

# Prevention and Resolution of Conflicts in Social Navigation - a Survey

Reuth Mirsky      Xuesu Xiao      Justin Hart      Peter Stone

April 2020

## Abstract

With the approaching goal of having robots collaborate in shared human-robot environments, navigation in this context becomes both crucial and desirable. Recent developments in robotics have encountered and tackled some of the challenges of navigating in mixed human-robot environments, and in recent years we observe a surge of related work that specifically targets the question of how to handle conflicts between agents in social navigation. These contributions offer models, algorithms, and evaluation metrics, however as this research area is inherently interdisciplinary, many of the relevant papers are not comparable and there is no standard vocabulary between the researchers. The main goal of this survey is to bridge this gap by proposing such a common language, using it to survey existing work, and highlighting open problems. It starts by defining a conflict in social navigation, and offers a detailed taxonomy of its components. This survey then maps existing work while discussing papers using the framing of the proposed taxonomy. Finally, this paper propose some future directions and problems that are currently in the frontier of social navigation to help focus research efforts.

## 1 Introduction

Enabling autonomous robots to navigate in the presence of people and/or other robots has been studied for the past 70 years. Grey Walter built robotic “turtles” that could navigate on their own [143]. These robots, named Elmer and Elsie, were an exercise in minimalism and an attempt to demonstrate 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 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 constructed from hard-coded rules or learned from a relatively clean dataset [12, 52, 119]. Solving the problem of navigating in such a setting — in the presence of other robots or humans — is a complex, cross-disciplinary challenge with facets of 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

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. [111] gave a compact description of socially-aware navigation: *Socially-aware navigation is the strategy exhibited by a social robot which identifies and follows social conventions (in terms of management of space) in order 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.*

This survey focuses on three requirements that separate “Social Navigation” from more general navigation. These requirements are:

1. There exists an autonomously navigating agent. The agent has a specific, reachable navigational goal that it needs to reach.
2. There exists one (or more) human pedestrian(s).
3. Navigation takes place in natural environments “in the wild”, rather than in specifically built environments or labs.

Many works discuss challenges that occur when only one or two of these requirements are met. Tele-operation 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 artificial settings or simulations rather than in real world environments, such that 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). While this survey cites many works in which not all of the above requirements hold, because of their contributions to our understanding of social navigation, our main focus is on works that meet all three requirements.

This survey focuses on interactions with people which require the robot to reason about an encounter, specifically a *conflict*, in the context of social navigation. 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. All of these challenges, however, are orthogonal to the challenge of avoiding conflicts with other pedestrians. Better understanding conflict 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 speed by at least one of the parties, they will collide.

By this definition not all conflicts end in a collision, but every collision is preceded by a conflict that is not resolved. In this survey, we define collision avoidance as “conflict resolution.” (See more on the meaning of resolution of conflicts in the next section.) **This survey covers work where the authors include conflicts in their models, and makes the distinction between research toward avoiding such conflicts altogether (prevention) and research that models responses to such conflicts (resolution).**

Other works have presented ideas that overlap those in this survey, but from different perspectives. As previously mentioned, there are numerous works on the topic of social navigation that

focus on elements such as joint or group navigation [57, 85, 106, 146], giving navigational instructions [133, 147], detecting dynamic objects [27, 61], acting in human-like manner in social contexts such as waiting in line [90] or distributing flyers [116], and other factors not discussed in this survey; except in cases that are specifically in service to the purpose of assisting in detecting or avoiding conflicts. Here we detail the major related surveys, both to provide a reference for readers who are interested in those additional points of view and to define the scope of this survey.

Kruse et al. [68] highlight the 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, which requires this survey to narrow focus somewhat as we update their coverage of the topic. Our focus is on the narrower topic of conflicts that arise between a robot and pedestrians. Hoogendoorn and Bovy [47] 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.

Charalampous et al. [16] 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 operating in close proximity with humans, or as we refer to it, robots in conflict situations. Specifically, we aim to provide basic definitions that will be used to standardize future works on the problem of robots that navigate in environments in close proximity with humans.

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

## 2 Taxonomy

This section systematically lays out the taxonomy with which we survey each of the central elements of social navigation: models (Section 3) and algorithms (Section 4). For each work we discuss, we identify eight attributes as listed in Table 1. Below we discuss this list of attributes (in bold) and the values (in italics) they can take (we write in parentheses the shorthand version of the values as they appear in the tables in Sections 3 and 4). 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.

Table 1: The Social Navigation Taxonomy

| Attributes                       | Values   |
|----------------------------------|--|
| <b>Prevention vs. Resolution</b> | <i>Prevention (P) / Resolution (R) / Both (B) / Neither (N)</i>  |
| <b>Robot Role</b>                | <i>Reactor (R) / Initiator (I) / Both (B) / Neither (N)</i>  |
| <b>Number of Agents</b>          | <i>Absolute Number (Abs = # of agents) / Density (D = #/m<sup>2</sup>)</i>   |
| <b>Observability</b>             | <i>Full / Partial / Depth / RGB</i>  |
| <b>Motion Control</b>            | <i>SFM / ORCA / ROS / Human / Other</i>  |
| <b>Communication</b>             | <i>None (N) / Indirect (I) / Direct (D)</i>  |
| <b>Experiment Type</b>           | <i>Simulation (Sim) / Robot-Only (R) / Human Experiment (HE) / Survey (Sur)</i>  |
| <b>Agent Type</b>                | <i>Human-Robot (H-R) / Human-Agent (H-A) / Human-Human (H-H) / Homogeneous Agents (Hom) / Heterogeneous Agents (Het)</i> |

## 2.1 Taxonomy Attributes and Values

**Prevention vs. Resolution** *Prevention (P) / Resolution (R) / Both (B) / Neither (N)*. This attribute identifies work on planning ahead to avoid conflicts (*Prevention*) vs. work on solving conflicts when they are imminent (*Resolution*).

**Robot Role** *Reactor (R) / Initiator (I) / Both (B) / Neither (N)*. The robot can infer and react to the human’s plan (*Reactor*) or it can actively influence the human’s plan (*Initiator*). Some works try to compromise between the two (*Both*) and a few papers, such as human-only experiments, do not discuss a robot strategy (*Neither*).

**Number of Agents** *Absolute Number (Abs) / Density (D)*. Several papers deal with a one-on-one interaction and some 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 of pedestrians, we include the navigating robot in the count, so it can be compared 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 papers rely on an existing motion controller, and a 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 <sup>1</sup> (*ROS*), 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 or relevant (such as research projects that use cellular automata, point-based navigation, Dijkstra’s algorithm, or other types of search for motion planning) or in which the motion

<sup>1</sup><https://www.ros.org/>

control is novel and is a major part of the paper’s contribution (such as SIPP [139], Social Momentum [80], or LM-SARL [17]). We mark these cases as “other”, but 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 [26] and stigmergy [7]. *Direct* communication means that there is some mechanism that was added to the robot to allow communication. See Figure 1 for examples.

**Experiment Type** *Simulation (Sim) / Robot-Only (R) / Human Experiment (HE) / Survey (Sur)*. Many researchers run experiments in *Simulation* as part of their empirical evaluation, either as the only type of evaluation or in addition to real-world experiments. *Robot-Only* experiments are defined as experiments in the real world that do not involve people. *Human Experiments* are real world experiments which include people; often 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: Human experiment, robot-only, simulation. The exception for this policy is for surveys, which usually accompany additional experiments, and hence are mentioned together with another experiment type.

**Agent Type** *Human-Robot (H-R), Human-Agent (H-A), Human-Human (H-H), Homogeneous Agents (Hom), Heterogeneous Agents (Het)*. This survey focuses on social navigation between 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 approaches are 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 papers are included which provide a fundamental understanding of purely-human navigation and present evaluations that do not involve robots at all (*Human-Human*).

Some of the attributes and the values chosen to be presented here are not intuitive. Developing these ideas requires careful reasoning. Here, we discuss these ideas and explain their rationale. First, **Prevention vs. Resolution** is an important attribute, but it is non-trivial to classify a robot’s behavior as being either “prevention” or “resolution.” On the one hand, it is clear that prevention and resolution are different tasks that can direct the robot’s behavior. On the other hand, the difference between these tasks can be elusive. Several papers identify a specific behavior that is unique to resolution of conflicts. Reynolds [109] defined “unaligned collision avoidance” as a scenario in which the agent “will steer laterally to turn away from the potential collision. It will also accelerate forward or decelerate backwards to get to the indicated site before or after the predicted collision”. Park et al. [102] referred to a similar scenario: “The combination of *speed reduction* and the *change of the heading direction* is only used as a last resort when the pedestrians are too close to each other or when the pedestrian has winded far off the original path, and the angle to the destination is large enough such that side stepping is unnatural.” We identify these scenarios as resolution. In most cases, papers discussing resolution involve a (1) reduction in speed and (2) a change in heading by a robot or person. These papers usually also model a single point of interaction between the robot and the person. Papers on prevention, on the other hand, discuss navigational

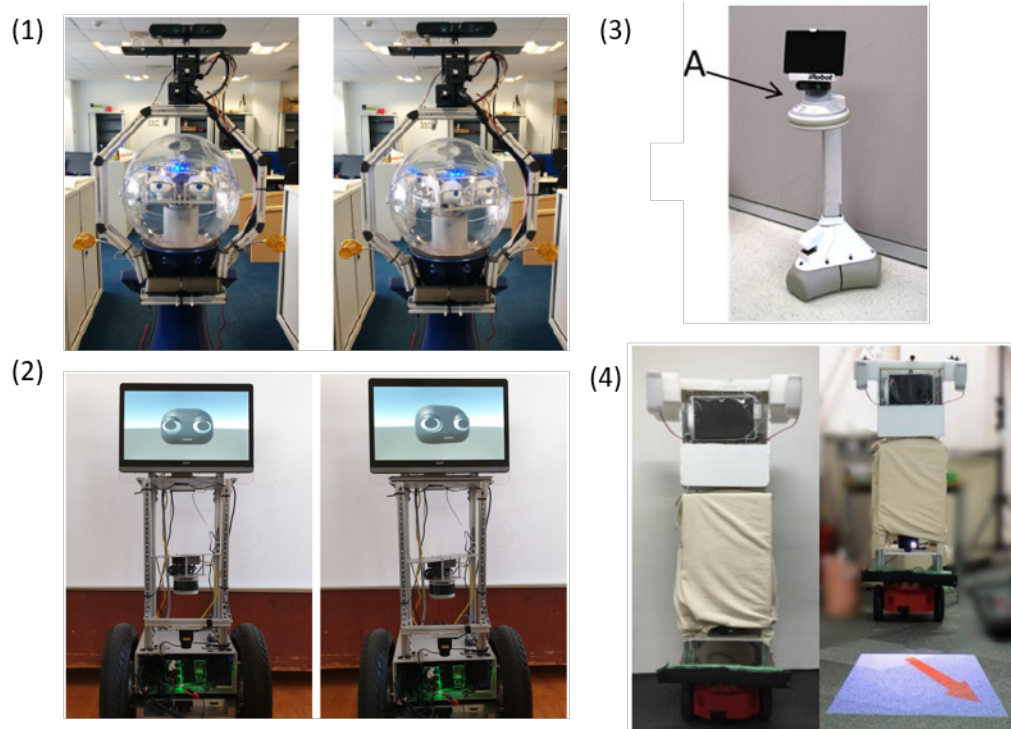


Figure 1: Various direct communication behaviors: (1) mechanical gaze [81]; (2) virtual gaze [41]; (3) sensor rotation [31]; (4) arrow signaling [117]

techniques that can be implemented from farther away which inform the local navigation process, such as a trajectory change. There are, however, some papers that we categorize as resolution even though they involve only (1) or (2) from above. For example, some papers use a communicative cue for signaling the robot’s intended trajectory (see *Direct Communication*). We categorize these papers as resolution, because the distance between the human and the robot is short enough that without the communicative cue, both a reduction in speed (1) and a change in heading (2) would need to take place.

