International Journal of Social Robotics (2021) 13:1851–1877
https://doi.org/10.1007/s12369-020-00720-2

Break the Ice: a Survey on Socially Aware Engagement for Human–Robot First Encounters

João Avelino[1] · Leonel Garcia-Marques[2] · Rodrigo Ventura[1] · Alexandre Bernardino[1]

Accepted: 26 October 2020 / Published online: 8 January 2021

Abstract
Society is starting to come up with exciting applications for social robots like butlers, coaches, and waiters. However, these robots face a challenging task: to meet people during a first encounter. This survey explores the literature that contributes to this task. We define a taxonomy based on psychology and sociology models: Kendon's greeting model and Greenspan's model of social competence. We use Kendon's model as a framework to compare and analyze works that describe robotic systems that engage with people. To categorize individual skills, we use three components of Social Awareness that belong to Greenspan's model: Social Sensitivity, Social Insight, and Communication. Under each section, we highlight some research gaps and propose research directions to address them. Through our analysis, we suggest significant research directions for enhanced first encounters. First, social scripts need to be evaluated under equal conditions. Second, interaction management and tracking for first encounters should consider state and observation uncertainties. Third, perception methods need lighter and robust integration in mobile platforms. Fourth, methods to explicitly define social norms are still scarce. Finally, research on social feedback and interaction recovery may fill the gaps of imperfect first encounters.

Keywords Survey · Human–robot interaction · Social robots · First encounters · Social feedback

1 Introduction

Timidly, mobile social robots are starting to appear in social contexts. We define them as embodied agents designed to engage in social interaction that can navigate autonomously in their environment, combining the definitions of social robots [40] and of mobile robots [104]. Contrary to virtual characters on screens, computers, and smartphones, their embodiment allows them to be proactive members of society and to improve human engagement [70,92,116]. It comes as no surprise that industry and academia are exploring the marketing advantages of these systems. For instance, companies and institutions have deployed mobile robotic butlers to approach and guide people in their facilities (SIGA[1] Robots in Santander's headquarters, in Madrid, Spain), greet visitors (Viva[2] robots in Pavilhão do Conhecimento, in Lisbon, Portugal) and serve food and drinks in restaurants and events (for instance, the Ginger[3] robot, in Kathmandu, Nepal). Another important application for these systems is assistance to humans in elderly care centers. Given the unprecedented increasing gap between supply and demand of care services, robots like Vizzy [82], Mbot [129], and GrowMu [91] have been used to help the staff to entertain, persuade, and motivate seniors to participate in activities and physical exercises. Albeit with distinct goals, all these robots share a common

This work was funded with Grant SFRH/BD/133098/2017, from Fundação para a Ciência e a Tecnologia, and supported by the LARSyS - FCT Project UIDB/50009/2020.

✉ João Avelino
javelino@isr.tecnico.ulisboa.pt

Leonel Garcia-Marques
garcia_marques@sapo.pt

Rodrigo Ventura
rodrigo.ventura@isr.tecnico.ulisboa.pt

Alexandre Bernardino
alex@isr.tecnico.ulisboa.pt

[1] Institute for Systems and Robotics, Instituto Superior Técnico, University of Lisbon, Lisbon, Portugal

[2] Faculty of Psychology, University of Lisbon, Lisbon, Portugal

[1] https://www.cnet.com/news/ferrari-red-robots-greet-visitors-to-santander-bank/.

[2] https://www.idmind.pt/presentation-of-robot-viva/.

[3] https://www.euronews.com/2018/11/27/nepal-s-digital-restaurant-where-guests-are-served-by-robots.

task: to meet and engage humans into interaction in a possible first encounter.

This survey's objective is to study the achievements and limitations of robot skills to initiate first encounters. First, we define a taxonomy, models, and necessary social skills based on social cognition literature. Then, we analyze robotic systems on first encounters and relate their implementations to the taxonomy. Considering the proposed taxonomy, we address the state-of-the-art of individual social skills necessary for first encounters, identify research gaps, and provide future directions.

1.1 Human–robot First Encounters and Why They Matter

In the scope of this survey, a first encounter is the first interaction between a physical robot and a human. We are especially interested in situations where the robot has no information about the humans with whom it interacts. We can classify these as Zero Acquaintance Encounters (ZAE) [5] from the perspective of the robot. Zero Acquaintance is defined in the literature as a condition in which the agent/human has never interacted with the target or observed the target in social interaction [5,65].

The first encounter between a robot and a human is the cornerstone for both short-term engagement and long-term interactions. Their potential importance can be drawn from human–human studies that report that first encounters determine the direction of relationships and whether people wish to meet each other afterward [100]. Humans spontaneously start forming impressions and judgments about each other [5], and these impressions can last for a significant time after the encounter [122]. These judgments and impressions are influenced by several powerful effects known in the social cognition literature. For instance, the primacy effect [10] is a phenomenon that biases people into recalling/crediting earlier information more than later information. Thus, people can make negative judgments if a robot misbehaves in the first interaction moments, which will affect their trust in the robot [134]. Another example is the incongruency effect [50,51,119,120], that states that people tend to better recall expectancy-incongruent information than congruent information. Even though these effects relate to the impression formation of humans, researchers have shown that humans evaluate and judge artificial social entities (like robots and virtual characters) as they do with other humans [93,98]. In their recent HRI study, Paetzel et al. [87] observed that participants determined the robot's competence in the first minutes of interaction, and it remained stable over the following sessions, a result that highlights the importance of a first impression in human–robot interaction. Hypothetically, if a human expects a robot to follow certain social norms and it breaks them, the human would strongly recall this event due to both effects, even if the remainder of the interaction was pleasant. Given these insights, it is natural to assume that the design and development of robotic skills that enhance the quality of zero-acquaintance encounters are of the utmost importance for human–robot interaction and trust.

In addition to the previous application-related motivations, this is also a fascinating topic from a scientific point of view. It involves a complex set of perception and action skills, research on how to integrate them in common frameworks, and knowledge from social sciences and human behavior. Definitively a multi-disciplinary challenge.

1.2 Survey Motivation

During ZAE's, the robot needs to be able to understand the social context, perceive signals, express them, and respect social norms. In this context, robots do not have a personalized model of the humans with whom they are going to interact with, but still need to comply with human expectations of social behaviors. These systems need to leverage on the body of knowledge of social sciences and Human–robot interaction studies. It is necessary to understand which skills are involved in the process, how to manage them, and understand their current technological limitations and maturity. To our knowledge, this problem has not been surveyed from this perspective before. Past surveys focused on individual skills, which are challenging research problems themselves. The application of those skills is usually broader than ZAE's.

An example is the ability to manage space during interactions (proxemics) and social navigation, which the robot needs to respect during ZAEs. This skill makes the robot follow the social norm of respecting others' personal space. Rios-Martinez and co-authors [101] surveyed this topic in a thoughtful review of theories and research on social robot navigation for both focused and unfocused interactions.

Communication is another example. It is an essential part of the interaction between social beings during a ZAE since it lets both parties signal their intentions of interacting or not, usually through its nonverbal modalities. Recently, Saunderson et al. [108] surveyed existing works focused on non-verbal communication in human–robot interaction. They studied works under the proxemics, kinesics, haptics, chronemics, and their combinations. They paid attention to both sensing and action, as well as human reactions and perceptions of robots employing these modes.

The final example is that of behavior adaptation. During a ZAE, the robot may need to accommodate to the target of interaction. For instance, if the person displays discomfort with the robot's distance, it should be able to update its belief of "appropriate distance" and act accordingly. This topic has attracted a keen interest in the research community, as reported by Rossi and colleagues in their survey on user profiling and behavioral adaptation [103]. Their classifica-

tion scheme splits both topics into physical, cognitive, and social subdomains. They review cues used to profile people as well as the robotic skills and methods to adapt their behavior to that user profile. A more recent survey from Martins and colleagues [75] explores robot adaptation on non-physical interaction behaviors. They propose a taxonomy that they use to categorize analyzed works under three categories: (i) adaptive systems with no user model, (ii) systems based on static user models, and (iii) systems based on dynamic user models. They cover a large number of works on ongoing interactions between people and robots, mainly during tasks. Ahmad and colleagues [2] surveyed existing works on robot adaptation to human actions. They covered robot adaptation in the following domains: health care and therapy, education, public domains and work environments, and homes.

