Szafir et al. [153] equip quad-rotor drones with LEDs mounted in a ring at the base, providing four different signal designs along this strip. They found that their LEDs improve participants' ability to quickly infer the intended motion of the drone. Shrestha et al. [141] perform a study in which a robot crosses a human's path, indicating its intended path with an arrow projected onto the floor. They demonstrate their method to be effective in expressing the robot's intended trajectory. Fernandez et al. [37] introduce the concept of a "passive demonstration", to disambiguate the intention of a robot's LED turn signal. Watanabe et al. [170] evaluate a robotic wheelchair that autonomously navigates the environment with and without intentional communication. They show that passengers and pedestrians found intentional communication intuitive and helpful for passing-by actions.

Robot Gaze as Signal. Several contributions build on the fact that humans infer other people's movement trajectories from their gaze direction [114]^{R1,R2}, and from the relationship between head pose and gaze direction [68]^{R2,R3}. Norman [113]^{R2,R3} speculates that bicycle riders know how to avoid collisions with pedestrians since pedestrian motion can be predicted by gaze. Similarly, Unhelkar et al. [163] found that head pose is a significant predictor of the direction that a person intends to walk.

Following a similar line of thought, Khambhaita et al. [73] propose a motion planner that coordinates head motion to the path a robot will take 4 seconds in the future. In a video survey in which their robot approaches a T-intersection in a hallway, they found that study participants are significantly more able to determine the intended path of the robot in terms of the left or right branch of the intersection when the robot uses the gaze cue as opposed to when it does not. Using a different gaze cue, Lynch et al. [93]^{R1,R2} perform a study in a virtual environment in which virtual agents establish mutual gaze with participants during path-crossing events in a virtual hallway, finding no significant effect in helping participants to disambiguate their paths from those of the virtual agents.

Fiore et al. [39] propose an analysis of human interpretation of social cues in hallway navigation. Their study design included different proxemic and gaze cues that were implemented by rotating the sensors of the robot. Their results show that cues associated with the robot's proxemic behavior were found to significantly affect participant perceptions of the robot's social presence while cues associated with the robot's gaze behavior were not found to be significant. However, Fernandez et al. [37] show that people can adapt to LED-based cues after watching a demonstration of its use, and May et al. [99] present a robot that was able to convey its intention using a mechanical signal but not using a gaze cue. Hart et al. [55] challenge these previous results by providing a different naturalistic gaze cue using a virtual agent head which is added to a mobile robot platform, and compared its performance against a similar robot with an LED turn signal. The results of this work suggest that people can perceive the naturalistic gaze cue and react to it. These conflicting results can be attributed to the vast differences in signal implementation between the different experiments.

Table 6 summarizes the taxonomy values for algorithms that focus on conveying the robot's intention to a human.

4.3 Mediating Conflicts in Navigational Intentions

Karamouzas et al. [70] identify a power-law interaction that is based not on the physical separation between pedestrians but on their projected time to a potential future collision and is therefore fundamentally anticipatory in nature. This finding highlights that there is value in understanding and mediating between the human's navigational goal and the robot's.

Murakami et al. [107] propose to smooth a wheelchair's trajectory to avoid colliding with pedestrians. Kruse et al. [81, 82] investigate classic navigation algorithms that create erratic trajectories near obstacles that make a robot look confused. To address this challenge, they use context-dependent cost functions and directional cost functions that help a robot solve spatial conflicts. One result, for example, is adjusting the robot's velocity instead of its path. Silva and Fraichard [143] tackle the mediation problem using the notion of motion effort and how it should be shared between the robot and the person to avoid collisions. To that end, their approach learns a robot behavior using Reinforcement Learning that enables it to mutually solve the collision avoidance problem during simulated trials. Svenstrup et al. [150] propose a modified RRT for navigation in human environments assuming access to full-state information. The proposed RRT planner plans with a potential field representation of the world, with a potential model designed for moving humans. Alternatively, recent work by Truc et al. [160] focused on drone navigation around people. This work introduced a human-aware 3D reactive planner for drone navigation. This planner is based on stochastic optimization of two criteria: discomfort due to the proximity of the drone to pedestrians, and visibility of the drone.

A different line of research combines social navigation and person following. This combination can work in several directions: both Müller et al. [105], Topp and Christensen [157] present collision avoidance algorithms that are utilized in the context of following one particular person through a populated environment. Alternatively, in Yao et al. [174], the robot leverages the planning of other pedestrians and follows them instead of searching for a solution on its own.

Table 7 summarizes the taxonomy values for mediation algorithms for social navigation discussed in this subsection.

Table 6. An Overview of the Different Intention-conveying Algorithms Used in Social Navigation

Year	Paper	Role	# Agents	Obs.	Motion Control	Com.	Exp. Type	Agent Type
2009	Nummenmaa et al. [114]	I	Abs = 2	Partial	Other (Hand Coded)	D	Sim	H-A
2013	Fiore et al. [39]	I	Abs = 2	Depth	Other (Hand Coded)	I + D	Sim	H-A
2015	May et al. [99]	I	Abs = 2	RGB + Depth	Other (A*)	D	Lab	H-R
2015	Szafir et al. [153]	I	Abs = 2	RGB + Depth	Other (Hand Coded)	D	Lab + Sur	H-R
2015	Unhelkar et al. [163]	N	Abs = 1	Full	Other (SIPP)	N	Sim	Hom
2015	Watanabe et al. [170]	I	Abs = 2	Depth	ROS	D	Lab	H-R
2016	Khambhaita et al. [73]	I	Abs = 2	RGB + Depth	ROS	D	Lab + Sur	H-R
2018	Baraka and Veloso [4]	I	Abs = 2	RGB + Depth	ROS	D	Lab + Sur	H-R
2018	Fernandez et al. [37]	I	Abs = 2	Depth	Other (Hand Coded)	D	Lab	H-R
2018	Lynch et al. [93]	I	Abs = 2	Full	Other (Hand Coded)	D	Sim	H-A
2018	Shrestha et al. [140]	I	Abs = 2	Full	Other (Hand Coded)	D	Lab + Sur	H-R
2020	Hart et al. [55]	I	Abs = 2	Depth	Other (Hand Coded)	D	Lab	H-R

Role refers to the robot role, **Obs.** is observability, **Com.** refers to communication, and **Exp. type** is the experiment type.

5 EVALUATING AN INTERACTION

The numerous different metrics and evaluation methods used in social navigation make apparent the need to standardize them. This section is meant to provide tools and metrics to evaluate new research in social navigation concerning the existing literature and with our proposed taxonomy to provide context for evaluation. As we are surveying an interdisciplinary area, many of the metrics used so far for evaluation were adapted from other research areas (e.g., Human-Computer Interfaces, psychology, physics, mechanical engineering, and more). To pinpoint the most common and useful metrics, we discuss only the metrics that were used in the papers that were presented in the tables in Sections 3 and 4. For each metric we present, we mention the taxonomy attributes that are the most relevant and can directly affect the values of the metric. For example, measuring group formation directly depends on the **Number of Agents** in the environment, since if there is only one pedestrian it cannot form a group. Table 8 summarizes this evaluation according to the different aspects of the interaction: properties of the interaction itself, actions taken by the human or the robot, emergent behaviors, algorithmic properties, and others. This last aspect includes both qualitative evaluation and prediction accuracy, which is a very common metric to estimate the proficiency of *obstacle detection*, a preliminary step before the actual interaction.

5.1 Interaction Properties

This subsection discusses measurements that are related to the nature of the interaction itself, and are meant to evaluate how successful and efficient an interaction is. These metrics are readily apparent to an outside observer, and external to the robot and the human.

Table 7. An Overview of the Different Mediation Algorithms Used in Social Navigation

Year	Paper	Role	# Agents	Obs.	Motion Control	Com.	Exp. Type	Agent Type
2002	Murakami et al. [107]	B	Abs = 2	RGB + Depth	Other (Hand Coded)	I	Lab	H-R
2005	Topp and Christensen [157]	R	Abs = 4	Depth	Other (Person Tracking)	N	Lab	H-R
2008	Müller et al. [105]	R	Abs = 7	Depth	Other (A^* + Person Tracking)	N	Lab	H-R
2010	Svenstrup et al. [150]	R	Abs = 39	Full	Other (Modified RRT)	N	Sim	H-R
2013	Ferrer et al. [38]	R	Abs = 10	Depth	SFM	N	ItW	H-R
2013	Guzzi et al. [52]	B	Abs = 6	RGB	Other (Hand Coded)	N	R	R-R
2014	Karamouzas et al. [70]	R	D = 0.27-2.5	Full	Other (Hand Coded)	N	Sim	Hom
2014	Kruse et al. [82]	B	Abs = 2	Full	Other (Hand Coded)	I	Lab + Sur	H-R
2017	Silva and Fraichard [143]	B	Abs = 2	Full	ROS	N	Sim	Hom
2019	Yao et al. [174]	R	Abs = 6	RGB + Depth	Other (Geometry based)	N	Lab	H-R
2022	Truc et al. [160]	R	Abs = 2	Full	Other (Hand Coded)	N	Sim	Het

Role refers to the robot role, **Obs.** is observability, **Com.** refers to communication, and **Exp. type** is the experiment type.