**Observability** is an important factor to consider, especially when discussing simulations, which will explicitly model the observations that can be made by agents acting in the scene. Many simulations assume that the robot (or pedestrians around it) has full (ground truth) observability. Other simulations restrict observability in an artificial way, to suit what a robot would realistically be able to sense (partial observability). In the discussion of these papers, it is important to keep in mind that these observation modalities may not be realistic to implement on real robots thus impacting how these contributions must be interpreted in the context of real-world embodied social navigation.

With respect to the **Communication** attribute, we make the distinction between communication that is indirect or direct and communication that is implicit or explicit. Implicit communication is communication that is conveyed by people regardless of the intention to communicate (the interpretation of eye gaze is implicit), and explicit communication communication that takes place intentionally (speech is explicit) [22]. Because robots do not tend to naturally communicate anything implicitly (for example, not all robots have “eyes” and those that do do not necessarily need to turn them to “look” at something, and do not naturally turn them to where they are about to walk), 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 [140]), implicit-direct (e.g., gaze change on a virtual head [1]), and explicit-direct (e.g., arrow projections on the floor [144]).

Some papers present more than one set of experiments, for example, both *Human-Robot* in a human studies and *Heterogenous agents* in simulation. As this survey focuses on social navigation, we choose to highlight *Human-Robot* experiments. Thus, in papers that present more than one set of experiments, for our **Experiment Type** and **Agent Type** attributes we choose the values that are relevant to the set of experiments which is the most similar to a *Human-Robot* interaction: *Human-Robot* > *Robot-Only* > *Simulation* > *Survey*.

We also wish to 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., prevention vs. resolution). Usually, the abstract attributes and their values depend on the concrete attributes. Figure 2 presents the hierarchical structure of these attributes. 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 the beginning. 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 whether the interaction will be **Prevention vs. Resolution** and the **Agent Type** which can perceive the communication channel chosen.

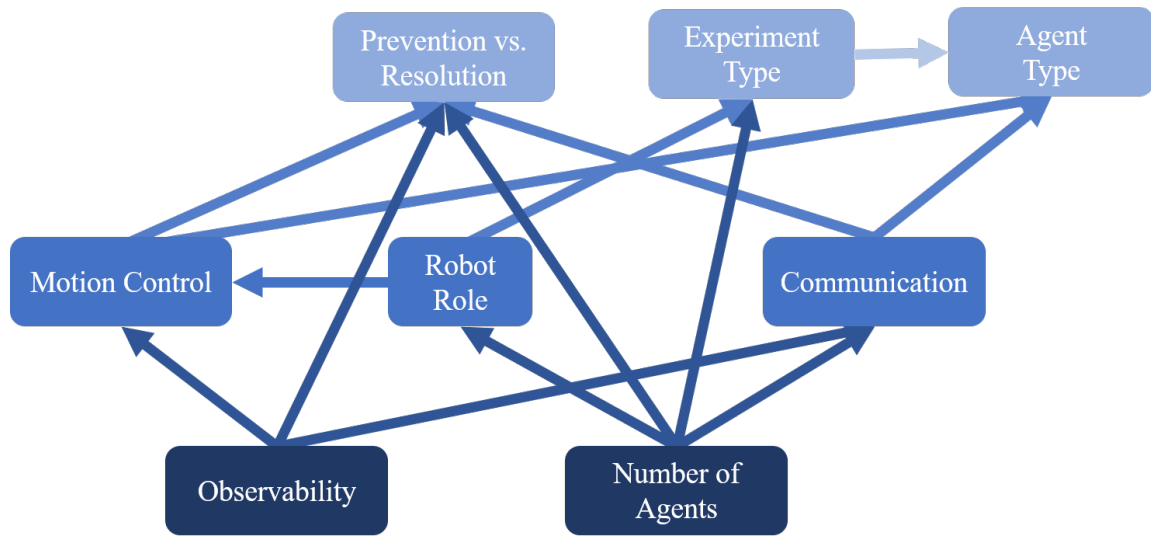


Figure 2: The hierarchical structure of the taxonomy’s attributes

## 2.2 Additional Concepts

There are some additional concepts that are worth mentioning, but which we decided to exclude from explicit inclusion in our taxonomy. As research and discussion on social navigation progresses, this taxonomy could be extended to include these attributes.

First, one seemingly-important factor to consider in the taxonomy is **collision type**. When referring to collisions, the majority of work mention head-on collisions or side collisions, while rear-end collisions are the least commonly investigated type. In the papers presented in this survey, there is not a single work that explicitly discusses only one type of collision; rather there are several papers that propose ways to categorize collisions according to the required response from pedestrians and / or the robot. Reynolds [109] 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 running into each other.*” Separation is similar to 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.” Mavrogiannis et al. [80] discuss the point in space and time where agents collide and call this point “entanglement.” This concept raises an additional question about the concrete implementation of this collision point - what’s considered close enough to be an entanglement in a social context (e.g., Mavrogiannis et al. [79] segmented their experiments with  $d \leq 1$  meter between the robot and the human). Thus, 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 a leg? Overall, the definition for collision varies between researchers and may be a parameter that can be adjusted.

Another common discussion point is **context awareness and semantic mapping**. Many papers discuss the need for mobile social robots to be aware of their context [16]. 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 [64]. 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 the interaction between a mobile robot and pedestrians is how people **react to humans vs. robotic counterparts**. Will a human interaction with another human produce a similar or different response from an interaction with a robot? 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, although it is not unanimously agreed upon. In their survey on proxemics for social navigation, Rios-Martinez et al. [111] 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 [9] have identified that people feel comfortable when a robot moves in speeds that are between  $0.254m/s$  and  $0.381m/s$ , while a normal walking speed for young human is about  $1m/s$ . This difference in speed 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** is useful and will be used in our discussions in the next sections, but is not included as an attribute in the taxonomy. How a robot can best communicate with humans is a rich and versatile research area; and is taken into consideration through **observability** and **communication** in the taxonomy. To discuss such communication more accurately, we make the distinction between cues and signals [142]. Cues are the low-level inputs that the robot can receive or send, such as: gaze, position, language, etc. 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 resolution, and the way to implement them in a robot (or detect them in a human) is through social cues.

One attribute that is relevant in a broader context than social navigation is **focused vs. unfocused interaction**. Goffman [37] 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. [111] use this attribute to identify different types of navigational behaviors in robots: minimizing probability of encounter, avoiding collisions, passing people, staying in line, approaching humans, following people, and walking side-by-side. Because the papers in this survey revolve around conflicts, the robot and the human do not share focus, and hence all included papers involve strictly unfocused interactions. Focused vs. unfocused interaction are not considered as part of the taxonomy.

Lastly, the topic of differences in navigation with **independent pedestrians vs. groups vs. crowds** has enjoyed recent popularity. Most social navigation papers 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 the group size distribution highly depends on context (a casual Saturday afternoon stroll vs. a workday morning commute) [21]. Recent research has demonstrated that in many contexts, more than 50% of pedestrians are not travelling alone, but in groups [85]. Thus, the context in which navigation takes place determines whether it is necessary to consider a surrounding

Table 2: An overview of the different multiagent based models used in social navigation. **P. vs. R.** is Prevention vs. Resolution, **Role** refers to the robot role, **Obs.** is observability, **Com.** refers to communication, and **Exp. type** is the experiment type.

| Year | Paper                      | P vs. R | Role | # Agents | Obs.        | Motion Control     | Com. | Exp. Type | Agent Type |
|------|----------------------------|---------|------|----------|-------------|--------------------|------|-----------|------------|
| 1997 | Musse and Thalmann [89]    | R       | R    | Abs=10   | Full        | Other (Hand Coded) | N    | Sim       | Hom        |
| 2005 | Strassner and Langer [126] | N       | N    | Abs=2    | Partial     | Other (Hand Coded) | N    | Sim       | Hom        |
| 2010 | Foka and Trahanias [32]    | P       | R    | Abs=6    | Depth       | Other (POMDP)      | I    | HE        | H-R        |
| 2011 | Van Den Berg et al. [140]  | P       | R    | Abs=1000 | Full        | ORCA               | I    | Sim       | Hom        |
| 2013 | Bandyopadhyay et al. [3]   | P       | R    | Abs=4    | RGB + Depth | SFM                | N    | HE + Sim  | H-R        |
| 2014 | Bonneaud and Warren [8]    | R       | R    | Abs=20   | Full        | Other              | N    | Sim       | Hom        |
| 2014 | Okal and Arras [97]        | R       | R    | Abs=176  | Partial     | SFM                | N    | Sim       | Hom        |
| 2016 | Godoy et al. [36]          | P       | B    | Abs=100  | Full        | Other              | N    | HE + Sim  | R-R        |
| 2019 | Chen et al. [17]           | P       | R    | Abs=6    | Full        | Other (LM-SARL)    | N    | HE        | Hom        |

crowd. There is a surge of recent work that investigate conflicts with a groups rather than with a single pedestrian or a crowd with no social formations. This subarea of social navigation is not yet mature enough to identify either general research trends or the conditions under which such navigational groups form. As such, this attribute has not been included in the taxonomy.

### 3 Models

This section details various models used for social navigation. The discussion is grouped according to three main underlying models: Multiagent systems, human-inspired models, and physics-based models (specifically, the social force model and other force modelling). Each of these categories represents a different set of assumptions — as well as a different research community — that each model comes from. Navigation in multiagent systems is usually designed with the premise that agents navigating in an environment are homogeneous. These papers include multi-robot navigation models and crowd modelling. Thus, a social navigation model that aims to build on this work would be required to reason about agents with different, sometimes unknown, behaviors. Other models are inspired from insights about human navigation. These papers provide measurements and rules that explain how people navigate amongst themselves, and a social navigation model is required to translate these rules into robot motion and perception. We specify papers leveraging the social force model as a separate category that is inspired from the physical modelling of forces. Many models have been proposed which build upon the seminal work by Helbing and Molnar [44] with additional types of forces. Finally, some of the papers sit at an intersection between two categories. In these cases, the work is categorized according to the type of motion control associated with the chosen model, such that it is grouped with work that uses similar motion control.

#### 3.1 Multiagent Systems

Two research communities that have contributed significantly to the study of the problem 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. Due to these different goals, multi-robot work usually comes from the perspective of a single interaction (or only a few) in mind under realistic constraints; while the challenge of

crowd modelling is to model interactions between hundreds and thousands of agents simultaneously, though the perception and movement restrictions on those agents tends to not be grounded in the physical constraints that both robots and real people must contend with.