This survey arises as an attempt to organize available literature and identify gaps and research directions to solve the problem of first encounters. We intend to contribute to the literature by attempting to answer the following question: "How far are social robots from being able to engage with strangers in feedback sensitive and socially acceptable way in first encounters?". We will do so by proposing a taxonomy based on the social cognition literature, using Kendon's model of greetings and Greenspan's model of social awareness. The taxonomy derived from Kendon's model allows us to compare robotic systems in first encounters, which have distinct taxonomies. With Greenspan's model, we categorize and overview the state-of-the-art of required social skills. Our line of work assumes that social robots, like humans, cannot engage people perfectly the whole time, thus needing to be able to understand human feedback and adapt accordingly. With this question in mind, we intend this survey to be a useful asset for researchers that aim to make robots capable of smooth engagement with people and "break the ice" in first interactions while being able to recognize social norm violations and adopt corrective actions.

1.3 Survey Objectives and Scope

With this survey, we intend to study achievements and limitations in socially aware engagement during first encounters between robots and humans. Our focus on zero-acquaintance encounters means that we only cover works that describe robotic systems that meet and open interaction without previously known personalized user models. Thus, the robot has zero-acquaintance with the person and must resort to models of knowledge of social norms and scripts. We will address this subject from the robot's perspective, pinpointing current shortcomings, challenges, and possible research directions. Even though we focus on the technological side, we take advantage of the valuable knowledge reported by interaction studies as well as studies in the areas of psychology and social cognition.

First encounters can be extremely diverse, as a result of multiple robot types and interaction contexts. Here, we focus on mobile social robots that are minimally anthropomorphic. This definition implies that robots need to be able to navigate and have a design that allows them to mimic at least a minor set of human social behaviors. Vizzy, MBOT, Robovie [55], GrowMu, Sanbot, and Pepper are notable examples of such robots (Fig. 1). Our survey assumes social norms play a pivotal role in first encounters, where an agent has no information about the other's preferences.As such, we limit the scope of the survey to interactions with adults and seniors without cognitive impairments and casual social encounters in uncrowded scenes. We assume that most members of this group follow social norms and can recognize when others break them. There is one pivotal moment of human–robot interaction that we examine in this work: the interaction opening set of perception-action iterations that lead to interaction. We do not focus on interactions past this point since they can be remarkably broad, ranging from dialogues to touch interaction. Therefore, these interaction topics should be addressed in individual surveys. As a reference, Mavridis [78] published a review of verbal and non-verbal communication in human–robot conversations. Finally, even though we concentrate on 1-to-1 interaction, a social robot needs to be aware of its surroundings, needing to detect and enter in groups of people, if the target is part of a group.

2 Taxonomy and Survey Organization

The start of a pleasant meeting between people requires them to recognize each other as social entities and be willing to interact. That implies that both agents follow social norms during an interaction. Social norms are so important to humans that people are willing to incur self-costs to punish deviant behavior [39]. Nonetheless, they are informal and can exist with no kind of sanction for someone not following them. Given their importance in the process, we recall the definition proposed by Malle et al. [74].

Definition 1 *Social norm* "... an instruction to (not) perform action A in context C, provided that a sufficient number of individuals in the community (i) indeed follow this instruction and (ii) demand of each other to follow the instruction".

Remark 1 When we refer to social norms throughout our work, we refer to those that occur due to the natural interaction of people and are not enforced by a legal system.

Thus, it is relevant for a social robot to follow appropriate social norms when meeting people, acting according to people's expectations toward socially competent agents. However, knowledge about social norms does not tell the robot how to plan their actions and behave in a specific social

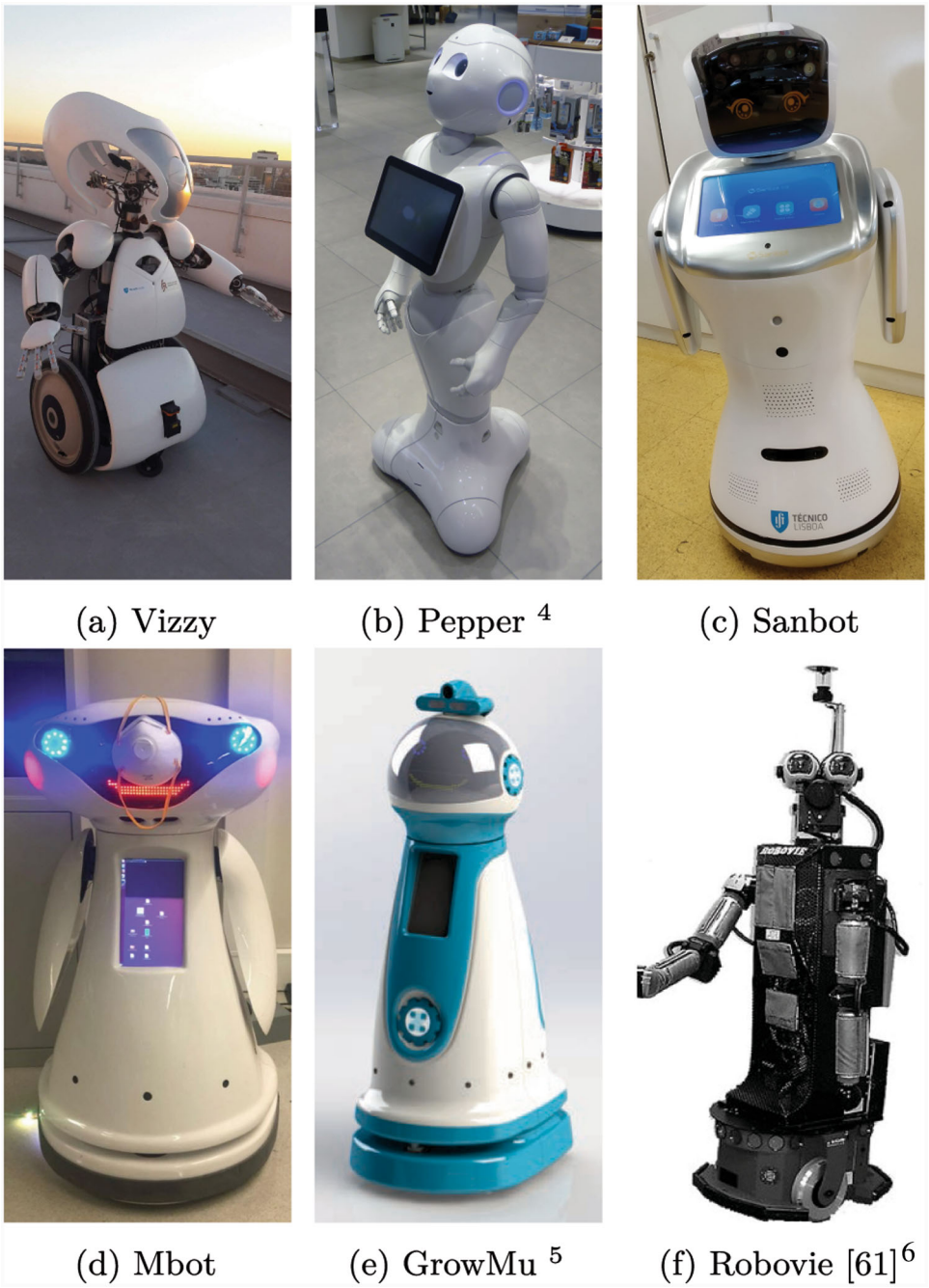

Fig. 1 Examples of minimally anthropomorphic mobile social robots considered in this survey. [4]https://en.wikipedia.org/wiki/Pepper_(robot). [5]https://cordis.europa.eu/project/id/643647/reporting. [6]Robovie developed by ATR

context, like meeting someone. This process is especially challenging during a ZAE since people have no information about each other. Before any interaction, each party will create a visually based impression on the other according to their preconceived beliefs, supported by social norms and cultural information. Yet, these norms might not be sufficient to plan the sequence of appropriate behaviors. Schank [109] claims that people resort to sequential behavioral patterns observed in their community during specific contexts: they follow social scripts. Once people identify the interaction type, they activate a script that embeds social norms and specifies a sequence of actions that humans should perform as the interaction progresses [17]. Social scripts can be simple or complex. Along with this work, we will use the following definition, adapted from [1,52]:

Definition 2 *Social script* a mental construct that contains information about the plans and sequences of actions appropriate and expected from the participants of a social situation.