Conflicts Count is one of the most common approaches to estimating the success of an interaction. This measurement is quantified in several ways: by counting desirable outcomes vs. undesirable outcomes, by counting accidents, or by counting interactions that ended without the robot reaching its goal. In this category, we also consider experiments that counted how many times the robot was required to replan [105] and how many targets it was able to reach in total [52]. This measure is affected by the **Number of Agents**, the **Experiment Type**, and the evaluated **Agent Type**.

Speed is another very common metric used to evaluate an interaction. In general, faster velocities imply that the robot was able to navigate confidently without slowing down. Many researchers used this metric to complement conflict count, to account for cases where a robot may reach its goal quickly but frequently collides with walls. As a reference point, the robot's speed is usually compared to the average pedestrian speed (1.3 ± 0.2 m/s), but this value depends on whether they walk alone or in a group, as group size affects speed more than density level [104]. Gérin-Lajoie et al. [46] measured similar results for natural walking around dynamic obstacles (1.44 ± 0.17 m/s). Accordingly, this measurement is greatly affected by the **Robot's Role** in the interaction, the **Number of Agents**, the **Experiment Type**, and the **Agent Type**.

Path Time is a way to measure the velocity of the robot throughout a full interaction. As the robot might accelerate or decelerate, recording the total time that it took the robot to reach its goal is a simple way to measure its performance. One unique metric that is also relevant to throughput is "social work", defined by Ferrer et al. [38]. This metric measures the total work done by the robot

Table 8. An Overview of the Different Metrics Used in the Surveyed Papers to Evaluate a Social Interaction

Evaluation Type	Metric Evaluated	Relevant Works
Interaction Properties	Conflicts Count	Murakami et al. [107], Pacchierotti et al. [119], Müller et al. [105], Kirby et al. [75],Svenstrup et al. [150], Tamura et al. [155], Diego and Arras [32],Bandyopadhyay et al. [3], Park et al. [122], Ma et al. [94],Guzzi et al. [52], Unhelkar et al. [163], Godoy et al. [48],Okal and Arras [118], Kretzschmar et al. [80], Khambhaita et al. [73], Fernandez et al. [37], Li et al. [85], Everett et al. [36],Ding et al. [33], Long et al. [87], Lynch et al. [93], Jiang et al. [63], Yao et al. [174], Jin et al. [65], Meng et al. [101], Chen et al. [23], Hart et al. [55], Liang et al. [86], Lu et al. [91], Gupta et al. [51]
	Speed	Helbing and Molnar [58], Gérin-Lajoie et al. [46], Karamouzas et al. [69], Moussaïd et al. [103], Kruse et al. [82], Unhelkar et al. [163], Kretzschmar et al. [80], Long et al. [87], Liang et al. [86]
	Path Time	Helbing and Molnar [58], Pacchierotti et al. [119], Karamouzas et al. [69], Foka and Trahanias [40], Bandyopadhyay et al. [3], Ferrer et al. [38], Godoy et al. [48], Chen et al. [24], Chen et al. [25], Tai et al. [154], Everett et al. [36], Ding et al. [33], Long et al. [87], Jiang et al. [63], Jin et al. [65], Kanazawa et al. [67], Chen et al. [23], Liang et al. [86], Lu et al. [91], Gupta et al. [51]
	Path Length	Helbing and Molnar [58], Pacchierotti et al. [119], Karamouzas et al. [69], Henry et al. [59], Luber et al. [92], Rios-Martinez et al. [131], Lu et al. [90], Vasquez et al. [165], Okal and Arras [118], Ding et al. [33], Jiang et al. [63], Long et al. [87], Nardi and Stachniss [111], Liang et al. [86]
	Acceleration	Helbing and Molnar [58], Bonneaud and Warren [11]
	Avoidance Distance	Luber et al. [92], Kruse et al. [82], May et al. [99], Kim and Pineau [74], Kretzschmar et al. [80], Chen et al. [24], Tai et al. [154], Lynch et al. [93], Jin et al. [65], Kanazawa et al. [67], Lu et al. [91]
	Smoothness	Helbing and Molnar [58], Gockley et al. [47], Karamouzas et al. [69], Park et al. [122], Guzzi et al. [52], Vasquez et al. [165], Karamouzas et al. [70], Okal and Arras [118], Truc et al. [160]
Robot/Human Actions	Degrees Turned	Helbing and Molnar [58], Karamouzas et al. [69], Bonneaud and Warren [11], Truong and Ngo [161]
	Gaze Fixations	Nummenmaa et al. [114], Kitazawa and Fujiyama [78]
	Gaze-Movement Angle	Murakami et al. [107], Pacchierotti et al. [119], Müller et al. [105], Kirby et al. [75],Svenstrup et al. [150], Diego and Arras [32], Bandyopadhyay et al. [3], Ratsamee et al. [128], Park et al. [122], Ma et al. [94],Guzzi et al. [52], Unhelkar et al. [163], Godoy et al. [48],Okal and Arras [118], Kretzschmar et al. [80], Khambhaita et al. [73],Fernandez et al. [37], Li et al. [85], Everett et al. [36],Ding et al. [33], Long et al. [87], Lynch et al. [93], Yao et al. [174],Jin et al. [65], Meng et al. [101], Chen et al. [23],Hart et al. [55], Liang et al. [86]
	Head Orientation	Patla et al. [123], Ratsamee et al. [128], Unhelkar et al. [163]
	Body Position	Patla et al. [123], Unhelkar et al. [163]
Emergent Behaviors	Lane Emergence	Helbing and Molnar [58], Bennewitz et al. [7], Loscos et al. [89], Van Den Berg et al. [164], Karamouzas et al. [70]
	Group Formation	Musse and Thalmann [108], Moussaïd et al. [104], Swofford et al. [151]
	Maximal density	Bandyopadhyay et al. [3], Ma et al. [94], Mead and Matarić [100]
Algorithmic Properties	Computation Time	Sisbot et al. [145], Moussaïd et al. [104], Van Den Berg et al. [164], Silva and Fraichard [143], Ding et al. [33]
	Model Prediction	Kuderer et al. [84], Okal and Arras [117], Kim and Pineau [74], Kretzschmar et al. [80], Bera et al. [8], Silva and Fraichard [143], Yao et al. [174], Nardi and Stachniss [111], Meng et al. [101]
User Experience	Attitude	Jeffrey and Mark [62], May et al. [99], Baraka and Veloso [4], Senft et al. [137], Chen et al. [26]
	Acceptance and Social Presence	Gockley et al. [47],Kruse et al. [82], Szafir et al. [153], Watanabe et al. [170], Khambhaita et al. [73], Kretzschmar et al. [80], Chen et al. [26]
	Comfort	Murakami et al. [107], Kruse et al. [82], Vasquez et al. [165], Watanabe et al. [170], Shrestha et al. [140]
	Trust	Chen et al. [26]
No Interaction Evaluation		Reynolds [130], Strassner and Langer [149], Topp and Christensen [157], Ohki et al. [116], Pandey and Alami [120], O'Callaghan et al. [115], Gómez et al. [50], Papadakis et al. [121], Charalampous et al. [21]

and the summation of the work done by each person in the scene. Kanazawa et al. [67] examined the total waiting time that the robot had experienced during the interaction. This measure depends on the **Robot's Role**, the **Number of Agents** in the environment, and the **Experiment Type**.

Path Length provides another perspective about the interaction, and is correlated with speed and path time: by counting any two of these three metrics (Speed, Path Time, and Path Length) one can get a reasonable estimation of the third. As such, this metric is also affected by the same attributes as the other two metrics: the **Robot's Role**, the **Number of Agents**, and the **Experiment Type**.

Acceleration is a way to measure the changes in the robot's behavior throughout the interaction. A robot that accelerates or decelerates several times in an interaction is an indication that it has to replan or adjust to avoid a conflict. This metric is highly affected by the **Robot's Role** and the **Number of Agents**.

Smoothness is a generalization for several metrics that measure the total energy that was put into the interaction by the robot or the human. Successful interactions are expected to require less energy than unsuccessful interactions, which force the robot to replan. Smoothness can be evaluated in several ways, including acceleration/deceleration over time, total kinetic energy used [122], path irregularity (how many unnecessary turns were taken) [52], cumulative heading change [118], and the integral of the square of the curvature to measure the smoothness of a pedestrian's path [69]. This measure is influenced by the **Robot's Role**, the **Observability** that can enable the robot to plan better ahead, and the **Motion Control** used.

Avoidance Distance is a way to measure how close the robot came to a conflict or a full collision with a human. Usually, a robot that can avoid pedestrians from afar is considered more successful than a robot that almost reaches collision [150]. However, this success sometimes creates a tradeoff between the total length of the path the robot needs to take and the smoothness of the path. This metric is affected by the **Robot's Role**, the **Number of Agents**, and the **Motion Control** used that might have its own predefined distance-keeping restrictions.