Many have considered the challenge of multi-robot navigation [145]. As 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. [140] present the principle of optimal reciprocal collision avoidance (ORCA) that provides a sufficient condition for multiple robots to avoid collisions among one another, and thus can guarantee collision-free navigation. Chen et al. [17] model human-robot and human-human interactions, then infer relative importance through a pooling module via a self-attention mechanism, finally planning motions.

A branch of multi-robot research was inspired by planning under uncertainty, using Markov Decision Processes (MDPs). Foka and Trahanias [32] model hot points of human navigation, a probabilistic prediction of the person’s destination. In their work, they use a Partially Observable MDP (POMDP) solved online at each time step to determine which actions the robot actually performs. Bandyopadhyay et al. [3] specifically model human intention with Mixed Observability MDP (MOMDP) and then plan the motion of a robot in this setting.

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 [89] propose a model of crowd behavior where agent behavior is determined using a predefined set of rules. Strassner and Langer [126] use behavioral rules for modelling each person’s behavior in a crowd. Such behaviors include perceiving, storing, and forgetting knowledge. Bonneaud and Warren [8] 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 [97] present a model for crowd behavior in which groups are formed. Their representation gives each individual an internal state, and under a set of predefined conditions pedestrians can choose to walk together.

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

### 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 modelling of human behavior, which then leverages these models for improving robot navigation. One highly cited paper empirically evaluates human behavior in situations of obstacle avoidance [23]. Their work investigates the relation between object avoidance and finding one’s aimpoint in a series of human studies. Their results are summarized in a decision-tree to facilitate reasoning about collision detection with other objects (static or moving) and defined the concept of Gaze-Movement Angle (GMA) — the angle between one’s gaze and one’s direction of movement — which can be used to estimate where a collision might occur. As a different way to estimate the expected collision point, Carel [10] defined  $\tau$  to be the time to bypass a dynamic obstacle (human or not). Moussaïd et al. [84] use  $\tau$  to heuristically plan how to navigate in a way that avoids collisions. Park et al. [102] claim that GMA-based collision prediction has several advantages over the time-to-contact ( $\tau$ ) 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 [63] investigate the Information Process

Table 3: An overview of the different human-inspired and psychology-based models used in social navigation. **P. vs. R.** is Prevention vs. Resolution, **Role** refers to the robot role, **Obs.** is observability, **Com.** refers to communication, and **Exp. type** is the experiment type.

| Year | Paper                      | P vs. R | Role | # Agents | Obs.      | Motion Control     | Com. | Exp. Type | Agent Type |
|------|----------------------------|---------|------|----------|-----------|--------------------|------|-----------|------------|
| 1995 | Cutting et al. [23]        | R       | R    | Abs=2    | Partial   | Other (Hand Coded) | I    | Sim + Sur | H-H        |
| 1998 | Jeffrey and Mark [48]      | R       | R    | Abs=4+   | Partial   | Human              | N    | Sim       | H-H        |
| 1999 | Patla et al. [103]         | B       | R    | Abs=2    | Partial   | Human              | N    | HE        | H-H        |
| 1999 | Reynolds [109]             | R       | R    | Abs=2    | Full      | None               | N    | None      | Hom        |
| 2002 | Bennewitz et al. [5]       | P       | R    | Abs=2    | Depth     | Other (Learned)    | N    | R         | H-R        |
| 2008 | Gérin-Lajoie et al. [34]   | R       | R    | Abs=2    | None      | Human              | N    | HE        | H-H        |
| 2010 | Henry et al. [45]          | P       | R    | None     | Depth     | Other (A*)         | N    | Sim       | Hom        |
| 2010 | Kitazawa and Fujiyama [63] | R       | R    | Abs=4    | Partial   | Human              | N    | HE        | H-H        |
| 2011 | Moussaïd et al. [84]       | R       | R    | Abs=96   | Partial   | Other (Hand Coded) | N    | HE + Sim  | Hom        |
| 2011 | O’Callaghan et al. [95]    | P       | R    | Abs=2    | Depth     | Other (Planner)    | N    | HE        | H-R        |
| 2012 | Rios-Martinez et al. [110] | P       | R    | Abs = 6  | RGB+Depth | ROS                | N    | Sim       | Hom        |
| 2013 | Lu et al. [74]             | P       | R    | Abs = 2  | Partial   | ROS                | N    | Sim       | Het        |
| 2013 | Park et al. [102]          | R       | R    | D=0.1-1  | Partial   | Other (Hand Coded) | I    | Sim       | Hom        |
| 2014 | Charalampous et al. [15]   | P       | R    | Abs=2    | RGB+Depth | Other              | N    | HE        | H-R        |
| 2014 | Papadakis et al. [101]     | N       | I    | Abs=2    | RGB+Depth | None               | I    | HE        | H-R        |
| 2014 | Vasquez et al. [141]       | P       | R    | Abs=6+   | Full      | Other (Dijkstra)   | I    | Sim       | H-A        |
| 2015 | Unhelkar et al. [139]      | R       | R    | Abs=2    | Full      | Other (SIPP)       | N    | HE + Sim  | H-A        |
| 2016 | Mead and Matorić [82]      | N       | I    | Abs=2    | RGB       | None               | D    | HE        | H-R        |
| 2016 | Truong and Ngo [137]       | P       | R    | Abs = 4  | RGB+Depth | Other (D*)         | N    | HE + Sim  | H-R        |
| 2020 | Senft et al. [115]         | R       | R    | Abs = 2  | Depth     | Other (Hand Coded) | I    | HE        | H-R        |

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 3). For example, Park et al. [102] 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 [34, 40, 48]. While the original formulation of personal space is depicted by Hall [40] as a concentric circle, later work extends that to an egg shape [42], ellipse [44], or as asymmetrical (smaller on the dominant side) [34]. 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  $m^2$  [85]. Rios-Martinez et al. [110] incorporate both personal space and IPS-based constraints into an adaptive optimization algorithm to enable more human-like navigation. Truong and Ngo [137] propose a comprehensive framework that reasons about pedestrians’ extended personal space and the 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 a sudden trajectory change as one expects to see in conflict resolution. Patla et al. [103] 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. [139] discretized walking trajectories as a decision problem regarding which target a person would walk toward. They incorporated this information into an anytime path planner [91] and evaluated this enhanced planner in simulation. Holman et al. [46] extend this predictive model to incorporate gaze. Senft et al. [115] 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

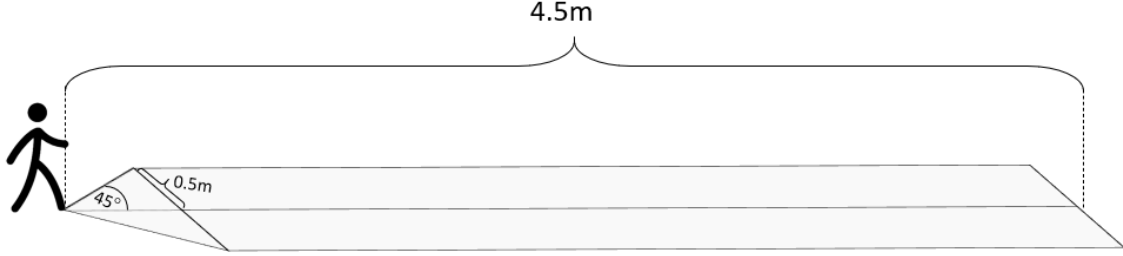


Figure 3: Information Process Space - the visual processing coverage of pedestrians, as measured by [63] and depicted by Rios-Martinez et al. [111].

Table 4: An overview of the different physics-inspired models used in social navigation. **P. vs. R.** is Prevention vs. Resolution, **Role** refers to the robot role, **Obs.** is observability, **Com.** refers to communication, and **Exp. type** is the experiment type.

| Year | Paper                   | P vs. R | Role | # Agents    | Obs.    | Motion Control       | Com. | Exp. Type | Agent Type |
|------|-------------------------|---------|------|-------------|---------|----------------------|------|-----------|------------|
| 1995 | Helbing and Molnar [44] | P       | R    | D=0.3       | Full    | SFM                  | N    | Sim       | H-A        |
| 2003 | Loscos et al. [73]      | R       | R    | Abs=6000    | Partial | Other (Hand Coded)   | N    | Sim       | Hom        |
| 2009 | Karamouzas et al. [55]  | R       | R    | Abs=1000    | Full    | SFM                  | N    | Sim       | Hom        |
| 2010 | Moussaïd et al. [85]    | N       | N    | D=0.03-0.25 | Full    | SFM                  | N    | HE + Sim  | H-H        |
| 2010 | Svenstrup et al. [127]  | P       | R    | Abs=40      | Full    | Other (Modified RRT) | I    | Sim       | Hom        |
| 2020 | Swofford et al. [128]   | N       | I    | Abs=18      | RGB     | ROS                  | N    | HE        | H-R        |

of three steps: step, slide, and rotate.

All of the contributions mentioned above leverage insights from empirical studies on humans and robots to manually construct models for social navigation. However, together with the increasing abilities of machine learning, different learning techniques have been used to automatically learn models of navigation in social contexts. Lu et al. [74] propose a planning model that can be tuned to match different social navigation contexts. Bennewitz et al. [5] learn motion patterns of people that can be used for trajectory prediction in social robots. Henry et al. [45] extend this approach by modeling partial trajectories. More recently, Vasquez et al. [141] 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 also present results for the task of socially-compliant robot navigation in crowds, evaluating two different IRL approaches and several feature sets in large-scale simulations.

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

### 3.3 Physics-based Model

Researchers have also used models inspired by physics to represent the dynamics and interactions among different moving agents. Helbing and Molnar [44] were the first to propose the Social Force Model (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 the SFM models to handle additional forces: Karamouzas et al. [55] add an evasive force that uses collision prediction and avoidance, which makes agents

more proactive and anticipatory than the classical SFM. Moussaïd et al. [85] propose several group-related forces that help model pedestrians that walk in a group, and Swofford et al. [128] use a Deep Affinity Network (DANTE) to predict the likelihood that two individuals in a scene are part of the same conversational group, with consideration for the social context in which these interactions take place. A different type of force inspired work uses potential fields attached to moving pedestrians [127]. 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 for 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 are included which 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 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 the 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 traces. Offline inference does not happen in real-time and usually reasons about more than one trajectory during the learning phase.

#### 4.1.1 Online Inference

Cutting et al. [23] offer an early attempt to evaluate the trajectory of a passerby by calculating their gaze-movement angle (GMA) and reacting to it. The robot designed by Tamura et al. [132] detects pedestrians by using a laser range finder and tracks using a Kalman filter. They apply a social force model 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. [35] 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 is more similar to the participants’ expectations.