With these insights in mind, one can ask: have researchers studied social scripts that allow people to infer if others are open for engagement? Indeed, Kendon [64] observed that

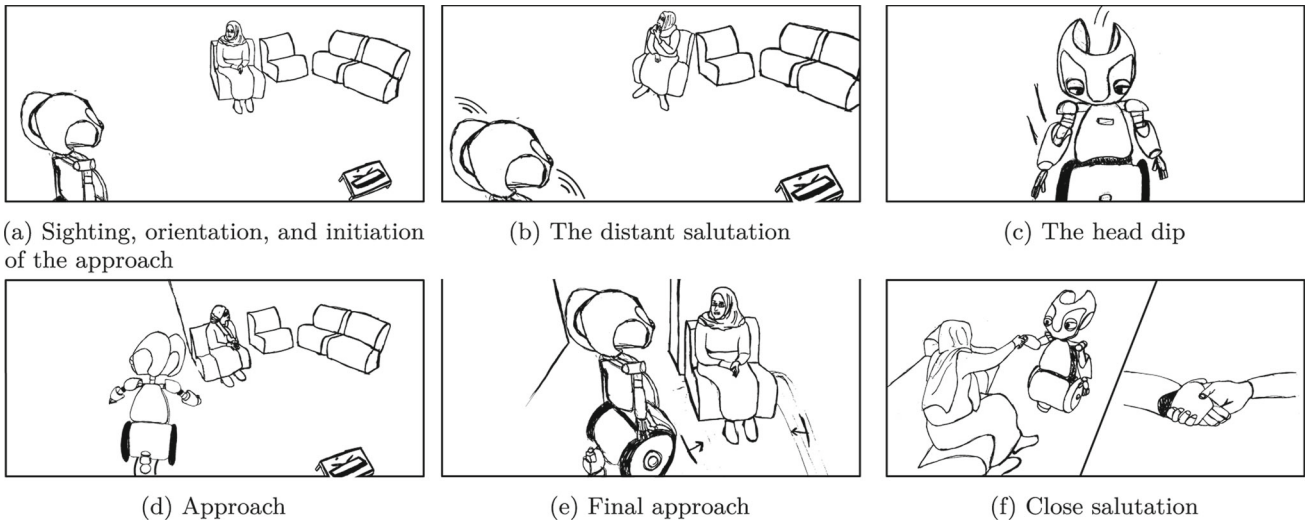

Fig. 2 Storyboard with a possible application of Kendon's greeting model in Human–robot interaction

humans followed a sequence of greeting rituals when meeting someone new, that although with distinct behaviors, follow the same structure across cultures. This process involves the interchange of social cues that ground the participants' interaction intentions and establishes which are the appropriate social norms to use through that interaction or future interactions [66]. Kendon's model is composed of six steps that we analyze in Sect. 2.1. We note that when we refer to "greetings" we are not addressing the individual act of saluting someone, but the full script used to start an interaction. Our definition was adapted from [34,64].

Definition 3 *Greeting* a ritual consisting of a sequence of interaction behaviors observed when people come into another's presence.

Greetings involve an exchange of social cues in the form of non-verbal signals that vary due to culture or the meeting context [9]. During a ZAE, these differences may occur in the management of space, gestures, and salutations. Hall [49] reports notable examples of differences in proxemics and gaze, with comparisons between several cultures. For instance, he argued that the German culture has a stricter notion of space and intrusion than the American culture. Differences can be so extreme between cultures that deviant behaviors in one culture can be considered normal in others. Gaze interactions between the American and English cultures are a notable example observable between two close cultures [49]. While the English keep their gaze fixed on the target to demonstrate that they are paying full attention, Americans find that behavior uncomfortable, preferring to advert their gaze frequently. Even when a social norm has the same positive or negative connotation among several communities, they can follow it with different levels of rigidity (norm tightness [44]).

It is not feasible to enumerate and encode a list of all of them for a robot to follow, due to the number of possible contexts [43]. Moreover, they can also evolve due to external factors. The replacement of handshakes with elbow-bumps during the salutation due to the COVID-19 pandemic exemplifies that.

Thus, creating a positive impact during a ZAE requires much more than following social scripts in an open-loop fashion. Socially aware robots need to perceive social feedback. The literature reports that it can be displayed through both verbal [18] and non-verbal cues [36].

Definition 4 *Social feedback* an evaluative response to a social actor's actions, in a specific social context, displayed through social cues.

Besides allowing a robot to track the interaction state on a social script, the ability to detect social feedback allows the robot to understand whether its behaviors were appreciated or violated people's expectations. We believe this understanding is fundamental to create a positive perception in humans during ZAEs. Since the public has a general perception of robots as competent beings, people can interpret failures and social norm violations as incongruent behaviors, leading to the incongruency effect. However, Jerónimo et al [56] reported that the incongruency effect vanished if the person learned about a personality trait that explained the incongruent behavior. Thus, we believe that a robot capable of understanding social feedback from humans can employ recovery strategies that can enhance the human–robot interaction experience.

For a robot to follow social scripts during a ZAE, it needs to have a set of social skills to perceive and act, thus Social Awareness. To make a comprehensive survey on the technological side of ZAEs, we need to identify relevant skills

and analyze their current implementation strengths and limitations. We make use of Greenspans's definition of social awareness.

Definition 5 *Social Awareness* "... the individual's ability to understand people, social events, and the processes involved in regulating social events."

2.1 Opening Interaction: the Greeting

Focused interaction between people usually starts with a greeting [34,66]. Kendon proposed a model for greetings between humans composed of the six multimodal steps illustrated in Fig. 2. We will now describe Kendon's model as described in his book [64], and discuss the necessary skills to allow a social robot to follow it.

Remark 2 We make a clear distinction between greetings and salutation. We consider the first as the social scripts composed of several interaction steps to initiate interaction. Salutations are the individual gestures or utterances that explicitly signal one's intent to interact (for instance, saying "Hi" and performing a handshake).

Remark 3 We use the term social actor to refer to both humans and social robots.

2.1.1 Sighting, Orientations, and Initiation of the Approach

The first step of the greeting ritual is crucial for its success. First, it requires social actors to recognize others as someone they wish to greet and the conditions to do it. Thus, a robotic social actor needs to be able to detect, track, identify people, and be aware of its surroundings. In this work, we call this set of skills: social context inference. According to Kendon's observations, humans will not approach a target before the target acknowledges their presence. They display this acknowledgment through gaze, which highlights another essential perception skills: gaze and visual field of view estimation. The ways humans get the target to acknowledge their presence depend on several factors: urgency, roles, the goal of the greeting, and their current activity. For instance, Yoshioka et al. [136] claim that the target's activity plays a significant role on engagement behaviors of humans. They found significant differences in speech distances and approach trajectories for distinct perceptions of how much concentrated the target was. It is thus fundamental for a competent social robot to detect human activities, groups, and estimate whether people can be interrupted or not. Kendon reported the following strategies to get the target's attention:

- Orient only head toward the target, but not the body, and wait for gaze signals.
- Synchronize movements with those of target's while averting gaze, to lower the risk of explicit rejection.
- Get the other's attention by calling, making gestures, coughing, or knocking on doors.
- Interrupt the other's activity directly, in urgent cases.

The following necessary skills are needed to employ these strategies: speech, gesture generation, natural gaze control, and body pose control. Humans can halt the greeting in this step without significant social consequences.