5.2 Robot/Human Actions

While the previous subsection considered measurements of the interaction as a whole, in this subsection we discuss measures that evaluate the actions taken by the robot or the human.

Degrees Turned As part of an interaction, either the robot or the human (or both) turn to avoid a collision. Evaluation which consists of this measurement usually tracks the degrees of the lane change of either party. This measure will be highly affected by the **Robot's Role** which will determine who will turn, the **Number of Agents** in the environment, and the **Motion Control** used.

Gaze is a general measurement, in which several different aspects can be evaluated, including fixation count and length [114], and the GMA [30]. Kitazawa and Fujiyama [78] investigated gaze patterns in a collision avoidance scenario with multiple pedestrians moving in a wide hallway-shaped area. They show that pedestrians pay much more attention to the ground surface to detect potential immediate environmental hazards than fixating on obstacles. Therefore, most of their fixations fall within a cone-shaped area rather than a semicircle, and the attention paid to approaching pedestrians is not as high as that to static obstacles. Metrics that involve gaze are affected by the **Robot's Role**, **Observability**, **Communication** protocols that the human should be aware of, the **Experiment Type**, and **Agent Type** which can all have great effects on gaze patterns.

Head Orientation and Body Positions are ways to capture some intermediate value between the degrees turned in practice, and the changes in GMA. Recently, Kitagawa et al. [77] leveraged people's reliance on such cues and incorporated similar body rotations into an omnidirectional robot to improve the way pedestrians perceive its performance. These metrics are highly affected by the **Robot's Role** in the interaction, the **Communication** channel used, and the **Agent Type**.

5.3 Emergent Behaviors

Several experiments have been designed to identify specific movement patterns and flow patterns that emerge during the execution of social navigation algorithms or to mimic human movement patterns that emerge in these contexts [7, 89]. In many cases, these patterns are in the form of lanes [58] or group clusters.

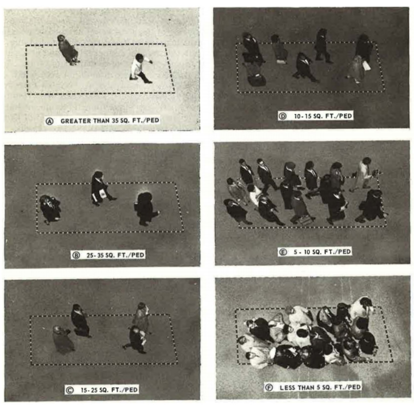

Fig. 5. Levels of Service from A to F: How crowded is the environment (taken from Fruin [42]).

Lane Emergence is a phenomenon that exists in human crowds—whenever an environment becomes crowded enough, people will likely follow the path of others who are going in the same direction [47, 174]. For several algorithms deployed in crowded environments, the researchers were able to detect the emergence of lanes in a robotic navigation context, and considered this behavior as a sign of success, since lanes are usually an efficient way to navigate in crowds. This measure is affected by the **Number of Agents**, the **Experiment Type**, and the evaluated **Agent Type**.

Group Formation is another phenomenon whose appearance implies the success of the interaction. However, unlike lane emergence, group formation is usually an explicit objective of a work that discusses these types of interactions: such work focuses on understanding how groups of pedestrians move together [104], and investigating whether a robot can seamlessly join such a group [108], bypass it [151], or disperse it [26]. This measure is affected by the **Number of Agents** and the **Agent Type**.

Maximal Density is a metric frequently used in simulations to stress-test an agent's ability to navigate in an environment with multiple other agents. When shifting to the real world, Fruin [42]^{R2,R3} proposed six levels of crowdness, which is referred to as *Level of Service*, as depicted in Figure 5. When comparing to human-only navigation, the average density of people in a non-crowded environment was evaluated to 0.03 pedestrians per m^2, and in a moderately crowded environment, there are 0.25 pedestrians per m^2 [104]. Notice that density, or the **Number of Agents** is an attribute in this survey's taxonomy—in this specific section, we only refer to evaluation that uses density as a metric, rather than as a controlled variable.

5.4 Algorithmic Properties

The previous subsections focused on measuring physical quantities, either about the interaction as a whole or about one of the parties. In this subsection, we focus on more algorithmic aspects of the interaction. The metrics presented here can often be measured internally by the robot.

Computation Time in social navigation refers to the robot's processing time. As the robot should perform in real-time, there is a need to evaluate whether the robot can process the required information, plan, and execute its plan on time. Two different components that are measured by computation time are interaction processing, which is usually measured in milliseconds [164], and learning (in data-driven approaches), which is usually measured in learning episodes for achieving a desired behavior [33]. Computation time is influenced by the **Number of Agents**, **Experiment Type**, and **Agent Type**.

Model Prediction is a crucial part of every social navigation interaction: to properly act, the robot should first be able to accurately predict the behavior of other agents in the environment. Some contributions focus solely on improving the part of the interaction that involves understanding the environment given sensor information, and accurately predicting trajectories [84, 101]^{R2,R3}, while others evaluate the prediction of pedestrian trajectories interleaved with robot execution [8, 111]. This metric is influenced by the **Robot's Role** in the interaction, **Observability**, and **Agent Type**.

5.5 User Experience Evaluation

So far, the evaluation metrics discussed can be evaluated using quantitative measures. Some contributions focus on analyzing an interaction and identifying theoretical concepts, thus having no empirical evaluation, while others focus more on user reports (e.g., self-report measures of comfort level) or provide a qualitative evaluation of an interaction. Survey Questions are the most common approach to elicit information from users about how they perceive an interaction with an agent or a robot. These metrics consist of trust levels during the interaction and afterward [135], social presence [5, 6, 54, 166], attitude towards the robot [12, 16, 158, 171], and more. The most relevant attributes that affect the human's reported experience during the interaction are **Robot Role**, **Communication**, and **Agent Type**.

Attitude covers research that evaluates the emotions of the human in the interaction and includes some common surveys that measure attitude, such as the Godspeed series [171] and RoSAS [16]. *Acceptance and Social Presence* refer to the evaluation of the agency of the robot during the interaction, and it is evaluated using the **Perceived Social Intelligence** (**PSI**) Scales [5] and others. *Trust* refers to the extent to which the human trusts the robot to behave socially. *Comfort* refers to the perceived safety and legibility of the robot from the perspective of the human. Torta et al. [158] identify specific values for this comfort zone (182 cm from a sitting person and 173 cm from a standing person). Syrdal et al. [152] present an empirical evaluation of the role of video prototyping and evocation as a good way to evaluate non-functional aspects of HRI. Comfort is also often evaluated through the lens of proxemics [118, 150], which is related to avoidance distance that was discussed earlier, but can encompass additional information about the interaction. For example, Hall [53] identifies different interaction ranges, as manifested in North American cultures: intimate space (up to 0.45m), personal space (1.2m), social space (3.6m), and public space (7.6m). These values are known to change when mapping these distances to HRIs [112]. The comfortable distance from a robot is often considered 0.2m, and arrival tolerance 0.5m, as reported by Kruse et al. [83] and Chen et al. [23].

5.6 Missing or Cursory Evaluation

is a category designated for papers that make only a theoretical contribution, such as classifying different abstract types of interactions [130] or ones that provide a limited qualitative analysis of an interaction [157]. Accordingly, research with no empirical evaluation might be affected by all attributes of the taxonomy, depending on the subject of the analysis.

5.7 Simulations and Resources

So far, this section discussed specific metrics and evaluation methods that have been used in social navigation. One of the goals of this discussion is to promote better comparisons between different contributions in the field. Another way to promote this goal is by using existing simulations or resources that can have a similar baseline. In this subsection, we identify some of the recent efforts to create social navigation benchmarks and evaluation frameworks.

Carton et al. [17] propose a framework for the analysis of human trajectories and show that humans plan their navigation trajectories in a similar fashion when walking past a robot or a human.

Simulations are commonly used to evaluate a social navigation algorithm or model (39 of the 75 surveyed papers used simulations), either as a preliminary step to physical navigation or as a completely independent task. Next, we point out several available simulation tools that can be used to evaluate new contributions. Loscos et al. [89] created a rule-based simulation that can model up to 10, 000 pedestrians in an urban environment. Treuille et al. [159] offered a real-time crowd model based on continuum dynamics, which can facilitate large-scale simulations for navigation. Heigeas et al. [57] presented a simulation platform where pedestrians act according to a physics-based particle force interaction model. Recently, Khambhaita et al. [73] created a simulated benchmark for social navigation tasks instead of physical experiments. This simulation is implemented with OpenAI Gym. Tsoi et al. [162] presents a testing platform that combines ROS and Unity into a social navigation testbed. In this platform's current version, it can measure whether or not the robot reaches its goal, time to goal, collisions with static objects, final distance to the goal, collisions with pedestrians, and closest distance to pedestrians.