Table 5: An overview of the different inference algorithms used in social navigation. **P. vs. R.** is Prevention vs. Resolution, **Role** refers to the robot role, **Obs.** is observability, **Com.** refers to communication, and **Exp. type** is the experiment type.

| Year | Paper                    | P vs. R | Role | # Agents | Obs.        | Motion Control         | Com. | Exp. Type | Agent Type |
|------|--------------------------|---------|------|----------|-------------|------------------------|------|-----------|------------|
| 2006 | Pacchierotti et al. [99] | R       | R    | Abs=3    | Depth       | Other (Hand Coded)     | N    | Sim       | Hom        |
| 2007 | Gockley et al. [35]      | P       | R    | Abs=2    | Depth       | Other (CVM)            | D    | HE        | H-R        |
| 2007 | Sisbot et al. [122]      | P       | R    | Abs=2    | RGB + Depth | Other (HAMP)           | I    | HE        | H-R        |
| 2009 | Kirby et al. [60]        | P       | R    | Abs=4    | Depth       | Other (A*)             | N    | Sim       | Hom        |
| 2010 | Ohki et al. [96]         | P       | R    | Abs=5    | Full        | Other (Hand Coded)     | N    | R         | H-R        |
| 2010 | Pandey and Alami [100]   | P       | R    | Abs=2    | Full        | Other (Hand Coded)     | N    | HE        | H-R        |
| 2010 | Tamura et al. [132]      | R       | R    | Abs=2    | Depth       | SFM                    | N    | HE        | H-R        |
| 2011 | Diego and Arras [24]     | P       | R    | Abs=5    | None        | Other (Modified TSP)   | N    | HE        | H-R        |
| 2012 | Kuderer et al. [69]      | P       | R    | Abs=3    | Full        | Other (learned)        | N    | HE        | H-R        |
| 2012 | Luber et al. [75]        | P       | R    | Abs=2    | Full        | Other (RMP)            | N    | Sim       | H-A        |
| 2013 | Ratsamee et al. [107]    | P       | R    | Abs=2    | RGB + Depth | SFM                    | N    | HE        | H-R        |
| 2014 | Gómez et al. [38]        | P       | R    | Abs=5    | Full        | Other (Planning)       | N    | Sim       | R-R        |
| 2016 | Kim and Pineau [59]      | R       | R    | Crowd    | RGB + Depth | Other (Costmap Search) | I    | HE        | Hom        |
| 2016 | Kretzschmar et al. [65]  | P       | R    | Abs=3    | Depth       | Other (RPROP)          | I    | HE        | Hom        |
| 2016 | Okal and Arras [98]      | P       | R    | Abs=4    | Depth       | ROS                    | I    | HE        | Hom        |
| 2016 | Pfeiffer et al. [104]    | P       | R    | Abs=891  | RGB         | Other (Max Entropy)    | I    | HE        | Hom        |
| 2017 | Bera et al. [6]          | P       | R    | D<=2     | Full        | Other (SocioSense)     | N    | Sim       | Het        |
| 2017 | Chen et al. [18]         | P       | R    | Abs=10+  | RGB         | Other (Learned)        | N    | HE        | H-R        |
| 2017 | Chen et al. [19]         | P       | R    | Abs=6    | Full        | Other (Learned)        | N    | Sim       | R-R        |
| 2018 | Ding et al. [25]         | P       | R    | Abs=20   | Depth       | None                   | N    | Sim       | R-R        |
| 2018 | Everett et al. [28]      | P       | R    | Abs=10+  | RGB + Depth | Crowd                  | N    | HE        | H-R        |
| 2018 | Jiang et al. [49]        | R       | B    | Abs=2    | RGB         | Other (Hand Coded)     | N    | Sim       | H-A        |
| 2018 | Li et al. [70]           | P       | R    | Abs=3+   | Depth       | Other (Learned)        | N    | HE        | H-R        |
| 2018 | Long et al. [72]         | P       | R    | Abs=100  | Depth       | Other (Learned)        | N    | Sim       | R-R        |
| 2018 | Tai et al. [131]         | P       | R    | Abs=3    | Depth       | Other (Learned)        | N    | HE        | H-R        |
| 2019 | Jin et al. [51]          | P       | R    | Abs=4    | Depth       | Other (Learned)        | N    | HE        | H-R        |
| 2019 | Meng et al. [83]         | N       | N    | Abs=1    | RGB         | None                   | N    | R         | R-R        |
| 2019 | Nardi and Stachniss [92] | N       | N    | Abs=1    | Full        | Other (Hand Coded)     | N    | Sim       | R-R        |
| 2020 | Liang et al. [71]        | P       | R    | Abs=10+  | RGB + Depth | Other (Learned)        | N    | HE        | H-R        |

Others have leveraged human gaze to infer the trajectory of pedestrians. Gaze is a very strong communicative cue used by human, in the context of collaborative settings in general [13] and for navigation in particular [2]. It has been shown that not only people can partially understand gaze cues from very young age, but also chimpanzees and dogs [105, 124]. 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. [125] 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. [123] 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 [121] and as a cue for interacting with copilot systems in cars [49, 50], also with the aim of inferring 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 [113]. 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 which work without such equipment [112] may soon allow us to perform the inverse of the robot experiments presented here, with the robot reacting to human gaze. Ratsamee et al. [107] 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’s trajectories during execution, some leverage these trajectories to learn and infer how a human would react in a social navigation interaction. Pacchierotti et al. [99] design 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 **Inverse Reinforcement Learning (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 straight forward application of IRL is by Kim and Pineau [59], to learn a cost function that respects social variables over features extracted from a RGB-D sensor. This work used IRL to infer cost functions in a social navigation context: navigational features were firstly 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 local path using the learned cost function to respect social variables. Obstacle avoidance was still handled by a low-level controller. Okal and Arras [98] tackle cost function representation at a global level in social context: they developed a graph structure and used Bayesian IRL to learn the cost for this representation. With the learned global representation, traditional global planner ( $A^*$ ) planned a global path over this graph, and POSQ steer function for differential-drive mobile robots served as a local planner. Henry et al. [45] use Inverse Reinforcement Learning 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. [104] to model agents’ trajectories for planning and by Kretzschmar et al. [65] to infer the parameters of the navigation model that matches the observed behavior in expectation. Kuderer et al. [69] also use human demonstrations, but instead of using a Markov Decision Process, they elicit features from the human trajectories, and then use Entropy maximization for choosing the robot’s behavior. Luber et al. [75] use unsupervised learning from surveillance data to learn motion patterns and augment this knowledge to a motion planner.

Sisbot et al. [122] create a human aware motion planner (HAMP) that is explicitly given a cost model for safety and for legibility, and the robot reasons about the joint cost of these two properties in its planning process. Cost were also implicitly defined by identifying an underlying model from demonstrated data. Kirby et al. [60] model human social conventions at the global planning stage. This trait 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 on 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 resolve and prevent conflicts, i.e. to realize collision avoidance. As a continuation of previous CADRL work [19], Chen et al. [18] 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). Within the same line of research, but to relax the assumption of other agents’ dynamics, Everett et al. [28] propose GA3C-CADRL to use 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. [51] contain 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 trained by RL [25]. Tai et al. [131] use Generative Adversarial Imitation Learning (GAIL) to learn continuous actions from depth image and desired force toward the target. This improved safety and efficiency upon pure BC. Li et al. [70] propose a new problem, Socially Concomitant Navigation (SCN), in addition to collision avoidance in traditional social navigation: the robot also needs to consider the motion of its companion so as 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. [6] create SocioSense, a social navigation algorithm that categorize pedestrians according to psychological traits (e.g. *shy*, *tense*) and adjusts the robot’s velocity according the the pedestrians around it.

To observe social rules when navigating in densely populated environments, Yao et al. [146] 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 through providing a local goal to the local planner. Other components of the existing navigation pipeline, e.g. state estimation, collision avoidance, etc., were still functioning as is. 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. [71] develop CrowdSteer, a 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. [78] 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.

## 4.2 Conveying the Robot’s Goal to the Human

Dragan et al. [26] formally define the concepts of legibility (motion that allows the observer to

Table 6: An overview of the different intention-conveying algorithms used in social navigation. **P. vs. R.** is Prevention vs. Resolution, **Role** refers to the robot role, **Obs.** is observability, **Com.** refers to communication, and **Exp. type** is the experiment type.

| Year | Paper                  | P vs. R | Role | # Agents | Obs.        | Motion Control     | Com.  | Exp. Type | Agent Type |
|------|------------------------|---------|------|----------|-------------|--------------------|-------|-----------|------------|
| 2009 | Nummenmaa et al. [94]  | R       | I    | Abs=2    | Partial     | Other (Hand Coded) | D     | Sim       | H-A        |
| 2013 | Fiore et al. [31]      | R       | I    | Abs=2    | Depth       | Other (Hand Coded) | I + D | HE        | H-R        |
| 2015 | May et al. [81]        | R       | I    | Abs=2    | RGB + Depth | Other (A*)         | D     | HE        | H-R        |
| 2015 | Szafir et al. [130]    | P       | I    | Abs=2    | RGB + Depth | Other (Hand Coded) | D     | HE + Sur  | H-R        |
| 2015 | Unhelkar et al. [139]  | P       | N    | Abs=1    | Full        | Other (SIPP)       | N     | Sim       | Hom        |
| 2015 | Watanabe et al. [144]  | R       | I    | Abs=2    | Depth       | ROS                | D     | HE        | H-R        |
| 2016 | Khambhaita et al. [58] | R       | I    | Abs=2    | RGB + Depth | ROS                | D     | HE + Sur  | H-R        |
| 2018 | Baraka and Veloso [4]  | P       | I    | Abs=2    | RGB + Depth | ROS                | D     | HE + Sur  | H-R        |
| 2018 | Fernandez et al. [29]  | R       | I    | Abs=2    | Depth       | Other (Hand Coded) | D     | HE        | H-R        |
| 2018 | Lynch et al. [76]      | R       | I    | Abs=2    | Full        | Other (Hand Coded) | D     | Sim       | H-A        |
| 2018 | Shrestha et al. [117]  | R       | I    | Abs=2    | Full        | Other (Hand Coded) | D     | HE + Sur  | H-R        |
| 2020 | Hart et al. [41]       | R       | I    | Abs=2    | Depth       | Other (Hand Coded) | D     | HE        | H-R        |

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 by the work by Sciutti et al. [114], who survey the concept of “motor resonance” between an acting robot and an observing human. Kitagawa et al. [62] recently presented a motion planning algorithm for omni-directional robots to resemble human movements in a time-efficient manner.

Many contributions use verbal signals for guidance [133]. Jeffrey and Mark [48] 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. [147] 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 [14]. To deal with this challenge, many contributions take inspiration from the theory of proxemics [40] as a non-verbal way to convey intent or restriction. Rios-Martinez et al. [111] investigate the comfort zone of people when a robot approaches them and Torta et al. [135] 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 a number of 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. [117] augment their robot with projection indicators to signal the robot’s intended path. Szafir et al. [130] 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. [118] perform a study in which a robot crosses a humans’ 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. [29] introduce the concept of a “passive demonstration,” in order to disambiguate the intention of a robot’s LED turn signal. Watanabe et al. [144] evaluate a robotic wheelchair that autonomously navigates the environment with and without intention communication. They show that passengers and walking people found intention 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 [94], and from the relationship between head pose and gaze direction [54]. Norman [93] speculates that bicycle riders know how to avoid collisions with pedestrians since pedestrian motion can be predicted by gaze. Similarly, Unhelkar et al. [139] 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. [58] propose a motion planner which 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. [76] 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. [31] 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. [29] show that people are able to adapt to LED-based cues after watching a demonstration of its use, and May et al. [81] present a robot that was able to convey their intention using a mechanical signal but not using a gaze cue. Hart et al. [41] 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 are able to 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. [56] 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 a value in understanding and mediating between the human’s navigational goal and the robot’s.