2.1.2 Distance Salutation

In this state, both parties officially signal that they initiated the greeting script. From this point, the greeting can either come to an end, if none of the parties intend to have further interaction ("greetings in passing") or continue to other script stages. Thus, it is necessary to track the greeting state to predict how it is going to evolve. The form of salutation can be a relevant predictor, which can be a combination of the following actions:

- Wave
- Smile
- Call
- Head movements:
  - Nod
  - Head toss
  - Head lower

Both parties may perform those salutations, which means that a social robot needs the skills of gesture recognition, facial expression detections, in addition to those we mentioned before.

This stage can be followed either by the head dip, approach, final approach, or close salutation. The distance salutation can occur just before the close salutation if both parties are bound to pass close to one another (for instance, moving toward one another in a corridor).

2.1.3 Head Dip

In this script stage, the social actor bends the neck forward, lowering the head. According to Kendon's observations, it is more likely to occur if humans have to adjust their body orientation to approach the target and does not happen after a distant salutation that does not lead to further interaction.

2.1.4 Approach

The approach is a stage where, either both parties or just one, actively move toward the other. During this step, humans may display:

- Grooming behaviors
- Gaze aversion, which is more salient in the social actor that moves more
- Body cross, which is a gesture where the social actor that walks a greater distance brings one or both arms forward briefly.

From these descriptions, we can identify an extra skill for social robots: socially aware navigation.

2.1.5 Final Approach

The final approach occurs when both parties are closer than 3.5 m and just before the close salutation. During this stage, we can observe the following behaviors:

- Verbal salutation
- Mutual smiling
- Mutual gazing
- Gestures where the participants show their hand palm

As the robot will be getting closer to the target in this phase, it should be able to execute a socially acceptable trajectory, and how to enter a group of people.

2.1.6 Close Salutation

The close salutation is the final stage of the greeting script. Here, the participants come to a halt, orient their hands toward each other, and salute each other verbally and non-verbally. Non-verbal salutations may involve body contact and are culturally dependent. Notable examples include:

- Handshakes
- Fist bumps
- Kiss on cheeks
- Hugs
- Bows
- Head nodding

Finally, both parties adjust their relative positions. According to Hall's proxemic theory [49], these distances signal the person's psychological proximity. At this stage, the greeting script ends. From this description, we can identify the following skills: salutation detection and performance.

Opening an encounter with a greeting is transversal between cultures, but the sequence length of Kendon's model varies according to several factors. Besides the cultural differences in the close salutation (for instance, handshakes, hugs, or kisses), the execution of each part of the model depends on how acquainted the parties are (being shorter, the emotionally closer they are) and context. Schiffrin [110] observed that the process is not always linear since failures in human perception can lead them to repeat some behaviors or even cancel the greeting with an apology. Social actors can fail and violate social norms during an interaction, which can elicit reactions from people [12]. Thus, the robot should be able to detect them and recover from interaction failures, since research as shown that it will improve people's perceptions of the robot [30]. We identify this skill as social feedback detection. Thus, these observations show us that the first encounter between people is a complex set of communication and perceptual skills.

2.2 Categorizing Social Skills with Greenspan's Model

Analysis of Kendon's model shows that a robot requires a multidisciplinary set of socially aware skills to engage with someone. The robot needs to infer the context and appropriate social norms, detect social cues and people's feedback, and communicate through verbal and non-verbal behaviors. To perform a structured and useful survey, we need a proper categorization of research works related to these skills. We find inspiration in Greenspans's theoretical/conceptual model of Social Competence to set a taxonomy for human–robot zero-acquaintance encounters. Greenspan [47] categorized these abilities under the Social Awareness competence group. Social Awareness is composed of three categories of skills: (i) Social sensitivity, (ii) Social insight, and (iii) Communication. This model was proposed during studies related to children with mental disabilities. Even though several theoretical models for Social Competence exist in the literature [25,31,35,45], we believe Greenspan's model serves a simple but efficient tool to categorize robots' social skills for zero-acquaintance encounters.

2.2.1 Model Description

The social sensitivity component of Greenspan's model deals with the capabilities to perceive and understand social agents, objects, and events. It has two sub-components: social inference and role-taking. The social inference ability consists of correctly classifying social situations, gatherings, and context. Role-taking is the ability to understand the viewpoints and feelings of others.

Social insight is the ability to interpret and understand the processes that govern social events and evaluate them. It splits into three sub-components. The first one is social comprehension, which is the ability to understand social models

and processes, like relationships, social classes, norms, and reciprocity. The second sub-component is psychological insight, which consists of the capability to understand people's motivations and personalities. Moral judgment is the third sub-component and consists of skills related to ethics, morality, and intentionality.

Social communication is a set of skills to deliver information to other social actors and influence their behaviors. It is composed of referential communication and the social problem-solving sub-components. Referential communication is the set of verbal and non-verbal skills necessary to communicate one's thoughts and feelings. Social problem solving is the ability to influence others toward one's goals and to resolve conflicts.

2.2.2 Assigning Necessary Skills for First Encounters to Greenspan's Model

We now categorize the required skills to open and close the interaction, under Greenspan's model. Each one of them will belong to one of the model's three categories, and then we will either use the sub-dimensions as sub-categories or create new ones. We do this to keep the structure simple and avoid unnecessary nested sub-categories.

We propose to group the social context inference, gaze & VFOA estimation, group detection, interruptibility estimation, and role-taking skills under the social sensitivity category. All of these abilities capture the social context. We note that social context inference is composed of a set of atomic skills that we will not discuss individually: detect/track/identify people, objects, activities, and facial expressions. Here, we are interested in how researchers integrated these skills to detect and represent the social context. Role-taking will designate the robot's ability to understand people's feedback and reactions toward it.

Under the social insight category, we address the social comprehension skills of socially aware navigation and understanding of social norms. We propose to associate them with social comprehension split into implicitly and explicitly defined social comprehension. The first deals with models that encode social norms implicitly, like costmaps in socially aware navigation. The second addresses methods and models where social norms are explicitly defined.

Our proposal for the communication category is to use its sub-categories of referential communication and social problem-solving. The first sub-category deals with the gestures used for non-verbal communication, salutations, gaze gestures, and their dynamics. Social problem-solving addresses robot behavior adaptation to social feedback.

2.3 Survey Structure

This survey is structured as follows. In Sect. 3, we present the methodology to survey research works related to our topic. Since we wrote this survey with a top-down approach in mind, we will start by addressing existing papers which focus on robots that engage people on possible first encounters. Afterward, we will review the needed skills, categorizing them with Greenspan's model. Thus, Sect. 4 analyses research works with robots engaging people, compares their social scripts with Kendon's greeting model, and summarizes their engagement success. The following three sections describe works categorized under each of Greenspan's components of social awareness. Section 5 describes works under the social sensitivity component. Those describe methods that perceive the social context and signals. Section 6 focuses on the social insight component, presenting papers that developed methods that model social interaction and norms. Then, Sect. 7 focuses on the communication component and presents works that developed nonverbal communication skills and strategies. We finish this survey with conclusions and research directions in Sect. 8.

3 Survey Method

Our survey followed a methodology inspired by the insights of Webster and Watson [131] and recommendations of vom Broke and colleagues [19,20]. After defining this survey's scope, we iterated through loops of conceptualization, literature search, and literature analysis (Fig. 4). We selected a total of 64 papers to debut in this survey as a result of the iterative process (refer to Tables 2 and 3). It was unfeasible for us to keep track of the number of discarded papers, as well as used keywords, mainly due to the iterative method and forward / backward search. Nonetheless, we created a word cloud to represent the frequency of the fifty most common words in titles, author keywords, and INSPEC keywords of the surveyed papers, to guide researchers when they perform a further investigation in this subject (Fig. 3). In the following subsections, we describe our method in detail.