For various reasons, there are not many contributions that can be generalized to real-world interactions: First, robots can only be tested under similar conditions, meaning that an evaluation platform for large mobile robots will be different from one for smaller robots. Explicitly identifying how accurate a robotic design is (2D vs. 3D representation, joint movement, 3rd person vs. 1st person evaluation, etc.) is a key component in the design of any real-world robot experiment [152]. In addition, real HRIs require human presence, which introduces a lot of variability and cannot be just compiled into an algorithm that can be used repeatedly.

Mavrogiannis et al. [97] recently published a case study where people and robots navigated in a shared space. The robots used three distinct navigation strategies, executed by a telepresence robot (two autonomous, one teleoperated). The first is ORCA, a local collision-free motion planner for a large number of robots as proposed by Van Den Berg et al. [164] and the second is the **social momentum** (**SM**) planning framework, which estimates the most likely intended avoidance protocols of others based on their past behaviors, superimposes them, and generates an expressive and socially compliant robot action that reinforces the expectations of others regarding these avoidance protocols [98]. These two chosen navigational strategies are agnostic to the fact that the other agent is a human. This assumption leaves an opportunity for further investigation.

6 DISCUSSION

In this survey, we identified specific components that comprise a social navigation interaction and introduced a detailed taxonomy to provide researchers with a framework and a language for comparing and contrasting research in social navigation (Section 2). We then compiled a comprehensive list of papers that contribute to social navigation and discussed them according to their values given our taxonomy (Sections 3 and 4). Next, we surveyed the different measurements used to evaluate interaction in this context and highlighted the relations between these measurements and the taxonomy attributes (Section 5).

Social navigation is a growing research area. While we expect that the attributes we chose for the taxonomy will remain relevant in the years to come, additional attributes will be added and the focus of specific work might shift to deal with new settings. However, any progress in the field must be rooted in the fundamental components of social navigation as they are presented in this survey. In addition, the proposed taxonomy can serve as a framework that enables researchers to properly place their contributions with respect to other work and to provide better benchmarks, which we hope will lead to additional growth in this research area.

To conclude this survey and to consolidate its contributions into a coherent guide, we offer the readers the following checklist to assist with the design of social navigation interaction between a human and a robot. When introducing a new contribution to social navigation, potential aspects to consider include the following.

(1) **Taxonomy:** Identify the values your work has with respect to the taxonomy's attributes in this survey: Robot's Role, Number of Agents, Observability, Motion Control, Communication, Experiment Type, and Agent Type. As shown in this survey, the values of these attributes differ greatly among different papers; thus using this taxonomy is expected to help place new contributions within useful contexts and scopes.
(2) **Reliability:** Provide as many details as possible about the choices made in the design of the robot, and about the implementation details. For example, when reporting an absolute number of pedestrians, also report the size of the area in which the experiment was conducted.
(3) **Human Presence:** If your work consists of interaction with pedestrians, what is their level of familiarity with the robot before the interaction? As presented in this survey, often experiments with human subjects are conducted in the lab rather than in the wild, where the subjects are often the roboticists who designed the robot.
(4) **Context:** Identify what is exactly the context in which the interaction takes place. As with other decisions, the context in which the chosen design is utilized can affect the behavior of pedestrians.
(5) **Success:** If your work consists of empirical evaluation, identify in advance what is considered a success in an interaction. For example, if your work introduces a new indirect communication method, the success of the evaluation should properly isolate the effect of that method.
(6) **Evaluation:** Detail which metrics will be used to evaluate this success, and what values are these metrics expected to have. Based on the presented taxonomy and surveyed papers, evaluation can be placed in comparison to other existing work.

While the presented taxonomy and the above checklist can be useful resources, in Section 2 we mentioned some additional concepts that are not yet mature enough to be included in the taxonomy but might become more significant as the field grows. These concepts include an analysis of different collision types, context awareness, and semantic mapping, reactions to a robot vs. to a human, social cues and social signals, focused interaction, and navigating with groups of pedestrians. We see a surge of work that breaks traditional assumptions about pedestrian behavior in the context of social navigation [26, 106, 129], and these new settings may not be reflected using the existing attributes of the social navigation taxonomy. These papers are part of a fast-evolving field, in which we predict immense growth in the next decade. It is hence a good time to gather and map the knowledge that was already acquired, so it will also be easier to identify the differences when attacking new problem domains.

There are numerous open problems related to social navigation, given our current understanding and technological abilities: standardization of evaluation metrics and domains, context-aware navigation (workday vs. weekend), group understanding (avoid collision with a group participant), and adaptive navigation via machine learning (lifelong learning). Each problem offers many opportunities to leverage recent advances in machine learning, robotics, and HRIs and implement them in a social navigation context. For those interested in contributing to this research area, the above problems should serve as a promising starting point. More information about these problems can be found in Section 2.2.

Additionally, given the evaluation review provided in Section 5, we take a broad perspective of the field to highlight two existing gaps that provide an opportunity for researchers to make new, significant contributions to social navigation. First, most papers on social navigation focus on the quantifiable and technologically-focused aspects of navigation rather than on the experience of the human in the interaction. This gap calls for researchers to add a user-experience component to planned studies. Second, there is a consensus in social navigation papers that (perceived) conflicts are undesired. However, what precisely makes an interaction a conflict remains open, with different studies using a variety of metrics to account for collisions and near-collisions. The introduction of a set of clear, agreed-upon criteria that comprise a conflict would constitute a big advance for the social navigation community.

To conclude, we expect the field of social navigation to gain increased popularity and lead to more real-world applications during the next decade. This survey aims at helping lay the groundwork for these exciting developments by mapping existing approaches onto a novel taxonomy and providing a context for new contributions to social navigation.

REFERENCES

[1] Henny Admoni, Caroline Bank, Joshua Tan, Mariya Toneva, and Brian Scassellati. 2011. Robot gaze does not reflexively cue human attention. In *Proceedings of the Annual Meeting of the Cognitive Science Society*.

[2] Henny Admoni and Brian Scassellati. 2017. Social eye gaze in human-robot interaction: A review. *Journal of Human-Robot Interaction* 6, 1 (2017), 25–63.

[3] Tirthankar Bandyopadhyay, Kok Sung Won, Emilio Frazzoli, David Hsu, Wee Sun Lee, and Daniela Rus. 2013. Intention-aware motion planning. In *Proceedings of the Algorithmic Foundations of Robotics X*. Springer, 475–491.

[4] Kim Baraka and Manuela M. Veloso. 2018. Mobile service robot state revealing through expressive lights: Formalism, design, and evaluation. *International Journal of Social Robotics* 10, 1 (2018), 65–92. DOI: https://doi.org/10.1007/s12369-017-0431-x

[5] Kimberly A. Barchard, Leiszle Lapping-Carr, R. Shane Westfall, Andrea Fink-Armold, Santosh Balajee Banisetty, and David Feil-Seifer. 2020. Measuring the perceived social intelligence of robots. *ACM Transactions on Human-Robot Interaction* 9, 4 (2020), 1–29.

[6] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics* 1, 1 (2009), 71–81.

[7] Maren Bennewitz, Wolfram Burgard, and Sebastian Thrun. 2002. Learning motion patterns of persons for mobile service robots. In *Proceedings of the 2002 IEEE International Conference on Robotics and Automation (Cat. No. 02CH37292)*. IEEE, 3601–3606.

[8] Aniket Bera, Tanmay Randhavane, Rohan Prinja, and Dinesh Manocha. 2017. Sociosense: Robot navigation amongst pedestrians with social and psychological constraints. In *Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 7018–7025.

[9] Eric Bonabeau, Marco Dorigo, and Guy Theraulaz. 1999. *Swarm Intelligence: from Natural to Artificial Systems*. Number 1. Oxford University Press.

[10] Francisco Bonin-Font, Alberto Ortiz, and Gabriel Oliver. 2008. Visual navigation for mobile robots: A survey. *Journal of Intelligent and Robotic Systems* 53, 3 (2008), 263–296.

[11] Stephane Bonneaud and William H. Warren. 2014. An empirically-grounded emergent approach to modeling pedestrian behavior. In *Proceedings of the Pedestrian and Evacuation Dynamics 2012*. Springer, 625–638.

[12] Margaret M. Bradley and Peter J. Lang. 1994. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry* 25, 1 (1994), 49–59.

[13] John Travis Butler and Arvin Agah. 2001. Psychological effects of behavior patterns of a mobile personal robot. *Autonomous Robots* 10, 2 (2001), 185–202.

[14] Shiwei Lin, Ang Liu, Jianguo Wang, and Xiaoying Kong. 2022. A review of path-planning approaches for multiple mobile robots. *Machines* 10, 9 (2022), 773.

[15] Frederik Schewe and Mark Vollrath. 2023. Ecological interface design and head-up displays: The contact-analog visualization tradeoff. *Human Factors* 65, 1 (2023), 37–49.

[16] Colleen M. Carpinella, Alisa B. Wyman, Michael A. Perez, and Steven J. Stroessner. 2017. The robotic social attributes scale (RoSAS) development and validation. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-robot Interaction*. 254–262.

[17] Daniel Carton, Wiktor Olszowy, and Dirk Wollherr. 2016. Measuring the effectiveness of readability for mobile robot locomotion. *International Journal of Social Robotics* 8, 5 (2016), 721–741.