Murakami et al. [88] propose to smooth a wheelchair’s trajectory to avoid colliding with pedestrians. Kruse et al. [66, 67] investigate classic navigation algorithms that create erratic trajectories near

Table 7: An overview of the different mediation algorithms used in social navigation. **P. vs. R.** is Prevention vs. Resolution, **Role** refers to the robot role, **Obs.** is observability, **Com.** refers to communication, and **Exp. type** is the experiment type.

| Year | Paper                      | P vs. R | Role | # Agents   | Obs.        | Motion Control               | Com. | Exp. Type | Agent Type |
|------|----------------------------|---------|------|------------|-------------|------------------------------|------|-----------|------------|
| 2002 | Murakami et al. [88]       | R       | B    | Abs=2      | RGB + Depth | Other (Hand Coded)           | I    | HE        | H-R        |
| 2005 | Topp and Christensen [134] | R       | R    | Abs=4      | Depth       | Other (Person Tracking)      | N    | HE        | H-R        |
| 2008 | Müller et al. [86]         | R       | R    | Abs=7      | Depth       | Other (A* + Person Tracking) | N    | HE        | H-R        |
| 2010 | Svenstrup et al. [127]     | P       | R    | Abs=39     | Full        | Other (Modified RRT)         | N    | Sim       | H-R        |
| 2013 | Ferrer et al. [30]         | P       | R    | Abs=10     | Depth       | SFM                          | N    | HE        | H-R        |
| 2013 | Guzzi et al. [39]          | P       | B    | Abs=6      | RGB         | Other (Hand Coded)           | N    | R         | R-R        |
| 2014 | Karamouzas et al. [56]     | P       | R    | D=0.27-2.5 | Full        | Other (Hand Coded)           | N    | Sim       | Hom        |
| 2014 | Kruse et al. [67]          | P       | B    | Abs=2      | Full        | Other (Hand Coded)           | I    | HE + Sur  | H-R        |
| 2017 | Silva and Fraichard [120]  | P       | B    | Abs=2      | Full        | ROS                          | N    | Sim       | Hom        |
| 2019 | Yao et al. [146]           | P       | R    | Abs=6      | RGB + Depth | Other (Geometry based)       | N    | HE        | H-R        |

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 to solve spatial conflicts. One result, for example, is adjusting the robot’s velocity instead of its path. Silva and Fraichard [120] tackle the mediation problem using the notion of motion effort and how it should be shared between the robot and the person in order 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 our simulated trials. Svenstrup et al. [127] 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.

A different line of research combines social navigation and person following. This combination can work in several directions: both Müller et al. [86], Topp and Christensen [134] present collision avoidance algorithms that are utilized in the context of following one particular person through a populated environment. Alternatively, in Yao et al. [146], 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.

## 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 with respect to 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 other. 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. **Conflicts Count** is one of the most common approaches to estimate 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 [86] and how many targets it was able to reach in total [39]. What consists of a desirable outcome varies between researchers. For example, whether the goal of an algorithm is conflict **Prevention vs. Resolution** will determine whether a situation where the robot slowed down and changed its heading direction is a successful interaction. For prevention, reaching such a conflict is a failure; for resolution it is just the beginning of the interaction. This measure is also 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 as a complementary one for conflicts count, to account for cases such as a robot that can reach its goal quickly, but will collide walls most of the times. As a reference point, the robot’s speed is usually compared to the average speed of pedestrians ( $1.3 \pm 0.2$  m/s), but this value depends on whether they walk alone or in a group, as group size affects speed more than density level [85]. Gérin-Lajoie et al. [34] measured similar results for natural walking around dynamic obstacles ( $1.44 \pm 0.17$  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. [30]. This metric measures the total work done by the robot, and the summation of the work done by each person in the scene. Kanazawa et al. [53] looked at 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) you 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 had to replan or adjust to avoid a conflict. This metric is highly affected by whether the task is **Prevention**

**vs. Resolution** of conflicts, and also 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 [102], path irregularity (how many unnecessary turns were taken) [39], cumulative heading change [98], and the integral of the square of the curvature to measure the smoothness of a pedestrian’s path [55]. This measure is influenced by whether the task is **Prevention vs. Resolution**, 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 is able to avoid pedestrians from afar is considered more successful than a robot that almost reaches collision [127]. However, this success sometimes creates a tradeoff with the total length of the path the robot needs to take and the smoothness of the path. This metric is highly affected by whether the task is **Prevention vs. Resolution**, the **Robot’s Role**, the **Number of Agents**, and the **Motion Control** used that might have its own predefined distance-keeping restrictions.

**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 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 [94], and the Gaze-Movement Angle (GMA) [23]. Kitazawa and Fujiyama [63] investigated gaze patterns in a collision avoidance scenario with multiple pedestrians moving in a wide hallway shape area. They show that pedestrians pay much more attention to ground surface to detect potential immediate environmental hazards than fixating on obstacles, that most their fixations fall within a cone-shape area rather than semicircle, and that 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. [62] leveraged people’s reliance on such cues and incorporated similar body rotations into an omni-directional 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**.

## 5.2 Emergent Behaviors

Several experiments have been designed to identify specific movement patterns and flow patterns that emerge during execution of social navigation algorithms, or to mimic human movement patterns that emerge in these contexts [5, 73]. In many cases, these patterns are in the form of lanes [44] or

group clusters.



Figure 4: Levels of Service from A to F: How crowded is the environment (taken from Fruin [33])

lessly join such a group [89], bypass it [128], or disperse it [20]. This measure is affected by the **Number of Agents** and the **Agent Type**.

**Maximal Density** is a metric used a lot 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 [33] proposed to identify 6 levels of crowdness, which they refer to as *Level of Service*, as depicted in Figure 4. 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  $m^2$  [85]. 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.

### 5.3 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.

**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 [140], and learning (in data-driven approaches), which is usually measured in learning episodes for achieving a desired behavior [25]. Computation time is highly influenced by whether the task is **Prevention vs. Resolution** as resolution usually requires shorter timescales, **Number of Agents**, **Experiment Type** and **Agent Type**.

**Lane Emergence** is a phenomenon that exists in human crowds - whenever an environment becomes crowded enough, it is likely that people will follow the path of others who are going in the same direction [35, 146]. For several algorithms deployed in crowded environments, the researchers were able to detect the emergence of lanes in 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 [85], and are investigating whether a robot can seam-

**Model Prediction** is a crucial part of every social navigation interaction: in order to properly act, the robot should first be able to predict accurately 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 [69, 83], while others evaluate the prediction of pedestrian trajectories interleaved with robot execution [6, 92]. This metric is influenced by the **Robot’s Role** in the interaction, **Observability**, and **Agent Type**.

## 5.4 Other Evaluations

So far, all evaluation metrics were objective and could usually be quantitatively evaluated. Some contributions focus on analyzing an interaction and identifying theoretical concepts, thus have no empirical evaluation, while others test subjective quantities (e.g. comfort level) or provide a qualitative evaluation of an interaction.

**Survey Questions** is 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 comfort levels during the interaction [48, 81, 141], social presence [58, 99], expectation matching [35, 65], and more. With respect to comfort, Torta et al. [135] identify specific values for this comfort zone (182 cm from a sitting person and 173 cm from a standing person). Syrdal et al. [129] present an empirical evaluation of the role of video prototyping and evocation as a good way to evaluate non-functional aspects of HRI. Another type of subjective evaluation is of proxemics [98, 127], which is related to avoidance distance that was discussed earlier, but can encompass additional information about the interaction. For example, Hall [40] identifies different interaction ranges: Intimate space (up to 0.45m), personal space (1.2m), social space (3.6m), and public space (7.6m). When mapping these distances to human-robot interactions, the comfortable distance from a robot is 0.2m, and arrival tolerance 0.5m [17, 68]. A survey is also referred in this survey as an **Experiment Type**, hence this is the most related attribute.

**No Evaluation** is a category designated for papers that make only a theoretical contribution, such as classifying different abstract types of interactions [109] or ones that provide only a qualitative analysis of an interaction [134]. Accordingly, research with no empirical evaluation might be affected by all attributes of the taxonomy, depending on the subject of the analysis.

## 5.5 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. [11] propose a framework for the analysis of human trajectories, and show that human plan their navigation trajectory in a similar fashion when walking past a robot or a human.

Simulations are commonly used for the evaluation of 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 simulations that are available to use and can provide comparative evaluation for new contributions. Loscos et al. [73] created a rule-based simulation that can handle up to 10,000 pedestrians in an urban environment. Treuille et al.



Table 8: An overview of the different metrics used to evaluate a social interaction.

| Evaluation Type        | Metric Evaluated          | Relevant Works   |
|------------------------|---------------------------|--|
| Interaction Properties | Conflicts Count           | Murakami et al. [88], Pacchierotti et al. [99], Müller et al. [86], Kirby et al. [60], Svenstrup et al. [127], Tamura et al. [132], Diego and Arras [24], Bandyopadhyay et al. [3], Park et al. [102], Ma et al. [77], Guzzi et al. [39], Unhelkar et al. [139], Godoy et al. [36], Okal and Arras [98], Kretzschmar et al. [65], Khambhaita et al. [58], Fernandez et al. [29], Li et al. [70], Everett et al. [28], Ding et al. [25], Long et al. [72], Lynch et al. [76], Jiang et al. [49], Yao et al. [146], Jin et al. [51], Meng et al. [83], Chen et al. [17], Hart et al. [41], Liang et al. [71] |
|                        | Speed                     | Helbing and Molnar [44], Gérin-Lajoie et al. [34], Karamouzas et al. [55], Moussaïd et al. [84], Kruse et al. [67], Unhelkar et al. [139], Kretzschmar et al. [65], Long et al. [72], Liang et al. [71]  |
|                        | Path Time                 | Helbing and Molnar [44], Pacchierotti et al. [99], Karamouzas et al. [55], Foka and Trahanias [32], Bandyopadhyay et al. [3], Ferrer et al. [30], Godoy et al. [36], Chen et al. [18], Chen et al. [19], Tai et al. [131], Everett et al. [28], Ding et al. [25], Long et al. [72], Jiang et al. [49], Jin et al. [51], Kanazawa et al. [53], Chen et al. [17], Liang et al. [71]  |
|                        | Path Length               | Helbing and Molnar [44], Pacchierotti et al. [99], Karamouzas et al. [55], Henry et al. [45], Luber et al. [75], Rios-Martinez et al. [110], Lu et al. [74], Vasquez et al. [141], Okal and Arras [98], Ding et al. [25], Jiang et al. [49], Long et al. [72], Nardi and Stachniss [92], Liang et al. [71]   |
|                        | Acceleration              | Helbing and Molnar [44], Bonneaud and Warren [8]   |
|                        | Avoidance Distance        | Luber et al. [75], Kruse et al. [67], May et al. [81], Kim and Pineau [59], Kretzschmar et al. [65], Chen et al. [18], Tai et al. [131], Lynch et al. [76], Jin et al. [51], Kanazawa et al. [53],   |
|                        | Smoothness                | Helbing and Molnar [44], Gockley et al. [35], Karamouzas et al. [55], Park et al. [102], Guzzi et al. [39], Vasquez et al. [141], Karamouzas et al. [56], Okal and Arras [98]  |
| Robot / Human Actions  | Degrees Turned            | Helbing and Molnar [44], Karamouzas et al. [55], Bonneaud and Warren [8], Truong and Ngo [137]   |
|                        | Gaze Fixations            | Nummenmaa et al. [94], Kitazawa and Fujiyama [63]  |
|                        | Gaze-Movement Angle       | Murakami et al. [88], Pacchierotti et al. [99], Müller et al. [86], Kirby et al. [60], Svenstrup et al. [127], Diego and Arras [24], Bandyopadhyay et al. [3], Ratsamee et al. [107], Park et al. [102], Ma et al. [77], Guzzi et al. [39], Unhelkar et al. [139], Godoy et al. [36], Okal and Arras [98], Kretzschmar et al. [65], Khambhaita et al. [58], Fernandez et al. [29], Li et al. [70], Everett et al. [28], Ding et al. [25], Long et al. [72], Lynch et al. [76], Yao et al. [146], Jin et al. [51], Meng et al. [83], Chen et al. [17], Hart et al. [41], Liang et al. [71]                  |
|                        | Head Orientation          | Patla et al. [103], Ratsamee et al. [107], Unhelkar et al. [139]   |
|                        | Body Position             | Patla et al. [103], Unhelkar et al. [139]  |
|                        |                           |  |
| Emergent Behaviors     | Lane Emergence            | Helbing and Molnar [44], Bennewitz et al. [5], Loscos et al. [73], Van Den Berg et al. [140], Karamouzas et al. [56]   |
|                        | Group Formation           | Musse and Thalmann [89], Moussaïd et al. [85], Swofford et al. [128]   |
|                        | Maximal density           | Bandyopadhyay et al. [3], Ma et al. [77], Mead and Mataric [82]  |
| Algorithmic Properties | Computation Time          | Sisbot et al. [122], Moussaïd et al. [85], Van Den Berg et al. [140], Silva and Fraichard [120], Ding et al. [25]  |
|                        | Model Prediction          | Kuderer et al. [69], Okal and Arras [97], Kim and Pineau [59], Kretzschmar et al. [65], Bera et al. [6], Silva and Fraichard [120], Yao et al. [146], Nardi and Stachniss [92], Meng et al. [83]   |
| Other                  | Survey Questions          | Jeffrey and Mark [48], Murakami et al. [88], Pacchierotti et al. [99], Gockley et al. [35], Svenstrup et al. [127], Vasquez et al. [141], Kruse et al. [67], May et al. [81], Watanabe et al. [144], Szafir et al. [130], Okal and Arras [98], Kretzschmar et al. [65], Khambhaita et al. [58], Chen et al. [18], Baraka and Veloso [4], Shrestha et al. [117], Senft et al. [115]   |
|                        | No Interaction Evaluation | Reynolds [109], Strassner and Langer [126], Topp and Christensen [134], Ohki et al. [96], Pandey and Alami [100], O'Callaghan et al. [95], Gómez et al. [38], Papadakis et al. [101], Charalampous et al. [15]   |