3.1 Problem Identification

We identified the topic covered in this review through reading and discussion on human–robot interaction textbooks and journal papers. Most notably, Kanda and Ishiguro's book on human–robot interaction [60], Rios-Martinez et al.'s survey on proxemics in robotics [101], Shi et al.'s work on a flyer distributing robot [115], and Charalampous and colleagues' review on recent trends in socially aware navigation [26]. Thus, we reiterate the question on Sect. 1.2: "How far are social robots from being able to engage with strangers

Table 1 Taxonomies for robots engaging with people

Reference	Stage of Kendon's model						
	(1) Sighting, orientation, and initiation of the approach	(2) The distance salutation	(3) The head dip	(4) Approach	(5) Final approach	(6) Close salutation	Engagement success
Satake et al. [106] and Satake et al. [107]	**(1) Fiding an interaction target:** Select reachable & anticipate willingness to interact. No gesture	–	–	**(2) Interaction at a public distance:** Frontal approach	**(3) Initiating a conversation at a social distance:** Nonverbal intention to interact. Recognize acknowledgement	Greet people verbally	Engaged people: 56%
Shi et al. [115]	Compute approach utility to select target. Gaze at target.	–	–	Frontal approach target. Continuous gaze.	Reduce velocity with distance. Extend arm. Gaze. Verbally offer flyer.	–	Distributed flyers: Robot: 18% v.s. Human: 10%
Zhao et al. [140] (WoZ. Robot reacts to human approaching)	**Far field:** Raised eyes (facial expression)	–	–	N.A. (Human approaches robot)	**Mid field:** Smiling eyes. Voice greeting.	**Near field:** Smiling eyes & blush. Voice intro.	N.A.
Heenan et al. [54]	**Sighting:** Idle behaviors. Detect person. Attempt eye contact.	**Distance salutation:** Stand. Gaze at person. Wave.	–	**Approach:** Avoid eye contact. Move to personal space and then gaze at person.		**Close salutation:** Handshake & gaze & vocal greeting	N.A. (informal observations)
Foster et al. [41]	Select user paying attention. Gaze.	–	–	N.A. (Human approaches robot)	–	Gaze & verbal greeting	N.A.

Table 1 continued

Reference	Stage of Kendon's model (1) Sighting, orientation, and initiation of the approach	(2) The distance salutation	(3) The head dip	(4) Approach	(5) Final approach	(6) Close salutation	Engagement success
Brščić et al. [22]	**Wait and observe:** Gaze around. Select target. Gaze at person.	–	–	**Approach:** gaze and move toward person.		**Guidance service:** Verbal greeting. Offer guidance.	Engaged people: 87.5%
Kato et al. [62]	**Proactively waiting:** body and gaze oriented at target.	–	–	**Collaboratively-initiating:** move toward person and offer help just before stopping.		–	Engaged people: 87.2%
Saad et al. [105] (High enthusiasm mode)	**Select target:** select a target that is not engaged	**2) Draw attention (part 1):** Wave & verbal greeting.	–	**2) Draw attention (part 2):** Small approach movement (0.3 m)	–	–	Human attentiveness score (details on paper): Wave: 0.84 Wave & speech: 0.77 Wave & speech & approach: 0.95

Table 2 Papers covered in this survey (part 1)

Ref.	Robots engage people	Social sensitivity					Social insight		Communication	
		Social context inference	Group detection	Gaze & VFOA	Interrupt.	Role taking	Implicitly defined social comprehension	Explicitly defined social comprehension	Referential communication	Social problem solving
[125]		✓					✓			
[126]			✓				✓			
[135]							✓			
[99]							✓			
[106]	✓				✓					
[115]	✓						✓		✓	
[140]	✓									
[54]	✓									
[4]			✓							
[13]				✓						
[15]					✓					
[16]			✓							
[27]					✓					
[32]			✓							
[69]		✓								
[71]			✓							
[76]				✓						
[77]				✓						
[81]						✓				
[84]					✓	✓				
[94]						✓				✓
[95]						✓				✓
[96]			✓				✓			
[102]						✓				
[112]			✓							
[113]			✓							
[123]		✓								
[127]					✓					
[128]			✓							
[130]						✓				
[133]				✓						
[139]			✓							
[24]								✓		

in feedback sensitive and socially acceptable way in first encounters?"

3.2 Conceptualization of Topic

As a consequence of not finding an overview of the topic, we organized our survey guided by Kendon's model of human greetings [64] and Greenspan's model of social competence [47]. Even though the main topic remained unchanged, the scope evolved along the iterative process in order to become more specific and comprehensive.

3.3 Literature Search

We restricted our literature search to the following acadamic search engines and databases: IEEE Xplore, Scopus, Google Scholar, and Scinapse. The sets of keywords used to query the databases evolved with the scope redefinitions and with infor-

Table 3 Papers covered in this survey (part 2)

Ref.	Robots engage people	Social sensitivity					Social insight		Communication	
		Social context inference	Group detection	Gaze & VFOA	Interrupt.	Role taking	Implicitly defined social comprehension	Explicitly defined social comprehension	Referential communication	Social problem solving
[89]										✓
[90]								✓		
[48]							✓			
[79]							✓			
[6]									✓	
[7]									✓	
[57]									✓	
[58]									✓	
[59]									✓	
[85]									✓	
[86]									✓	
[83]									✓	
[11]									✓	
[117]									✓	
[33]							✓			
[41]	✓	✓								
[3]						✓				
[107]	✓				✓					
[22]	✓				✓					
[62]	✓				✓					
[105]	✓									
[67]						✓				
[73]						✓				
[72]						✓				
[21]						✓				
[46]						✓				
[138]		✓								
[63]				✓						
[124]						✓				
[68]						✓				
[38]						✓				

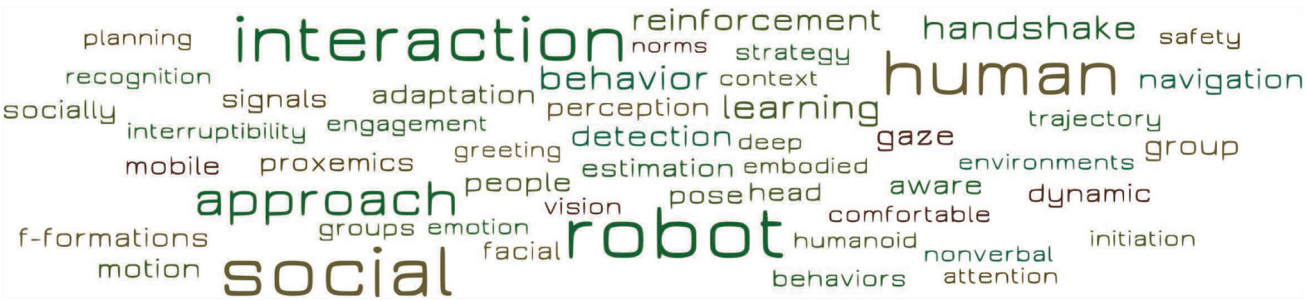

Fig. 3 The fifty most common words in the surveyed paper titles, author keywords, and INSPEC keywords. Word sizes represents their frequency