[18] Anthony R. Cassandra, Leslie Pack Kaelbling, and James A. Kurien. 1996. Acting under uncertainty: Discrete bayesian models for mobile-robot navigation. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems IROS'96*. IEEE, 963–972.

[19] Umberto Castiello. 2003. Understanding other people's actions: Intention and attention. *Journal of Experimental Psychology: Human Perception and Performance* 29, 2 (2003), 416.

[20] Elizabeth Cha, Yunkyung Kim, Terrence Fong, and Maja J. Mataric. 2018. A survey of nonverbal signaling methods for non-humanoid robots. *Foundations and Trends® in Robotics* 6, 4 (2018), 211–323.

[21] Konstantinos Charalampous, Christos Emmanouilidis, and Antonios Gasteratos. 2014. Social mapping on RGB-D scenes. In *Proceedings of the 2014 IEEE International Conference on Imaging Systems and Techniques (IST)*. IEEE, 398–403.

[22] Konstantinos Charalampous, Ioannis Kostavelis, and Antonios Gasteratos. 2017. Recent trends in social aware robot navigation: A survey. *Robotics and Autonomous Systems* 93 (2017), 85–104.

[23] Changan Chen, Yuejiang Liu, Sven Kreiss, and Alexandre Alahi. 2019. Crowd-robot interaction: Crowd-aware robot navigation with attention-based deep reinforcement learning. In *Proceedings of the 2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 6015–6022.

[24] Yu Fan Chen, Michael Everett, Miao Liu, and Jonathan P. How. 2017. Socially aware motion planning with deep reinforcement learning. In *Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 1343–1350.

[25] Yu Fan Chen, Miao Liu, Michael Everett, and Jonathan P. How. 2017. Decentralized non-communicating multiagent collision avoidance with deep reinforcement learning. In *Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 285–292.

[26] Zhiming Chen, Tingxiang Fan, Xuan Zhao, Jing Liang, Cong Shen, Hua Chen, Dinesh Manocha, Jia Pan, and Wei Zhang. 2021. Autonomous social distancing in urban environments using a quadruped robot. *IEEE Access* 9 (2021), 8392–8403.

[27] James S. Coleman and John James. 1961. The equilibrium size distribution of freely-forming groups. *Sociometry* 24, 1 (1961), 36–45.

[28] Jonathan Crespo, Jose Carlos Castillo, Oscar Martinez Mozos, and Ramon Barber. 2020. Semantic information for robot navigation: A survey. *Applied Sciences* 10, 2 (2020), 497.

[29] Yuchen Cui, Qiping Zhang, Brad Knox, Alessandro Allievi, Peter Stone, and Scott Niekum. 2021. The empathic framework for task learning from implicit human feedback. In *Conference on Robot Learning*, PMLR, 604–626.

[30] James E. Cutting, Peter M. Vishton, and Paul A. Braren. 1995. How we avoid collisions with stationary and moving objects. *Psychological Review* 102, 4 (1995), 627.

[31] Guilherme N. DeSouza and Avinash C. Kak. 2002. Vision for mobile robot navigation: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24, 2 (2002), 237–267.

[32] Gian Diego and Tipaldi Kai O. Arras. 2011. Please do not disturb! minimum interference coverage for social robots. In *Proceedings of the 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 1968–1973.

[33] Wenhao Ding, Shuaijun Li, Huihuan Qian, and Yongquan Chen. 2018. Hierarchical reinforcement learning framework towards multi-agent navigation. In *Proceedings of the 2018 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, 237–242.

[34] Anca D. Dragan, Kenton C. T. Lee, and Siddhartha S. Srinivasa. 2013. Legibility and predictability of robot motion. In *Proceedings of the 2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 301–308.

[35] Andreas Ess, Bastian Leibe, Konrad Schindler, and Luc Van Gool. 2009. Moving obstacle detection in highly dynamic scenes. In *Proceedings of the 2009 IEEE International Conference on Robotics and Automation*. IEEE, 56–63.

[36] Michael Everett, Yu Fan Chen, and Jonathan P. How. 2018. Motion planning among dynamic, decision-making agents with deep reinforcement learning. In *Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 3052–3059.

[37] Rolando Fernandez, Nathan John, Sean Kirmani, Justin Hart, Jivko Sinapov, and Peter Stone. 2018. Passive demonstrations of light-based robot signals for improved human interpretability. In *Proceedings of the 2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 234–239.

[38] Gonzalo Ferrer, Anais Garrell, and Alberto Sanfeliu. 2013. Social-aware robot navigation in urban environments. In *Proceedings of the 2013 European Conference on Mobile Robots*. IEEE, 331–336.

[39] Stephen M. Fiore, Travis J. Wiltshire, Emilio J. C. Lobato, Florian G. Jentsch, Wesley H. Huang, and Benjamin Axelrod. 2013. Toward understanding social cues and signals in human.robot interaction: effects of robot gaze and proxemic behavior. *Frontiers in Psychology* 4 (2013), 859.

[40] Amalia F. Foka and Panos E. Trahanias. 2010. Probabilistic autonomous robot navigation in dynamic environments with human motion prediction. *International Journal of Social Robotics* 2, 1 (2010), 79–94.

[41] Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. 2003. A survey of socially interactive robots. *Robotics and Autonomous Systems* 42, 3-4 (2003), 143–166.

[42] John J. Fruin. 1971. *Designing for Pedestrians: A Level-of-service Concept*. Number HS-011 999.

[43] Heba Gaber, Mohamed Marey, Safaa Amin, and Mohamed F. Tolba. 2017. Localization and mapping for indoor navigation: Survey. *Handbook of Research on Machine Learning Innovations and Trends* (2017), 136–160.
[44] Yuxiang Gao and Chien-Ming Huang. 2022. Evaluation of socially-aware robot navigation. *Frontiers in Robotics and AI* 8 (2022), 721317.
[45] Sourav Garg, Niko Sünderhauf, Feras Dayoub, Douglas Morrison, Akansel Cosgun, Gustavo Carneiro, Qi Wu, Tat-Jun Chin, Ian Reid, Stephen Gould, Peter Corke, and Michael Milford. 2020. Semantics for robotic mapping, perception and interaction: A survey. *Foundations and Trends® in Robotics* 8, 1–2 (2020), 1–224.
[46] Martin Gérin-Lajoie, Carol L. Richards, Joyce Fung, and Bradford J. McFadyen. 2008. Characteristics of personal space during obstacle circumvention in physical and virtual environments. *Gait and Posture* 27, 2 (2008), 239–247.
[47] Rachel Gockley, Jodi Forlizzi, and Reid Simmons. 2007. Natural person-following behavior for social robots. In *Proceedings of the ACM/IEEE International Conference on Human-robot Interaction*. 17–24.
[48] Julio Godoy, Ioannis Karamouzas, Stephen J. Guy, and Maria L. Gini. 2016. Moving in a crowd: Safe and efficient navigation among heterogeneous agents.. In *Proceedings of the IJCAI*. 294–300.
[49] Erving Goffman. 2008. *Behavior in Public Places*. Simon and Schuster.
[50] Javier V. Gómez, Nikolaos Mavridis, and Santiago Garrido. 2014. Fast marching solution for the social path planning problem. In *Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 1871–1876.
[51] Himanshu Gupta, Bradley Hayes, and Zachary Sunberg. 2022. Intention-aware navigation in crowds with extended-space POMDP planning. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*. 562–570.
[52] Jérôme Guzzi, Alessandro Giusti, Luca M. Gambardella, Guy Theraulaz, and Gianni A. Di Caro. 2013. Human-friendly robot navigation in dynamic environments. In *Proceedings of the 2013 IEEE International Conference on Robotics and Automation*. IEEE, 423–430.
[53] Edward Twitchell Hall. 1966. *The Hidden Dimension*. Vol. 609. Garden City, NY: Doubleday.
[54] Chad Harms and Frank Biocca. 2004. Internal consistency and reliability of the networked minds measure of social presence. In *Proceedings of the 7th Annual International Workshop: Presence*. Universidad Politecnica de Valencia Valencia, Spain.
[55] Justin Hart, Reuth Mirsky, Xuesu Xiao, Stone Tejeda, Bonny Mahajan, Jamin Goo, Kathryn Baldauf, Sydney Owen, and Peter Stone. 2020. Using human-inspired signals to disambiguate navigational intentions. In *Proceedings of the International Conference on Social Robotics*. Springer, 320–331.
[56] Leslie A. Hayduk. 1981. The shape of personal space: An experimental investigation. *Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences du Comportement* 13, 1 (1981), 87.
[57] Laure Heigeas, Annie Luciani, Joëlle Thollot, and Nicolas Castagné. 2003. A physically-based particle model of emergent crowd behaviors. In *Graphicon 2003-13th International Conference on Computer Graphics*, 1–9.
[58] Dirk Helbing and Peter Molnar. 1995. Social force model for pedestrian dynamics. *Physical Review E* 51, 5 (1995), 4282.
[59] Peter Henry, Christian Vollmer, Brian Ferris, and Dieter Fox. 2010. Learning to navigate through crowded environments. In *Proceedings of the 2010 IEEE International Conference on Robotics and Automation*. IEEE, 981–986.
[60] Blake Holman, Abrar Anwar, Akash Singh, Mauricio Tec, Justin Hart, and Peter Stone. 2021. Watch where you're going! gaze and head orientation as predictors for social robot navigation. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 6183–6190.
[61] Serge Hoogendoorn and Piet Bovy. 2003. Simulation of pedestrian flows by optimal control and differential games. *Optimal Control Applications and Methods* 24, 3 (2003), 153–172.
[62] Phillip Jeffrey and Gloria Mark. 1998. Constructing social spaces in virtual environments: A study of navigation and interaction. In *Proceedings of the Workshop on Personalised and Social Navigation in Information Space*. Stockholm: Swedish Institute of Computer Science, 24–38.
[63] Yu-Sian Jiang, Garrett Warnell, Eduardo Munera, and Peter Stone. 2018. A study of human-robot copilot systems for en-route destination changing. In *Proceedings of the 27th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN2018)*. Nanjing, China. Retrieved from http://www.cs.utexas.edu/users/ai-lab/?ROMAN18-Jiang
[64] Yu-Sian Jiang, Garrett Warnell, and Peter Stone. 2018. Inferring user intention using gaze in vehicles. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI)*. Boulder, Colorado. Retrieved from http://www.cs.utexas.edu/users/ai-lab/?ICMI18-Jiang
[65] Jun Jin, Nhat M. Nguyen, Nazmus Sakib, Daniel Graves, Hengshuai Yao, and Martin Jagersand. 2020. Mapless navigation among dynamics with social-safety-awareness: a reinforcement learning approach from 2d laser scans. In *IEEE International Conference on Robotics and Automation (ICRA'20)*, IEEE, 6979–6985.
[66] Leslie Pack Kaelbling. 2020. The foundation of efficient robot learning. *Science* 369, 6506 (2020), 915–916.