[136] offered a real-time crowd model based on continuum dynamics, which can facilitate large-scale simulations for navigation. Heïgeas et al. [43] presented a simulation platform where pedestrians act according to a physics-based particle force interaction model. Recently, Khambhaita et al. [58] created a simulated benchmark for social navigation tasks instead of physical experiments. This simulation is implemented with openAI Gym. Tsoi et al. [138] 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 goal, collisions with pedestrians, and closest distance to pedestrians.

Pertaining to real world interactions, there are not many contributions that can generalize due to several reasons: 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 (e.g. 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 [129]. In addition, real human-robot interactions 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. [79] recently published a testbed 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 Optimal Reciprocal Collision Avoidance (ORCA), a local collision-free motion planner for a large number of robots as proposed by Van Den Berg et al. [140] 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 [80]. These two chosen navigational strategies are agnostic to the fact that the other agent is a human. This assumption leaves an opportunity to investigate this problem further.

## 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 surveyed a comprehensive list of papers that contribute to social navigation and discussed them according to their values given our taxonomy (Sections 3 and 4). We then surveyed the different measurements used to evaluate an interaction in this context, and highlighted the relations between these measurements and the taxonomy attributes (Section 5).

Social navigation is a growing research area and we expect that while 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 to 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 an 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, the things to verify are:

1. **Taxonomy** Identify the values your work has with respect to the taxonomy’s attributes in this survey: Prevention vs. Resolution, Robot’s Role, Number of Agents, Observability, Motion Control, Communication, Experiment Type, and Agent type.
2. **Reliability** Provide as many details as possible about the choices made in the design of the robot, and about the implementation details.
3. **Human Presence** If your work consists of an interaction with pedestrians, what is their level of familiarity with the robot prior to the interaction?
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.
6. **Evaluation** Detail which metrics will be used to evaluate this success, and what values are these metrics expected to have.

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 [20, 87, 108], 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 an 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 charging into new problem domains.

There are numerous open problems related to social navigation, in which there are specific potential advances that are within reach, given our current understanding and technological abilities: standardization of evaluation metrics and domains, context-aware navigation (e.g., workday vs. weekend), group understanding (avoid collision with a group participant), and adaptive navigation via ML (lifelong learning). Each of these problems offers many opportunities that leverage recent advances in machine learning, robotics, and human-robot interactions and bring them into implementation in a social navigation context. For those interested in contributing to this research area, the above problems are a promising starting point. More information about these problems can be found in Subsection 2.2.

To conclude, we expect the field of social navigation to gain increasing interest and real-world applications in the next decade. To be prepared for these new challenges, we present this survey as a fundamental contribution to this research area, that will help in mapping the existing knowledge in social navigation.

## References

- [1] Admoni, H., Bank, C., Tan, J., Toneva, M., and Scassellati, B. (2011). Robot gaze does not reflexively cue human attention. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 33.
- [2] Admoni, H. and Scassellati, B. (2017). Social eye gaze in human-robot interaction: a review. *Journal of Human-Robot Interaction*, 6(1):25–63.
- [3] Bandyopadhyay, T., Won, K. S., Frazzoli, E., Hsu, D., Lee, W. S., and Rus, D. (2013). Intention-aware motion planning. In *Algorithmic foundations of robotics X*, pages 475–491. Springer.
- [4] Baraka, K. and Veloso, M. M. (2018). Mobile service robot state revealing through expressive lights: Formalism, design, and evaluation. *International Journal of Social Robotics*, 10(1):65–92.
- [5] Bennewitz, M., Burgard, W., and Thrun, S. (2002). Learning motion patterns of persons for mobile service robots. In *Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No. 02CH37292)*, volume 4, pages 3601–3606. IEEE.
- [6] Bera, A., Randhavane, T., Prinja, R., and Manocha, D. (2017). Sociosense: Robot navigation amongst pedestrians with social and psychological constraints. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 7018–7025. IEEE.
- [7] Bonabeau, E., Dorigo, M., Marco, D. d. R. D. F., Theraulaz, G., Theraulaz, G., et al. (1999). *Swarm intelligence: from natural to artificial systems*. Number 1. Oxford university press.
- [8] Bonneaud, S. and Warren, W. H. (2014). An empirically-grounded emergent approach to modeling pedestrian behavior. In *Pedestrian and evacuation dynamics 2012*, pages 625–638. Springer.
- [9] Butler, J. T. and Agah, A. (2001). Psychological effects of behavior patterns of a mobile personal robot. *Autonomous Robots*, 10(2):185–202.
- [10] Carel, W. (1961). Visual factors in the contact analog (publication r61elc60). *Ithaca, NY: General Electric Company Advanced Electronics Center*.
- [11] Carton, D., Olszowy, W., and Wollherr, D. (2016). Measuring the effectiveness of readability for mobile robot locomotion. *International Journal of Social Robotics*, 8(5):721–741.
- [12] Cassandra, A. R., Kaelbling, L. P., and Kurien, J. A. (1996). Acting under uncertainty: Discrete bayesian models for mobile-robot navigation. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS’96*, volume 2, pages 963–972. IEEE.
- [13] Castiello, U. (2003). Understanding other people’s actions: Intention and attention. *Journal of Experimental Psychology: Human Perception and Performance*, 29(2):416.
- [14] Cha, E., Kim, Y., Fong, T., Mataric, M. J., et al. (2018). A survey of nonverbal signaling methods for non-humanoid robots. *Foundations and Trends® in Robotics*, 6(4):211–323.
- [15] Charalampous, K., Emmanouilidis, C., and Gasteratos, A. (2014). Social mapping on rgb-d scenes. In *2014 IEEE International Conference on Imaging Systems and Techniques (IST) Proceedings*, pages 398–403. IEEE.

- [16] Charalampous, K., Kostavelis, I., and Gasteratos, A. (2017). Recent trends in social aware robot navigation: A survey. *Robotics and Autonomous Systems*, 93:85–104.
- [17] Chen, C., Liu, Y., Kreiss, S., and Alahi, A. (2019). Crowd-robot interaction: Crowd-aware robot navigation with attention-based deep reinforcement learning. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 6015–6022. IEEE.
- [18] Chen, Y. F., Everett, M., Liu, M., and How, J. P. (2017a). Socially aware motion planning with deep reinforcement learning. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1343–1350. IEEE.
- [19] Chen, Y. F., Liu, M., Everett, M., and How, J. P. (2017b). Decentralized non-communicating multiagent collision avoidance with deep reinforcement learning. In *2017 IEEE international conference on robotics and automation (ICRA)*, pages 285–292. IEEE.
- [20] Chen, Z., Fan, T., Zhao, X., Liang, J., Shen, C., Chen, H., Manocha, D., Pan, J., and Zhang, W. (2021). Autonomous social distancing in urban environments using a quadruped robot. *IEEE Access*, 9:8392–8403.
- [21] Coleman, J. S. and James, J. (1961). The equilibrium size distribution of freely-forming groups. *Sociometry*, 24(1):36–45.
- [22] Cui, Y., Zhang, Q., Allievi, A., Stone, P., Niekum, S., and Knox, W. B. (2020). The empathic framework for task learning from implicit human feedback. *arXiv preprint arXiv:2009.13649*.
- [23] Cutting, J. E., Vishton, P. M., and Braren, P. A. (1995). How we avoid collisions with stationary and moving objects. *Psychological review*, 102(4):627.
- [24] Diego, G. and Arras, T. K. O. (2011). Please do not disturb! minimum interference coverage for social robots. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1968–1973. IEEE.
- [25] Ding, W., Li, S., Qian, H., and Chen, Y. (2018). Hierarchical reinforcement learning framework towards multi-agent navigation. In *2018 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 237–242. IEEE.
- [26] Dragan, A. D., Lee, K. C., and Srinivasa, S. S. (2013). Legibility and predictability of robot motion. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 301–308. IEEE.
- [27] Ess, A., Leibe, B., Schindler, K., and Van Gool, L. (2009). Moving obstacle detection in highly dynamic scenes. In *2009 IEEE International Conference on Robotics and Automation*, pages 56–63. IEEE.
- [28] Everett, M., Chen, Y. F., and How, J. P. (2018). Motion planning among dynamic, decision-making agents with deep reinforcement learning. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3052–3059. IEEE.
- [29] Fernandez, R., John, N., Kirmani, S., Hart, J., Sinapov, J., and Stone, P. (2018). Passive demonstrations of light-based robot signals for improved human interpretability. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 234–239. IEEE.