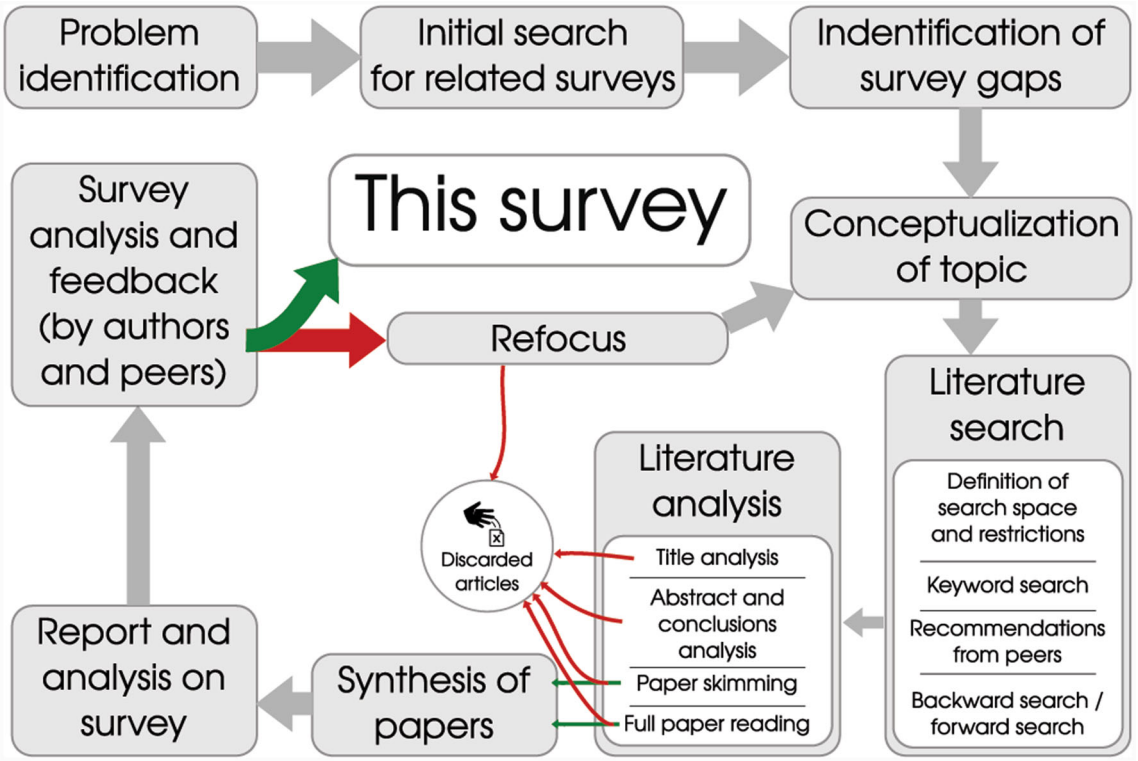

Fig. 4 The iterative survey method and its inner cycle. First we identified the research topic from books and discussions with colleagues. From those, we identified the challenge of socially aware human–robot engagement during first encounters. Search for surveys of this topic revealed a gap. Then we employed an iterative cycle of (re)conceptualization, literature search, literature analysis, paper synthesis, writting, and survey analysis

mation from the previous paper analysis. In addition to the active database searches, literature suggested by colleagues, peers, and reviewers was an extremely valuable asset in the process, since these were curated resources that introduced new keywords and search terms. Finally, the search process also had steps of backward and forward search. The backward search step consisted of collecting references cited by collected papers. The forward search step consisted of collecting papers that cited the already collected papers.

3.4 Literature Analysis

Since it is unfeasible to analyze all papers to a full extent, we used a method inspired in Subramanyam's work [121]. First, we analyze each paper's title and discard those where the title is clearly out of the scope of the survey cycle, i.e., those with title keyowrd that do not respect scope restrictions. Then, we analyze the abstract and conclusions of the remaining articles to clarify whether their topic fits. Afterward, we skim the selected papers. During the skimming process, we examined tables, figures, and scanned through the introduction and discussion. For some articles, it becomes possible to either make an informative summary or discard them with this data. Finally, we fully read and examine the remaining papers, either summarizing them or discarding them.

Regarding works on robots engaging with people, we only included those where the robot opens interaction with people without a personalized model. These can either be technological or HRI studies, as long as they describe the interaction stages in detail and present the robot's architecture. We excluded papers that focus on posterior moments of interaction and those that did not feature single minimally anthropomorphic robots.

As for the individual robotic skills, we only include those that implement the skills derived from Sect. 2 and categorized in Sect. 2.2.2. These can be works, that although not tested on autonomous robots, can be applied to them, as is the case for computer vision algorithms. Since we do not deal with the challenges of conversation management, we excluded papers that address speech synthesis, recognition, natural language processing, and dialogue management. However, we do not exclude works that use verbal and prosodic features since these can be relevant cues to detect feedback.

3.5 Final Cycle Steps

In the final cycle steps, we compiled the summarized papers into the survey, from which we identify literature gaps, draw conclusions, and reason about future directions. It was followed by a review and discussion process either within the authors or between authors and peers. This process is fundamental for the survey to converge into a helpful and comprehensive tool for future research.

4 Robots Engaging with People

The research topic of robots that engage with people is receiving a keen interest in the research community. Even though a considerable amount of works in the literature address the problem of a robot that engages with people, a significant amount of them focus solely on robot trajectories during the robot's approach [99,125,126,135]. However, as observed in Kendon's model, initiating an interaction with someone requires an interchange of social signals. Moreover, since people might not be expecting to be engaged by a robot, during a first encounter, being unable to reproduce and detect these social signals may lead to failed engagement attempts. Satake and colleagues [106,107] observed and categorized failed engagement attempts with Robovie at a shopping mall. These consisted of the following types:

1. *Unreachable* when the robot cannot get close to the person. It can happen due to actuator limits, or because the person was leaving.
2. *Unaware* when the person did not notice the robot's behaviors or did not recognize them as an attempt to interact.
3. *Unsure* when people notice the robot's actions but are not certain of the robot's intention to interact with them.
4. *Rejective* when people understand the robot's intentions but do not intend to interact.

Thus, Satake and colleagues [106,107] suggest that engaging robots should not approach people naively. As such, we now analyze past strategies for mobile social robots to initiate interaction. Since past works present distinct taxonomies to describe the social scripts that they follow, we use Kendon's model to compare these works under a single taxonomy. Moreover, since Kendon developed this greeting model from observations of humans, it also allows us to compare these works' social scripts with those observed in humans. We compared their respective taxonomies with Kendon's model is Table 1.

Distances between social actors play a relevant role not only on their psychological distance [49] but also on the displayed behaviors when initiating the interaction. All papers in Table 1 use them, whether the robot approaches people, or whether they approach it. For instance, Zhao and colleagues [140] tested the concept of "progressive interaction" with a three-stage model. Each stage relies on the person's distance to the robot to control its expressions and utterances: (i) the far field (from 4.2 m to 2.7 m); (ii) the mid field (from 2.7 m to 1.2 m); and (iii) the near field (less than 1.2 m). These stages compose their "progressive interaction" condition. In the far field, the robot displays facial expressions toward the person. Then, the robot verbally greets the person and uses more facial expressions in the mid field. Finally, once in the near field, the robot asks the person to talk with it. They report that people preferred the "progressive interaction condition" instead of passive behavior, where the robot waits for interaction. Distance may also mean that the robot cannot reach the target, and thus should cancel an engagement attempt that would fail due to unreachable targets, before it even begins. Computing the target's reachability is one of the first steps of Satake et al.'s [107] and Shi and colleagues' [115] works.

After knowing that the target can be reached, getting the target's attention and expressing the robot's intent to interact are two essential abilities. Researchers have done this in many ways. Showing high enthusiasm gestures can be an effective strategy to draw people's attention, as studied by Saad et al. [105]. They performed a study with Pepper at a building's entrance with mild (wave), moderate (wave & speech), and high (wave & speech & small approach movement) enthusiasm. They reported that people paid more attention to the robot when it showed high enthusiasm. Nonetheless, attempting to establish eye contact is the most common strategy among the analyzed papers [22,41,54,62,115], going in line with Kendon's description of the first stage of his model. Not only do robots attempt to get the user's attention through gaze, but it is also a cue of human intention to interact with them. For instance, the human gaze at the robot is used as an interaction opening signal by Pepper in the MuMMER project [41]. In that project, Pepper's role was to give direction to people at a shopping mall. It initiated interaction after detecting nearby people gazing at the robot and gazed back at them. Getting the target's attention addresses the "unaware" error type.

A socially aware approach has been seen in the literature either after both parties acknowledge each other's presence [22,54,62,115] or as a way to get the target's attention [106,107]. Satake and colleagues [107] carefully designed Robovie's approach behavior to show the robot's intent to interact when advertising shops to shopping mall's passerby. Their planner anticipates people's trajectories and computes a trajectory for a frontal approach toward a meeting point. With this behavior, they intended to reduce both "unaware" and "unsure" error types. They considerably reduced the number of "unaware" errors from 14% to 4% and of "unsure" errors from 24% to 18%, when compared with a strategy that

only navigates to people's positions. In total, they managed to engage with 56% of the approached people. Besides a frontal approach, gestures and appropriate velocities are also relevant. Shi et al. [115] gave Robovie the challenging task of flyer distribution. They first studied how humans do it and modeled their strategies. After computing a target selection plan that maximizes the number of reachable targets, Robovie gazed at its next target, moved toward her/him with continuous gaze, and extended it arm with the flyer while decelerating and verbally offering it. This last part is similar to Kendon's description of the final approach. The robot managed to distribute flyers to 18% of the engaged people, while a human could only distribute to 10%.