[67] Akira Kanazawa, Jun Kinugawa, and Kazuhiro Kosuge. 2019. Adaptive motion planning for a collaborative robot based on prediction uncertainty to enhance human safety and work efficiency. *IEEE Transactions on Robotics* 35, 4 (2019), 817–832.

[68] Anuradha Kar and Peter Corcoran. 2017. A review and analysis of eye-gaze estimation systems, algorithms and performance evaluation methods in consumer platforms. *IEEE Access* 5 (2017), 16495–16519.

[69] Ioannis Karamouzas, Peter Heil, Pascal Van Beek, and Mark H. Overmars. 2009. A predictive collision avoidance model for pedestrian simulation. In *Proceedings of the International Workshop on Motion in Games*. Springer, 41–52.

[70] Ioannis Karamouzas, Brian Skinner, and Stephen J. Guy. 2014. Universal power law governing pedestrian interactions. *Physical Review Letters* 113, 23 (2014), 238701.

[71] Haresh Karnan, Anirudh Nair, Xuesu Xiao, Garrett Warnell, Soeren Pirk, Alexander Toshev, Justin Hart, Joydeep Biswas, and Peter Stone. 2022. Socially compliant navigation dataset (SCAND): A large-scale dataset of demonstrations for social navigation. *IEEE Robotics and Automation Letters* 7, 4 (2022), 11807–11814.

[72] Deneth Karunarathne, Yoichi Morales, Takayuki Kanda, and Hiroshi Ishiguro. 2018. Model of side-by-side walking without the robot knowing the goal. *International Journal of Social Robotics* 10, 4 (2018), 401–420.

[73] Harmish Khambhaita, Jorge Rios-Martinez, and Rachid Alami. 2016. Head-body motion coordination for human aware robot navigation. In *Proceedings of the 9th International Workshop on Human-Friendly Robotics (HFR 2016)*. 8p.

[74] Beomjoon Kim and Joelle Pineau. 2016. Socially adaptive path planning in human environments using inverse reinforcement learning. *International Journal of Social Robotics* 8, 1 (2016), 51–66.

[75] Rachel Kirby, Reid Simmons, and Jodi Forlizzi. 2009. Companion: A constraint-optimizing method for person-acceptable navigation. In *Proceedings of the RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 607–612.

[76] Dmitry Mark Kit. 2012. *Change Detection Models for Mobile Cameras*. Ph.D. Dissertation.

[77] Ryo Kitagawa, Yuyi Liu, and Takayuki Kanda. 2021. Human-inspired motion planning for omni-directional social robots. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21)*. Association for Computing Machinery, New York, NY, USA, 34–42. DOI: https://doi.org/10.1145/3434073.3444679

[78] Kay Kitazawa and Taku Fujiyama. 2010. Pedestrian vision and collision avoidance behavior: Investigation of the information process space of pedestrians using an eye tracker. In *Proceedings of the Pedestrian and Evacuation Dynamics 2008*. Springer, 95–108.

[79] Ioannis Kostavelis and Antonios Gasteratos. 2015. Semantic mapping for mobile robotics tasks: A survey. *Robotics and Autonomous Systems* 66 (2015), 86–103.

[80] Henrik Kretzschmar, Markus Spies, Christoph Sprunk, and Wolfram Burgard. 2016. Socially compliant mobile robot navigation via inverse reinforcement learning. *The International Journal of Robotics Research* 35, 11 (2016), 1289–1307.

[81] Thibault Kruse, Patrizia Basili, Stefan Glasauer, and Alexandra Kirsch. 2012. Legible robot navigation in the proximity of moving humans. In *Proceedings of the 2012 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO)*. IEEE, 83–88.

[82] Thibault Kruse, Alexandra Kirsch, Harmish Khambhaita, and Rachid Alami. 2014. Evaluating directional cost models in navigation. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction*. 350–357.

[83] Thibault Kruse, Amit Kumar Pandey, Rachid Alami, and Alexandra Kirsch. 2013. Human-aware robot navigation: A survey. *Robotics and Autonomous Systems* 61, 12 (2013), 1726–1743.

[84] Markus Kuderer, Henrik Kretzschmar, Christoph Sprunk, and Wolfram Burgard. 2012. Feature-based prediction of trajectories for socially compliant navigation. In *Proceedings of the Robotics: Science and Systems*.

[85] Mingming Li, Rui Jiang, Shuzhi Sam Ge, and Tong Heng Lee. 2018. Role playing learning for socially concomitant mobile robot navigation. *CAAI Transactions on Intelligence Technology* 3, 1 (2018), 49–58.

[86] Jing Liang, Utsav Patel, Adarsh Jagan Sathyamoorthy, and Dinesh Manocha. 2021. Crowd-steer: Realtime smooth and collision-free robot navigation in densely crowded scenarios trained using high-fidelity simulation. In *Proceedings of the 29th International Conference on International Joint Conferences on Artificial Intelligence*, 4221–4228.

[87] Pinxin Long, Tingxiang Fanl, Xinyi Liao, Wenxi Liu, Hao Zhang, and Jia Pan. 2018. Towards optimally decentralized multi-robot collision avoidance via deep reinforcement learning. In *Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 6252–6259.

[88] Antonio M. López, Juan C. Alvarez, and Diego Álvarez. 2019. Walking turn prediction from upper body kinematics: A systematic review with implications for human-robot interaction. *Applied Sciences* 9, 3 (2019), 361.

[89] Celine Loscos, David Marchal, and Alexandre Meyer. 2003. Intuitive crowd behavior in dense urban environments using local laws. In *Proceedings of the Theory and Practice of Computer Graphics, 2003*. IEEE, 122–129.

[90] David V. Lu, Daniel B. Allan, and William D. Smart. 2013. Tuning cost functions for social navigation. In *Proceedings of the International Conference on Social Robotics*. Springer, 442–451.

[91] Xiaojun Lu, Hanwool Woo, Angela Faragasso, Atsushi Yamashita, and Hajime Asama. 2022. Socially aware robot navigation in crowds via deep reinforcement learning with resilient reward functions. *Advanced Robotics* 36, 8 (2022), 388–403.

[92] Matthias Luber, Luciano Spinello, Jens Silva, and Kai O. Arras. 2012. Socially-aware robot navigation: A learning approach. In *Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 902–907.

[93] Sean D. Lynch, Julien Pettré, Julien Bruneau, Richard Kulpa, Armel Crétual, and Anne-Héléne Olivier. 2018. Effect of virtual human gaze behaviour during an orthogonal collision avoidance walking task. In *Proceedings of the 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 136–142.

[94] J. Ma, Siu Ming Lo, W. G. Song, W. L. Wang, J. Zhang, and G. X. Liao. 2013. Modeling pedestrian space in complex building for efficient pedestrian traffic simulation. *Automation in Construction* 30 (2013), 25–36.

[95] Gonçalo S. Martins, Rui P. Rocha, Fernando J. Pais, and Paulo Menezes. 2019. Clusternav: Learning-based robust navigation operating in cluttered environments. In *Proceedings of the 2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 9624–9630.