- [30] Ferrer, G., Garrell, A., and Sanfeliu, A. (2013). Social-aware robot navigation in urban environments. In *2013 European Conference on Mobile Robots*, pages 331–336. IEEE.
- [31] Fiore, S. M., Wiltshire, T. J., Lobato, E. J., Jentsch, F. G., Huang, W. H., and Axelrod, B. (2013). Toward understanding social cues and signals in human–robot interaction: effects of robot gaze and proxemic behavior. *Frontiers in psychology*, 4:859.
- [32] Foka, A. F. and Trahanias, P. E. (2010). Probabilistic autonomous robot navigation in dynamic environments with human motion prediction. *International Journal of Social Robotics*, 2(1):79–94.
- [33] Fruin, J. J. (1971). *Designing for pedestrians: A level-of-service concept*. Number HS-011 999.
- [34] Gérin-Lajoie, M., Richards, C. L., Fung, J., and McFadyen, B. J. (2008). Characteristics of personal space during obstacle circumvention in physical and virtual environments. *Gait & posture*, 27(2):239–247.
- [35] Gockley, R., Forlizzi, J., and Simmons, R. (2007). Natural person-following behavior for social robots. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, pages 17–24.
- [36] Godoy, J., Karamouzas, I., Guy, S. J., and Gini, M. L. (2016). Moving in a crowd: Safe and efficient navigation among heterogeneous agents. In *IJCAI*, pages 294–300.
- [37] Goffman, E. (2008). *Behavior in public places*. Simon and Schuster.
- [38] Gómez, J. V., Mavridis, N., and Garrido, S. (2014). Fast marching solution for the social path planning problem. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1871–1876. IEEE.
- [39] Guzzi, J., Giusti, A., Gambardella, L. M., Theraulaz, G., and Di Caro, G. A. (2013). Human-friendly robot navigation in dynamic environments. In *2013 IEEE International Conference on Robotics and Automation*, pages 423–430. IEEE.
- [40] Hall, E. T. (1966). *The hidden dimension*, volume 609. Garden City, NY: Doubleday.
- [41] Hart, J., Mirsky, R., Xiao, X., Tejada, S., Mahajan, B., Goo, J., Baldauf, K., Owen, S., and Stone, P. (2020). Using human-inspired signals to disambiguate navigational intentions. In *International Conference on Social Robotics*, pages 320–331. Springer.
- [42] Hayduk, L. A. (1981). The shape of personal space: An experimental investigation. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 13(1):87.
- [43] Heïgeas, L., Luciani, A., Thollot, J., and Castagné, N. (2010). A physically-based particle model of emergent crowd behaviors. *arXiv preprint arXiv:1005.4405*.
- [44] Helbing, D. and Molnar, P. (1995). Social force model for pedestrian dynamics. *Physical review E*, 51(5):4282.
- [45] Henry, P., Vollmer, C., Ferris, B., and Fox, D. (2010). Learning to navigate through crowded environments. In *2010 IEEE International Conference on Robotics and Automation*, pages 981–986. IEEE.

- [46] Holman, B., Anwar, A., Singh, A., Tec, M., Hart, J., and Stone, P. (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)*, pages 6183–6190. IEEE.
- [47] Hoogendoorn, S. and Bovy, P. (2003). Simulation of pedestrian flows by optimal control and differential games. *Optimal control applications and methods*, 24(3):153–172.
- [48] Jeffrey, P. and Mark, G. (1998). Constructing social spaces in virtual environments: A study of navigation and interaction. In *Workshop on personalised and social navigation in information space*, pages 24–38. Stockholm: Swedish Institute of Computer Science.
- [49] Jiang, Y.-S., Warnell, G., Munera, E., and Stone, P. (2018a). 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.
- [50] Jiang, Y.-S., Warnell, G., and Stone, P. (2018b). Inferring user intention using gaze in vehicles. In *The 20th ACM International Conference on Multimodal Interaction (ICMI)*, Boulder, Colorado.
- [51] Jin, J., Nguyen, N. M., Sakib, N., Graves, D., Yao, H., and Jagersand, M. (2019). Mapless navigation among dynamics with social-safety-awareness: a reinforcement learning approach from 2d laser scans. *arXiv preprint arXiv:1911.03074*.
- [52] Kaelbling, L. P. (2020). The foundation of efficient robot learning. *Science*, 369(6506):915–916.
- [53] Kanazawa, A., Kinugawa, J., and Kosuge, K. (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):817–832.
- [54] Kar, A. and Corcoran, P. (2017). A review and analysis of eye-gaze estimation systems, algorithms and performance evaluation methods in consumer platforms. *IEEE Access*, 5:16495–16519.
- [55] Karamouzas, I., Heil, P., Van Beek, P., and Overmars, M. H. (2009). A predictive collision avoidance model for pedestrian simulation. In *International workshop on motion in games*, pages 41–52. Springer.
- [56] Karamouzas, I., Skinner, B., and Guy, S. J. (2014). Universal power law governing pedestrian interactions. *Physical review letters*, 113(23):238701.
- [57] Karunaratne, D., Morales, Y., Kanda, T., and Ishiguro, H. (2018). Model of side-by-side walking without the robot knowing the goal. *International Journal of Social Robotics*, 10(4):401–420.
- [58] Khambhaita, H., Rios-Martinez, J., and Alami, R. (2016). Head-body motion coordination for human aware robot navigation. In *9th International workshop on Human-Friendly Robotics (HFR 2016)*, page 8p.
- [59] Kim, B. and Pineau, J. (2016). Socially adaptive path planning in human environments using inverse reinforcement learning. *International Journal of Social Robotics*, 8(1):51–66.

- [60] Kirby, R., Simmons, R., and Forlizzi, J. (2009). Companion: A constraint-optimizing method for person-acceptable navigation. In *RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication*, pages 607–612. IEEE.
- [61] Kit, D. M. (2012). *Change detection models for mobile cameras*. PhD thesis.
- [62] Kitagawa, R., Liu, Y., and Kanda, T. (2021). Human-inspired motion planning for omnidirectional social robots. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '21, page 34–42, New York, NY, USA. Association for Computing Machinery.
- [63] Kitazawa, K. and Fujiyama, T. (2010). Pedestrian vision and collision avoidance behavior: Investigation of the information process space of pedestrians using an eye tracker. In *Pedestrian and evacuation dynamics 2008*, pages 95–108. Springer.
- [64] Kostavelis, I. and Gasteratos, A. (2015). Semantic mapping for mobile robotics tasks: A survey. *Robotics and Autonomous Systems*, 66:86–103.
- [65] Kretschmar, H., Spies, M., Sprunk, C., and Burgard, W. (2016). Socially compliant mobile robot navigation via inverse reinforcement learning. *The International Journal of Robotics Research*, 35(11):1289–1307.
- [66] Kruse, T., Basili, P., Glasauer, S., and Kirsch, A. (2012). Legible robot navigation in the proximity of moving humans. In *2012 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO)*, pages 83–88. IEEE.
- [67] Kruse, T., Kirsch, A., Khambhaita, H., and Alami, R. (2014). Evaluating directional cost models in navigation. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 350–357.
- [68] Kruse, T., Pandey, A. K., Alami, R., and Kirsch, A. (2013). Human-aware robot navigation: A survey. *Robotics and Autonomous Systems*, 61(12):1726–1743.
- [69] Kuderer, M., Kretschmar, H., Sprunk, C., and Burgard, W. (2012). Feature-based prediction of trajectories for socially compliant navigation. In *Robotics: science and systems*.
- [70] Li, M., Jiang, R., Ge, S. S., and Lee, T. H. (2018). Role playing learning for socially concomitant mobile robot navigation. *CAAI Transactions on Intelligence Technology*, 3(1):49–58.
- [71] Liang, J., Patel, U., Sathiamoorthy, A. J., and Manocha, D. (2020). Crowdsteer: Realtime smooth and collision-free robot navigation in dense crowd scenarios trained using high-fidelity simulation. *International Joint Conference on AI*.
- [72] Long, P., Fanl, T., Liao, X., Liu, W., Zhang, H., and Pan, J. (2018). Towards optimally decentralized multi-robot collision avoidance via deep reinforcement learning. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6252–6259. IEEE.
- [73] Loscos, C., Marchal, D., and Meyer, A. (2003). Intuitive crowd behavior in dense urban environments using local laws. In *Proceedings of Theory and Practice of Computer Graphics, 2003.*, pages 122–129. IEEE.



- [74] Lu, D. V., Allan, D. B., and Smart, W. D. (2013). Tuning cost functions for social navigation. In *International Conference on Social Robotics*, pages 442–451. Springer.
- [75] Lubner, M., Spinello, L., Silva, J., and Arras, K. O. (2012). Socially-aware robot navigation: A learning approach. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 902–907. IEEE.
- [76] Lynch, S. D., Pettré, J., Bruneau, J., Kulpa, R., Crétual, A., and Olivier, A.-H. (2018). Effect of virtual human gaze behaviour during an orthogonal collision avoidance walking task. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 136–142. IEEE.
- [77] Ma, J., Lo, S. M., Song, W., Wang, W., Zhang, J., and Liao, G. (2013). Modeling pedestrian space in complex building for efficient pedestrian traffic simulation. *Automation in Construction*, 30:25–36.
- [78] Martins, G. S., Rocha, R. P., Pais, F. J., and Menezes, P. (2019). Clusternav: Learning-based robust navigation operating in cluttered environments. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 9624–9630. IEEE.
- [79] Mavrogiannis, C., Hutchinson, A. M., Macdonald, J., Alves-Oliveira, P., and Knepper, R. A. (2019). Effects of distinct robot navigation strategies on human behavior in a crowded environment. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 421–430. IEEE.
- [80] Mavrogiannis, C. I., Thomason, W. B., and Knepper, R. A. (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*, pages 361–369.
- [81] May, A. D., Dondrup, C., and Hanheide, M. (2015). Show me your moves! conveying navigation intention of a mobile robot to humans. In *2015 European Conference on Mobile Robots (ECMR)*, pages 1–6. IEEE.
- [82] Mead, R. and Matarić, M. J. (2016). Perceptual models of human-robot proxemics. In *Experimental Robotics*, pages 261–276. Springer.
- [83] Meng, X., Ratliff, N., Xiang, Y., and Fox, D. (2020). Scaling local control to large-scale topological navigation. *IEEE International Conference on Robotics and Automation (ICRA)*.
- [84] Moussaïd, M., Helbing, D., and Theraulaz, G. (2011). How simple rules determine pedestrian behavior and crowd disasters. *Proceedings of the National Academy of Sciences*, 108(17):6884–6888.
- [85] Moussaïd, M., Perozo, N., Garnier, S., Helbing, D., and Theraulaz, G. (2010). The walking behaviour of pedestrian social groups and its impact on crowd dynamics. *PloS one*, 5(4):e10047.
- [86] Müller, J., Stachniss, C., Arras, K. O., and Burgard, W. (2008). Socially inspired motion planning for mobile robots in populated environments. In *Proc. of International Conference on Cognitive Systems*.
- [87] Murakami, H., Feliciani, C., Nishiyama, Y., and Nishinari, K. (2021). Mutual anticipation can contribute to self-organization in human crowds. *Science Advances*, 7(12):eabe7758.