Being able to detect if people are open for interaction can reduce the occurrence of "rejective" errors, as claimed by Brščić et al. [22], and Kato and colleagues [62]. Brscic et al. implemented a classifier that detected people with atypical trajectories and selected them as approach targets. They reasoned that those people might be lost and thus be open for the robot's help. The robot followed the steps in Table 1 during the approach. It managed to successfully engage in 87.2% of the attempts at a shopping mall. Similarly, Kato et al. estimated a store's customer's need for help from their trajectories. Robovie directed its body and gaze at likely targets and only initiated its approach movements when the person moved in its direction. It was successful in 87.2% of the attempts and significantly better than a passive approach (62.9%) and a proactive approach (42.7%).

Integrating all these behaviors and strategies is a challenging task. It requires accurate tracking and management. We argue that knowledge of social scripts will allow a robot to manage and track the interaction during first encounters. We believe that prior information about behaviors during the interaction will allow the robot to estimate its state given those that it observes, and to generate appropriate behaviors at each interaction step. Heenan and colleagues [54] implemented a state-machine model that integrates Kendon's greeting model and proxemics theory in the NAO robot. They argue that due to the lack of robust sensing capabilities, they needed to approximate the model to rely solely upon (i) presence; (ii) orientation; and (iii) location. Through informal observations, they report that even though the model is a good starting point for engaging people, it needs further development. They highlight that: (i) constant gaze can be awkward; (ii) robot pacing is important; and that (iii) the system needs to be more reliable to error situations, among others.They highlight that the system needs to be more reliable for error situations. Nonetheless, up to our knowledge, they were the first to explicitly follow Kendon's model to track and manage the interaction.

4.1 Research Gaps

The current state-of-the-art presents researchers with numerous opportunities to develop complex engagement behaviors for first encounters. Up to our knowledge, a small number of works attempt to implement models based on all steps of Kendon's model, or similar approaches. As noted, sensing capabilities are indeed a bottleneck for complex autonomous interactions.

Managing and tracking the meeting is also an open challenge. Even though some works take into consideration cases where the person does not intend to interact, the greeting steps depend on the context, with distinct steps for different circumstances. Moreover, the interaction might not be sequential. Humans might return to a previous level of the model, or skip a step depending on the social cues and mistakes that they make during the interaction.

5 Social Sensitivity

In this section, we survey existing works that perceive and understand humans, objects, and events. These skills compose the Social Sensitivity component of Greenspan's model. A social agent can use these perceptions to choose the best way to act (Communication component, Sect. 7) according to its models of social events (Social insight component, Sect. 6). We address architectures which detect low-level social information (like people, objects, their poses, and people's facial expressions) in Sect. 5.1—Social context inference. Then, we present works that estimate gaze direction and the visual field of attention (Sect. 5.2) followed by group detection (Sect. 5.3). Section 5.4 methods in the literature that deal with the challenging problem of interruptibility estimation, a significant cue for an agent that intends to interact. Finally, in Sect. 5.5 we address "role-taking", the ability to understand others' feelings and viewpoints. Humans can share this information through feedback. Thus, we focus on literature that proposes methods to estimate it. We end this section with an analysis of research gaps of social sensitivity.

5.1 Social Context Inference

The objective of [138] is to detect and track a large set of social signals to be used by a robotic head automata during dialogue HRI. They propose a system that tracks and stores a social scene. Their system uses RGB-D, RGB, illuminance, sound level, and temperature sensors. After low-level feature extraction, they perform (i) facial analysis, (ii) identity assignment, (iii) body analysis, and (iv) saliency detection. During facial analysis, they extract face positions and eye, nose, and mouth landmarks. They use this information to classify people's gender and estimate their age and facial

expressions. The system uses QR-codes to identify people and Kinect's skeleton tracking library to recognize a set of states/gestures (seating, standing, raising hands, crossing arms). Additionally, they also detect the saliency of image regions (interesting regions that attract human gaze). Afterward, they compile all this information into a single file, that describes the scene. This meta-scene file can then be used for HRI algorithms. This work is followed up by [69], where it becomes part of a cognitive architecture for robot face and head control.

The SPENCER project proposes an architecture for a mobile robot that guides people in an airport [123]. The robot can map and localize itself in very dynamic environments, and detects and tracks people and groups with laser and RGB-D data. It additionally detects objects and the spokesperson in order to guide a group of people to their destination, formulating the problem as a Mixed Observability Markov Decision Process.

On [125], the authors aim at creating a social navigation framework based on proxemics theory. The system's social awareness architecture detects and tracks humans with an RGB-D camera. The system estimates people's states (standing/sitting/moving), walking velocities, the field of view, interactions with "interesting objects" (with markers), and social interactions. The robot uses these data to create a cost map to navigate and approach people.

The MuMMER project [41] developed a complex system to infer the social context around the robot through audio-visual sensing. In the visual part, they extracted people's poses 2d skeleton poses using convolutional pose machines (OpenPose) [132] and OpenHeadPose [23] to estimate head poses. They kept track of people with face poses, colors, and OpenFace re-id features [8]. Additionally, they used a microphone array to perform voice localization. A multi-task neural network jointly performs speech/non-speech detection and sound localization, as proposed by He and colleagues [53]. Finally, they fuse both visual and audio location estimates assigning speech direction with the visually detected people to detect who is speaking. Their system also computes the visual focus of attention of each person based on estimated head poses with Sheiki and Odovez's work [114].

5.2 Gaze and Visual Field of Attention Detection

The human gaze is an important cue to detect human–human/object/robot interaction. Even though humans have a strong ability to estimate gaze accurately, it is still a difficult task for robots. Thus, it is receiving interest from the research community. Openface [13] is an example of an opensource framework for facial analysis that can estimate gaze. They use a method presented in [133], called eye-CLNF (Constrained Local Neural Field), trained on a synthetic training dataset of photo-realistic render of human eyes. Their approach achieves accurate results if the image of the subject's eyes has enough resolution. However, it fails with people who wear glasses or if their eyelids occlude the eye.

Recently, researchers created a rich dataset of people looking at a moving target (with known position) [63]. Then, they train/test using head crops and feed them to a backbone network (ImageNet pre-trained ResNet-18) that outputs 256 features to a bidirectional LSTM's with two layers and a fully connected layer. Their algorithm predicts gaze in spherical coordinates relative to the camera frame and the uncertainty of the gaze estimation. It has plausible results even when the eyes are not visible.

The visual field of attention (VFOA) is probably an even more important cue than gaze direction to reason about someone's ongoing activity and interactions. To estimate it, the authors of [76] propose a probabilistic formulation of the problem. They define target locations (objects or heads) and head orientations as observed random variables and VFOAs and gaze directions as latent random variables. They use a switching Kalman Filter approach and test it on two proposed datasets. More recently, they extended their work [77] to predict the VFOA when objects are outside of the image. Given people's heads' position and orientation, they create a top-down gaze heatmap that they feed into an encoder-decoder convolutional neural network. The output is an object heatmap that represents VFOA 3D locations from a top-down view.

5.3 Group Detection

A social robot should detect groups of people. The literature classifies groups of people into two distinct classes: semi-static groups of standing people and dynamic groups of people. It describes several techniques to detect semi-static groups of jointly focused interaction.

Perhaps the most commonly studied problem is the detection of standing conversational groups. For instance, one approach [16] uses people's 3D head orientation and proximity information to detect whether their view frustum intersects, thus assuming they are in a group. Hough Voting is a common strategy [32,112]. The idea is to associate a Gaussian probability density function that represents the probability of the o-space center, to each person in the scene. This set of distributions is used to vote for a given o-space center location.

Other works use game theory. The authors of [128] use people's position, orientation, and associated uncertainty to compute the most plausible region of attention. Then, they compute a pairwise affinity matrix for each person and extract the F-Formation as solutions of a non-cooperative clustering game over multiple frames.