[96] Christoforos Mavrogiannis, Francesca Baldini, Allan Wang, Dapeng Zhao, Pete Trautman, Aaron Steinfeld, and Jean Oh. 2023. Core challenges of social robot navigation: A survey. *ACM Transactions on Human-Robot Interaction* 12, 3 (2023), 1–39.

[97] Christoforos Mavrogiannis, Alena M. Hutchinson, John Macdonald, Patrícia Alves-Oliveira, and Ross A. Knepper. 2019. Effects of distinct robot navigation strategies on human behavior in a crowded environment. In *Proceedings of the 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 421–430.

[98] 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 2018 ACM/IEEE International Conference on Human-Robot Interaction*. 361–369.

[99] Alyxander David May, Christian Dondrup, and Marc Hanheide. 2015. Show me your moves! Conveying navigation intention of a mobile robot to humans. In *Proceedings of the 2015 European Conference on Mobile Robots (ECMR)*. IEEE, 1–6.

[100] Ross Mead and Maja J. Matarić. 2016. Perceptual models of human-robot proxemics. In *Proceedings of the Experimental Robotics*. Springer, 261–276.

[101] Xiangyun Meng, Nathan Ratliff, Yu Xiang, and Dieter Fox. 2020. Scaling local control to large-scale topological navigation. In *IEEE International Conference on Robotics and Automation (ICRA'20)*, IEEE, 672–678.

[102] Ronja Möller, Antonino Furnari, Sebastiano Battiato, Aki Härmä, and Giovanni Maria Farinella. 2021. A survey on human-aware robot navigation. *Robotics and Autonomous Systems* 145 (2021), 103837.

[103] Mehdi Moussaïd, Dirk Helbing, and Guy Theraulaz. 2011. How simple rules determine pedestrian behavior and crowd disasters. *Proceedings of the National Academy of Sciences* 108, 17 (2011), 6884–6888.

[104] Mehdi Moussaïd, Niriaska Perozo, Simon Garnier, Dirk Helbing, and Guy Theraulaz. 2010. The walking behaviour of pedestrian social groups and its impact on crowd dynamics. *PloS One* 5, 4 (2010), e10047.

[105] Jörg Müller, Cyrill Stachniss, Kai O. Arras, and Wolfram Burgard. 2008. Socially inspired motion planning for mobile robots in populated environments. In *Proceedings of the International Conference on Cognitive Systems*.

[106] Hisashi Murakami, Claudio Feliciani, Yuta Nishiyama, and Katsuhiro Nishinari. 2021. Mutual anticipation can contribute to self-organization in human crowds. *Science Advances* 7, 12 (2021), eabe7758.

[107] Yoshifumi Murakami, Yoshinori Kuno, Nobutaka Shimada, and Yoshiaki Shirai. 2002. Collision avoidance by observing pedestrians' faces for intelligent wheelchairs. *Journal of the Robotics Society of Japan* 20, 2 (2002), 206–213.

[108] Soraia Raupp Musse and Daniel Thalmann. 1997. A model of human crowd behavior: Group inter-relationship and collision detection analysis. In *Proceedings of the Computer Animation and Simulation'97*. Springer, 39–51.

[109] Yasushi Nakauchi and Reid Simmons. 2002. A social robot that stands in line. *Autonomous Robots* 12, 3 (2002), 313–324.

[110] Venkatraman Narayanan, Mike Phillips, and Maxim Likhachev. 2012. Anytime safe interval path planning for dynamic environments. In *Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 4708–4715.

[111] Lorenzo Nardi and Cyrill Stachniss. 2020. Long-term robot navigation in indoor environments estimating patterns in traversability changes. In *IEEE International Conference on Robotics and Automation (ICRA'20)*, IEEE, 300–306.

[112] Margot M. E. Neggers, Raymond H. Cuijpers, Peter A. M. Ruijten, and Wijnand A. IJsselsteijn. 2022. Determining shape and size of personal space of a human when passed by a robot. *International Journal of Social Robotics* (2022), 1–12.

[113] Don Norman. 2009. *The Design of Future Things*. Basic books.

[114] Lauri Nummenmaa, Jukka Hyönä, and Jari K. Hietanen. 2009. I'll walk this way: Eyes reveal the direction of locomotion and make passersby look and go the other way. *Psychological Science* 20, 12 (2009), 1454–1458.

[115] Simon T. O'Callaghan, Surya P. N. Singh, Alen Alempijevic, and Fabio T. Ramos. 2011. Learning navigational maps by observing human motion patterns. In *Proceedings of the 2011 IEEE International Conference on Robotics and Automation*. IEEE, 4333–4340.

[116] Takeshi Ohki, Keiji Nagatani, and Kazuya Yoshida. 2010. Collision avoidance method for mobile robot considering motion and personal spaces of evacuees. In *Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 1819–1824.

[117] Billy Okal and Kai O. Arras. 2014. Towards group-level social activity recognition for mobile robots. In *Proceedings of the IROS Assistance and Service Robotics in a Human Environments Workshop*.

[118] Billy Okal and Kai O. Arras. 2016. Learning socially normative robot navigation behaviors with bayesian inverse reinforcement learning. In *Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2889–2895.

[119] Elena Pacchierotti, Henrik I. Christensen, and Patric Jensfelt. 2006. Design of an office-guide robot for social interaction studies. In *Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 4965–4970.

[120] Amit Kumar Pandey and Rachid Alami. 2010. A framework towards a socially aware mobile robot motion in human-centered dynamic environment. In *Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 5855–5860.

[121] Panagiotis Papadakis, Patrick Rives, and Anne Spalanzani. 2014. Adaptive spacing in human-robot interactions. In *Proceedings of the 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2627–2632.

[122] Jin Hyoung Park, Francisco Arturo Rojas, and Hyun Seung Yang. 2013. A collision avoidance behavior model for crowd simulation based on psychological findings. *Computer Animation and Virtual Worlds* 24, 3-4 (2013), 173–183.

[123] A. E. Patla, A. Adkin, and T. Ballard. 1999. Online steering: Coordination and control of body center of mass, head and body reorientation. *Experimental Brain Research* 129, 4 (1999), 629–634. DOI: https://doi.org/10.1007/s002210050932

[124] Mark Pfeiffer, Ulrich Schwesinger, Hannes Sommer, Enric Galceran, and Roland Siegwart. 2016. Predicting actions to act predictably: Cooperative partial motion planning with maximum entropy models. In *Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2096–2101.

[125] Sören Pirk, Edward Lee, Xuesu Xiao, Leila Takayama, Anthony Francis, and Alexander Toshev. 2022. A protocol for validating social navigation policies. arXiv:2204.05443. Retrieved from https://arxiv.org/abs/2204.05443

[126] Daniel J. Povinelli, Donna T. Bierschwale, and Claude G. Cech. 1999. Comprehension of seeing as a referential act in young children, but not juvenile chimpanzees. *British Journal of Developmental Psychology* 17, 1 (1999), 37–60.

[127] Erwin Prassler, Dirk Bank, Boris Kluge, and M. Hagele. 2002. Key technologies in robot assistants: Motion coordination between a human and a mobile robot. *Transactions on Control, Automation and Systems Engineering* 4, 1 (2002), 56–61.

[128] Photchara Ratsamee, Yasushi Mae, Kenichi Ohara, Tomohito Takubo, and Tatsuo Arai. 2013. Human–robot collision avoidance using a modified social force model with body pose and face orientation. *International Journal of Humanoid Robotics* 10, 01 (2013), 1350008.

[129] Samantha Reig, Michal Luria, Janet Z. Wang, Danielle Oltman, Elizabeth Jeanne Carter, Aaron Steinfeld, Jodi Forlizzi, and John Zimmerman. 2020. Not some random agent: Multi-person interaction with a personalizing service robot. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. 289–297.

[130] Craig W. Reynolds. 1999. Steering behaviors for autonomous characters. In *Proceedings of the Game Developers Conference*. Citeseer, 763–782.

[131] Jorge Rios-Martinez, Alessandro Renzaglia, Anne Spalanzani, Agostino Martinelli, and Christian Laugier. 2012. Navigating between people: A stochastic optimization approach. In *Proceedings of the 2012 IEEE International Conference on Robotics and Automation*. IEEE, 2880–2885.

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

[133] Akanksha Saran, Srinjoy Majumdar, Elaine Schaertl Short, Andrea Thomaz, and Scott Niekum. 2018. Human gaze following for human-robot interaction. In *Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 8615–8621.

[134] Akanksha Saran, Elaine Schaertl Short, Andrea Thomaz, and Scott Niekum. 2020. Understanding teacher gaze patterns for robot learning. In *Proceedings of the Conference on Robot Learning*. PMLR, 1247–1258.

[135] Kristin E. Schaefer. 2016. Measuring trust in human robot interactions: Development of the "trust perception scale-HRI". In *Proceedings of the Robust Intelligence and Trust in Autonomous Systems*. Springer, 191–218.

[136] Alessandra Sciutti, Ambra Bisio, Francesco Nori, Giorgio Metta, Luciano Fadiga, Thierry Pozzo, and Giulio Sandini. 2012. Measuring human-robot interaction through motor resonance. *International Journal of Social Robotics* 4, 3 (2012), 223–234.