- [88] Murakami, Y., Kuno, Y., Shimada, N., and Shirai, Y. (2002). Collision avoidance by observing pedestrians’ faces for intelligent wheelchairs. *Journal of the Robotics Society of Japan*, 20(2):206–213.
- [89] Musse, S. R. and Thalmann, D. (1997). A model of human crowd behavior: Group inter-relationship and collision detection analysis. In *Computer Animation and Simulation’97*, pages 39–51. Springer.
- [90] Nakauchi, Y. and Simmons, R. (2002). A social robot that stands in line. *Autonomous Robots*, 12(3):313–324.
- [91] Narayanan, V., Phillips, M., and Likhachev, M. (2012). Anytime safe interval path planning for dynamic environments. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4708–4715. IEEE.
- [92] Nardi, L. and Stachniss, C. (2020). Long-term robot navigation in indoor environments estimating patterns in traversability changes. *ICRA*.
- [93] Norman, D. (2009). *The design of future things*. Basic books.
- [94] Nummenmaa, L., Hyönä, J., and Hietanen, J. K. (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):1454–1458.
- [95] O’Callaghan, S. T., Singh, S. P., Alempijevic, A., and Ramos, F. T. (2011). Learning navigational maps by observing human motion patterns. In *2011 IEEE International Conference on Robotics and Automation*, pages 4333–4340. IEEE.
- [96] Ohki, T., Nagatani, K., and Yoshida, K. (2010). Collision avoidance method for mobile robot considering motion and personal spaces of evacuees. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1819–1824. IEEE.
- [97] Okal, B. and Arras, K. O. (2014). Towards group-level social activity recognition for mobile robots. In *IROS Assistance and Service Robotics in a Human Environments Workshop*.
- [98] Okal, B. and Arras, K. O. (2016). Learning socially normative robot navigation behaviors with bayesian inverse reinforcement learning. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2889–2895. IEEE.
- [99] Pacchierotti, E., Christensen, H. I., and Jensfelt, P. (2006). Design of an office-guide robot for social interaction studies. In *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4965–4970. IEEE.
- [100] Pandey, A. K. and Alami, R. (2010). A framework towards a socially aware mobile robot motion in human-centered dynamic environment. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5855–5860. IEEE.
- [101] Papadakis, P., Rives, P., and Spalanzani, A. (2014). Adaptive spacing in human-robot interactions. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2627–2632. IEEE.

- [102] Park, J. H., Rojas, F. A., and Yang, H. S. (2013). A collision avoidance behavior model for crowd simulation based on psychological findings. *Computer Animation and Virtual Worlds*, 24(3-4):173–183.
- [103] Patla, A. E., Adkin, A., and Ballard, T. (1999). Online steering: coordination and control of body center of mass, head and body reorientation. *Experimental Brain Research*, 129(4):629–634.
- [104] Pfeiffer, M., Schwesinger, U., Sommer, H., Galceran, E., and Siegwart, R. (2016). Predicting actions to act predictably: Cooperative partial motion planning with maximum entropy models. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2096–2101. IEEE.
- [105] Povinelli, D. J., Bierschwale, D. T., and Cech, C. G. (1999). Comprehension of seeing as a referential act in young children, but not juvenile chimpanzees. *British Journal of Developmental Psychology*, 17(1):37–60.
- [106] Prassler, E., Bank, D., Kluge, B., and Hagele, M. (2002). Key technologies in robot assistants: Motion coordination between a human and a mobile robot. *Transactions on Control, Automation and Systems Engineering*, 4(1):56–61.
- [107] Ratsamee, P., Mae, Y., Ohara, K., Takubo, T., and Arai, T. (2013). Human–robot collision avoidance using a modified social force model with body pose and face orientation. *International Journal of Humanoid Robotics*, 10(01):1350008.
- [108] Reig, S., Luria, M., Wang, J. Z., Oltman, D., Carter, E. J., Steinfeld, A., Forlizzi, J., and Zimmerman, J. (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*, pages 289–297.
- [109] Reynolds, C. W. (1999). Steering behaviors for autonomous characters. In *Game developers conference*, volume 1999, pages 763–782. Citeseer.
- [110] Rios-Martinez, J., Renzaglia, A., Spalanzani, A., Martinelli, A., and Laugier, C. (2012). Navigating between people: A stochastic optimization approach. In *2012 IEEE International Conference on Robotics and Automation*, pages 2880–2885. IEEE.
- [111] Rios-Martinez, J., Spalanzani, A., and Laugier, C. (2015). From proxemics theory to socially-aware navigation: A survey. *International Journal of Social Robotics*, 7(2):137–153.
- [112] Saran, A., Majumdar, S., Short, E. S., Thomaz, A., and Niekum, S. (2018). Human gaze following for human-robot interaction. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 8615–8621. IEEE.
- [113] Saran, A., Short, E. S., Thomaz, A., and Niekum, S. (2019). Understanding teacher gaze patterns for robot learning. *arXiv preprint arXiv:1907.07202*.
- [114] Sciutti, A., Bisio, A., Nori, F., Metta, G., Fadiga, L., Pozzo, T., and Sandini, G. (2012). Measuring human-robot interaction through motor resonance. *International Journal of Social Robotics*, 4(3):223–234.

- [115] Senft, E., Satake, S., and Kanda, T. (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*, pages 539–547.
- [116] Shi, C., Satake, S., Kanda, T., and Ishiguro, H. (2018). A robot that distributes flyers to pedestrians in a shopping mall. *International Journal of Social Robotics*, 10(4):421–437.
- [117] Shrestha, M. C., Onishi, T., Kobayashi, A., Kamezaki, M., and Sugano, S. (2018). Communicating directional intent in robot navigation using projection indicators. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 746–751. IEEE.
- [118] Shrestha, M. C., Onishi, T., Kobayashi, A., Kamezaki, M., and Sugano, S. (2018). Communicating directional intent in robot navigation using projection indicators. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 746–751.
- [119] Siegwart, R., Nourbakhsh, I. R., and Scaramuzza, D. (2011). *Introduction to autonomous mobile robots*. MIT press.
- [120] Silva, G. and Fraichard, T. (2017). Human robot motion: A shared effort approach. In *2017 European Conference on Mobile Robots (ECMR)*, pages 1–6. IEEE.
- [121] Singh, R., Miller, T., Newn, J., Velloso, E., Vetere, F., and Sonenberg, L. (2020). Combining gaze and ai planning for online human intention recognition. *Artificial Intelligence*, page 103275.
- [122] Sisbot, E. A., Marin-Urias, L. F., Alami, R., and Simeon, T. (2007). A human aware mobile robot motion planner. *IEEE Transactions on Robotics*, 23(5):874–883.
- [123] Smith, K., Ba, S. O., Odobez, J., and Gatica-Perez, D. (2008). Tracking the visual focus of attention for a varying number of wandering people. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(7):1212–1229.
- [124] Soproni, K., Miklósi, Á., Topál, J., and Csányi, V. (2001). Comprehension of human communicative signs in pet dogs (*canis familiaris*). *Journal of comparative psychology*, 115(2):122.
- [125] Stiefelhagen, R., Finke, M., Yang, J., and Waibel, A. (1999). From gaze to focus of attention. In *International Conference on Advances in Visual Information Systems*, pages 765–772. Springer.
- [126] Strassner, J. and Langer, M. (2005). Virtual humans with personalized perception and dynamic levels of knowledge. *Computer Animation and Virtual Worlds*, 16(3-4):331–342.
- [127] Svenstrup, M., Bak, T., and Andersen, H. J. (2010). Trajectory planning for robots in dynamic human environments. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4293–4298. IEEE.
- [128] Swofford, M., Peruzzi, J., Tsoi, N., Thompson, S., Martín-Martín, R., Savarese, S., and Vázquez, M. (2020). Improving social awareness through dante: Deep affinity network for clustering conversational interactants. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1):1–23.

- [129] Syrdal, D. S., Otero, N., and Dautenhahn, K. (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*, pages 1–8.
- [130] Szaifir, D., Mutlu, B., and Fong, T. (2015). Communicating directionality in flying robots. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, HRI ’15, pages 19–26, New York, NY, USA. ACM.
- [131] Tai, L., Zhang, J., Liu, M., and Burgard, W. (2018). Socially compliant navigation through raw depth inputs with generative adversarial imitation learning. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1111–1117. IEEE.
- [132] Tamura, Y., Fukuzawa, T., and Asama, H. (2010). Smooth collision avoidance in human-robot coexisting environment. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3887–3892. IEEE.
- [133] Thrun, S., Beetz, M., Bennewitz, M., Burgard, W., Cremers, A. B., Dellaert, F., Fox, D., Haehnel, D., Rosenberg, C., Roy, N., et al. (2000). Probabilistic algorithms and the interactive museum tour-guide robot minerva. *The International Journal of Robotics Research*, 19(11):972–999.
- [134] Topp, E. A. and Christensen, H. I. (2005). Tracking for following and passing persons. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2321–2327. IEEE.
- [135] Torta, E., Cuijpers, R. H., and Juola, J. F. (2013). Design of a parametric model of personal space for robotic social navigation. *International Journal of Social Robotics*, 5(3):357–365.
- [136] Treuille, A., Cooper, S., and Popović, Z. (2006). Continuum crowds. *ACM Transactions on Graphics (TOG)*, 25(3):1160–1168. <https://howtorts.github.io/2014/01/09/continuum-crowds.html>.
- [137] Truong, X.-T. and Ngo, T.-D. (2016). Dynamic social zone based mobile robot navigation for human comfortable safety in social environments. *International Journal of Social Robotics*, 8(5):663–684.
- [138] Tsoi, N., Hussein, M., Espinoza, J., Ruiz, X., and Vázquez, M. (2020). Sean: Social environment for autonomous navigation. In *Proceedings of the 8th International Conference on Human-Agent Interaction*, pages 281–283.
- [139] Unhelkar, V. V., Pérez-D’Arpino, C., Stirling, L., and Shah, J. A. (2015). Human-robot co-navigation using anticipatory indicators of human walking motion. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6183–6190. IEEE.
- [140] Van Den Berg, J., Guy, S. J., Lin, M., and Manocha, D. (2011). Reciprocal n-body collision avoidance. In *Robotics research*, pages 3–19. Springer.
- [141] Vasquez, D., Okal, B., and Arras, K. O. (2014). Inverse reinforcement learning algorithms and features for robot navigation in crowds: an experimental comparison. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1341–1346. IEEE.

- [142] Vinciarelli, A., Pantic, M., and Bourlard, H. (2009). Social signal processing: Survey of an emerging domain. *Image and vision computing*, 27(12):1743–1759.
- [143] Walter, W. G. (1950). An imitation of life. *Scientific american*, 182(5):42–45.
- [144] Watanabe, A., Ikeda, T., Morales, Y., Shinozawa, K., Miyashita, T., and Hagita, N. (2015). Communicating robotic navigational intentions. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5763–5769. IEEE.
- [145] Yan, Z., Jouandeau, N., and Cherif, A. A. (2013). A survey and analysis of multi-robot coordination. *International Journal of Advanced Robotic Systems*, 10(12):399.
- [146] Yao, X., Zhang, J., and Oh, J. (2019). Following social groups: Socially compliant autonomous navigation in dense crowds. *arXiv preprint arXiv:1911.12063*.
- [147] Yedidsion, H., Deans, J., Sheehan, C., Chillara, M., Hart, J., Stone, P., and Mooney, R. J. (2019). Optimal use of verbal instructions for multi-robot human navigation guidance. In *International Conference on Social Robotics*, pages 133–143. Springer.