[137] Emmanuel Senft, Satoru Satake, and Takayuki Kanda. 2020. Would you mind me if I pass by you? Socially-appropriate behaviour for an omni-based social robot in narrow environment. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. 539–547.

[138] Chao Shi, Satoru Satake, Takayuki Kanda, and Hiroshi Ishiguro. 2018. A robot that distributes flyers to pedestrians in a shopping mall. *International Journal of Social Robotics* 10, 4 (2018), 421–437.

[139] Rathin Chandra Shit. 2020. Precise localization for achieving next-generation autonomous navigation: State-of-the-art, taxonomy and future prospects. *Computer Communications* 160 (2020), 351–374.

[140] Moondeep C. Shrestha, Tomoya Onishi, Ayano Kobayashi, Mitsuhiro Kamezaki, and Shigeki Sugano. 2018. Communicating directional intent in robot navigation using projection indicators. In *Proceedings of the 2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 746–751.

[141] M. C. Shrestha, T. Onishi, A. Kobayashi, M. Kamezaki, and S. Sugano. 2018. Communicating directional intent in robot navigation using projection indicators. In *Proceedings of the 2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 746–751. DOI: https://doi.org/10.1109/ROMAN.2018.8525528

[142] Roland Siegwart, Illah Reza Nourbakhsh, and Davide Scaramuzza. 2011. *Introduction to Autonomous Mobile Robots*. MIT Press.

[143] Grimaldo Silva and Thierry Fraichard. 2017. Human robot motion: A shared effort approach. In *Proceedings of the 2017 European Conference on Mobile Robots (ECMR)*. IEEE, 1–6.

[144] Ronal Singh, Tim Miller, Joshua Newn, Eduardo Velloso, Frank Vetere, and Liz Sonenberg. 2020. Combining gaze and AI planning for online human intention recognition. *Artificial Intelligence* 284 (2020), 103275.

[145] Emrah Akin Sisbot, Luis F. Marin-Urias, Rachid Alami, and Thierry Simeon. 2007. A human aware mobile robot motion planner. *IEEE Transactions on Robotics* 23, 5 (2007), 874–883.

[146] K. Smith, S. O. Ba, J. Odobez, and D. Gatica-Perez. 2008. Tracking the visual focus of attention for a varying number of wandering people. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30, 7 (2008), 1212–1229. DOI: https://doi.org/10.1109/TPAMI.2007.70773

[147] Krisztina Soproni, Ádám Miklósi, József Topál, and Vilmos Csányi. 2001. Comprehension of human communicative signs in pet dogs (canis familiaris). *Journal of Comparative Psychology* 115, 2 (2001), 122.

[148] Rainer Stiefelhagen, Michael Finke, Jie Yang, and Alex Waibel. 1999. From gaze to focus of attention. In *Proceedings of the International Conference on Advances in Visual Information Systems*. Springer, 765–772.

[149] Johannes Strassner and Marion Langer. 2005. Virtual humans with personalized perception and dynamic levels of knowledge. *Computer Animation and Virtual Worlds* 16, 3-4 (2005), 331–342.

[150] Mikael Svenstrup, Thomas Bak, and Hans Jørgen Andersen. 2010. Trajectory planning for robots in dynamic human environments. In *Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 4293–4298.

[151] Mason Swofford, John Peruzzi, Nathan Tsoi, Sydney Thompson, Roberto Martín-Martín, Silvio Savarese, and Marynel Vázquez. 2020. Improving social awareness through DANTE: Deep affinity network for clustering conversational interactants. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (2020), 1–23.

[152] Dag Sverre Syrdal, Nuno Otero, and Kerstin Dautenhahn. 2008. Video prototyping in human-robot interaction: Results from a qualitative study. In *Proceedings of the 15th European Conference on Cognitive Ergonomics: The Ergonomics of Cool Interaction*. 1–8.

[153] Daniel Szafir, Bilge Mutlu, and Terry Fong. 2015. Communicating directionality in flying robots. In *Proceedings of the 10th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '15)*. ACM, New York, NY, USA, 19–26. DOI: https://doi.org/10.1145/2696454.2696475

[154] Lei Tai, Jingwei Zhang, Ming Liu, and Wolfram Burgard. 2018. Socially compliant navigation through raw depth inputs with generative adversarial imitation learning. In *Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 1111–1117.

[155] Yusuke Tamura, Tomohiro Fukuzawa, and Hajime Asama. 2010. Smooth collision avoidance in human-robot coexisting environment. In *Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 3887–3892.

[156] S. Thrun, M. Beetz, M. Bennewitz, W. Burgard, A. B. Cremers, F. Dellaert, D. Fox, D. Haehnel, C. Rosenberg, N. Roy, and J. Schulte. 2000. Probabilistic algorithms and the interactive museum tour-guide robot minerva. *The International Journal of Robotics Research* 19, 11 (2000), 972–999.

[157] Elin Anna Topp and Henrik I. Christensen. 2005. Tracking for following and passing persons. In *Proceedings of the 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2321–2327.

[158] Elena Torta, Raymond H. Cuijpers, and James F. Juola. 2013. Design of a parametric model of personal space for robotic social navigation. *International Journal of Social Robotics* 5, 3 (2013), 357–365.

[159] Adrien Treuille, Seth Cooper, and Zoran Popović. 2006. Continuum crowds. *ACM Transactions on Graphics* 25, 3 (2006), 1160–1168. Retrieved from https://howtorts.github.io/2014/01/09/continuum-crowds.html

[160] Jérôme Truc, Phani-Teja Singamaneni, Daniel Sidobre, Serena Ivaldi, and Rachid Alami. 2022. KHAOS: A kinematic human aware optimization-based system for reactive planning of flying-coworker. In *Proceedings of the ICRA 2022*.

[161] Xuan-Tung Truong and Trung-Dung Ngo. 2016. Dynamic social zone based mobile robot navigation for human comfortable safety in social environments. *International Journal of Social Robotics* 8, 5 (2016), 663–684.
[162] Nathan Tsoi, Mohamed Hussein, Jeacy Espinoza, Xavier Ruiz, and Marynel Vázquez. 2020. SEAN: Social environment for autonomous navigation. In *Proceedings of the 8th International Conference on Human-Agent Interaction*. 281–283.
[163] 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 2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 6183–6190.
[164] Jur Van Den Berg, Stephen J. Guy, Ming Lin, and Dinesh Manocha. 2011. Reciprocal n-body collision avoidance. In *Proceedings of the Robotics Research*. Springer, 3–19.
[165] Dizan Vasquez, Billy Okal, and Kai O. Arras. 2014. Inverse reinforcement learning algorithms and features for robot navigation in crowds: An experimental comparison. In *Proceedings of the 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 1341–1346.
[166] Viswanath Venkatesh, Michael G. Morris, Gordon B. Davis, and Fred D. Davis. 2003. User acceptance of information technology: Toward a unified view. *MIS Quarterly* (2003), 425–478.
[167] Janardan Kumar Verma and Virender Ranga. 2021. Multi-robot coordination analysis, taxonomy, challenges and future scope. *Journal of Intelligent and Robotic Systems* 102, 1 (2021), 1–36.
[168] Alessandro Vinciarelli, Maja Pantic, and Hervé Bourlard. 2009. Social signal processing: Survey of an emerging domain. *Image and Vision Computing* 27, 12 (2009), 1743–1759.
[169] W. Grey Walter. 1950. An imitation of life. *Scientific American* 182, 5 (1950), 42–45.
[170] Atsushi Watanabe, Tetsushi Ikeda, Yoichi Morales, Kazuhiko Shinozawa, Takahiro Miyashita, and Norihiro Hagita. 2015. Communicating robotic navigational intentions. In *Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 5763–5769.
[171] Astrid Weiss and Christoph Bartneck. 2015. Meta analysis of the usage of the godspeed questionnaire series. In *Proceedings of the 2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 381–388.
[172] Xuesu Xiao, Bo Liu, Garrett Warnell, and Peter Stone. 2022. Motion planning and control for mobile robot navigation using machine learning: a survey. *Autonomous Robots* 46, 5 (2022), 569–597.
[173] Zhi Yan, Nicolas Jouandeau, and Arab Ali Cherif. 2013. A survey and analysis of multi-robot coordination. *International Journal of Advanced Robotic Systems* 10, 12 (2013), 399.
[174] Xinjie Yao, Ji Zhang, and Jean Oh. 2019. Following social groups: Socially compliant autonomous navigation in dense crowds. arXiv preprint arXiv:1911.12063 (2019).
[175] Harel Yedidsion, Jacqueline Deans, Connor Sheehan, Mahathi Chillara, Justin Hart, Peter Stone, and Raymond J. Mooney. 2019. Optimal use of verbal instructions for multi-robot human navigation guidance. In *Proceedings of the International Conference on Social Robotics*. Springer, 133–143.

Received 7 June 2021; revised 3 January 2024; accepted 1 February 2024