Graph-based methods currently have the best results in a recent evaluation with the GRODE metrics [111]. [113] developed a Graph-Cuts based method that uses proxemic information (position and orientation) to detect F-Formations on single images. Another graph-based method [139] aims at detecting levels of involvement in free-standing conversing groups for single images.

Most works that use RGB data use fixed ceiling cameras to maximize people's detection efficiency. However, some notable exceptions, like [4], detect groups of people from head-mounted RGB cameras. To avoid degrading the results, they first detect blur in the image and discard it if larger than a threshold. Then, they detect faces, compute each face's 3D pose. Finally, a correlation clustering method estimates groups taking temporal information, position, and orientation into account.

Dynamic group detection was also explored in the literature, but to a lesser extent. On [71], a system uses RGB-D data to detect and track people and dynamic groups. Their approach uses HOG's and HOD's to detect people and tracks them with a Multiple Hypothesis Tracker (MHT). A probabilistic SVM predicts social relations between detections and an extended version of MHT tracks groups. The full system is computationally heavy but able to run in real-time.

The authors of [96] propose two fast methods. The Link Method uses a static analysis based on proxemics and dynamic analysis to track pairwise relationships' evolution. The Interpersonal Synchrony Method runs over sliding-time windows and detects pair interactions through the intersection of the field of views. Then, it evaluates intergroup synchrony through the analysis of people's speeds.

In [126], the authors extend the Graph-Cuts method proposed by [113] to deal with dynamic groups. They do so, by adding velocity information to people's state and adding motion constraints to the algorithm.

5.4 Interruptibility Estimation

As reported in Sect. 4, knowing whether people are open for interaction can significantly improve engagement success. Thus, it makes it necessary for a robot to estimate interruptibility automatically.

People's poses and trajectories are significant cues to decide whether to engage with them or not. Thus, Satake and colleagues [106,107] developed an algorithm that classifies people's trajectories into four classes: (i) fast-walking, (ii) idle-walking, (iii) wandering, and (iv) stopping. With this information, their system predicts if the robot can approach a pedestrian, and chooses a pose to intercept them. Kato et al. [62] also use trajectories to understand when Robovie should engage with shop clients, based on their need for help. They trained an SVM to learn interaction intention, with 95.4% performance, from the following features:

- Distance to robot.
- Smallest robot frontal aperture angle that can cover the human trajectory.
- Deviation of velocity.
- Stop time.

To approach humans with atypical behaviors, Brščić and colleagues [22] trained an SVM classifier to detect those, based on two features: speed and predictability. The predictability feature represents how likely people are of going to a position, given a pedestrian motion model. Their detector of atypical behaviors achieved 91.4% accuracy.

Banerjee et al. propose [15] a system that estimates if people are interruptible. Their architecture extracts spatial information (position, orientation, head orientation, and gaze direction of a person), and sound (presence and orientation). Using video data, the researchers label objects near the target person. This data is fed into several machine learning algorithms to estimate the level of "interruptibility" (from 0 to 4).

Other works do not represent the social scene explicitly, using an end-to-end approach. On [84], the authors attempt to detect whether a person can be interrupted or not and the scene context (studying, dining, at lobby). They test two different sets of features: audio amplitude with image intensity, or GIST with volume and frequency features. With them, they train several classifiers: SVM, Naive Bayes, and Decision Trees (maximum of 78.07% accuracy for context and 70.64% for appropriateness). The authors of [27] trained a neural network that, given a detected person, creates a heatmap around the focus of interaction and a caption that describes the activity.

5.5 Role-taking

We believe that the capacity to recognize humans' feedback to actions is fundamental to a social robot during human–robot interaction. There is still scarce literature on robots that receive natural feedback from humans and learn from them. However, distinct feedback modalities have been explored in past works.

From an implementation point of view, one of the easiest ways for a robot to collect social feedback from humans is through button presses or interface clicks from an informed person. That is the case in the original paper presenting the TAMER framework [67], a reinforcement learning framework that takes users' feedback to shape their behaviors. MacGlashan and colleagues [73] trained a virtual dog to navigate a grid world environment through 5 buttons of feedback to test their proposed reinforcement learning algorithm. Another work [72] uses binary button feedback to make a virtual agent learn how to chase and catch a second one. They claim that the lack of feedback can be as informative as

explicit feedback and present a probabilistic model of how a trainer gives it. The work of Nigam and Riek [84], is yet another example where a robot receives button feedback. The robot uses this feedback to learn whether it interrupted people or not.

Facial expressions contain significantly more information than the previous modalities and do not require the user to touch the system. Broekens [21] estimated affect from facial expressions associating happiness to positive rewards and fear to punishing reward. These signals were collected from people watching an agent in a grid world environment. Social feedback improved the performance of the agent when compared with a condition without it. Gordon and colleagues [46] composed social feedback as a weighted sum of detected valence (three values) and engagement (binary). They used a commercial product to compute these variables from smiles, eyebrows, and lip motions and used the social feedback signal to train a robotic tutor to motivate children.

Other works estimate social feedback from body movements and poses. Mitsunaga et al. [81] present a work where they adapted the robot's behaviors (proxemics, gaze meeting ration, motion speed, and waiting time) with natural signals with a Policy Gradient Reinforcement Learning (PGRL) method in real-time. The robot uses the human's movement, time spent looking at the robot, and time spent before interaction. Trung and colleagues [124] used the 3d coordinates of the head shoulders and neck from data gathered in their previous work [80], to produce distinct feature sets used to train several classifiers. Their goal was to detect robot failures from people's reactions. These reactions can be seen as expressions of negative feedback since they are responses to the unintended robot states. Their best results were achieved using a KNN classifier trained with feature vectors composed of the average of differences between features over a 1 second time window. The authors claim that the classifier could be used in real-life scenarios if the detected person is part of the training set. However, it does not generalizes. More recently, Kontogiorgos et al. [68] used head movements, gaze, and speech features to detect reactions to robot generated speech failures during a task where a robot (either human-like or a device) instructed users to cook non-trivial recipes. The authors used a random forest classifier to classify segments of videos. The classifier was better at detecting "no failures" than "failures". Gaze features and head movements were found to be important when people dealt with a humanoid robot. Ritschel and colleagues [102] use a multimodal approach to get people's engagement. They intend their robot to adapt its personality (with different language behaviors) to keep the user engaged during the interaction. The robot has different levels of introversion and extroversion and estimates the user's engagement with a Dynamic Bayesian Network (DBN). They gather body data from a Kinect sensor and detect head tilt, head orientation, head touches, crossed arms, open arms, and lean postures.

Audio is yet another important modality, used, for instance, to detect laugher, a significant social signal. Although it is a complex signal related to both positive and negative feedback [38], it is a strong signal that, under normal conditions, implies that something happened. Weber et al. [130], developed a laugher detector for their reinforcement learning joke-telling algorithm. They analyzed an audio signal with a sliding window approach and classified voiced frames with a Support Vector Machine that used paralinguistic features. This system achieves 84% accuracy on laugher recognition on a person-independent evaluation. They also used video data to detect smiles through commercial software. They claim that both detectors' confidence can be an efficient estimator of laugher intensity.

Researchers have also combined several modalities to compute feedback. The Ph.D. thesis of Ahmad [3] containts such an example. It describes a behavior selection unit for a social robot engaging in a game with a child that uses a reinforcement learning based algorithm to set the robot's personality. The reward signal can be thought of as a form of social feedback: social engagement. It is computed using eye-gaze toward the robot, facial expressions, verbal responses, and simple gestures. Qureshi et al. [95] used detected smiles, successful handshakes (hand sensors), and eye contact detection to learn the most appropriate action given the state.

Finally, we also note that robots can potentially sense signals that are invisible to humans and use them as social feedback. The work of [127] and colleagues is such an example, were a robot uses EEG signals to detect user engagement and adapt the its speech behavior to keep a user interested in the game. This signal is used in an Inverse Reinforcement Learning approach as a complement to the user's score.

5.6 Research Gaps

Most works present a fixed pipeline of modules that infer specific signals for specific applications. Even though notable examples like [69,123,138] developed an architecture that gathers a significant amount of sensed signals, it seems that a central question remains open: which features are necessary for general social sensitivity, and how can we feasibly detect them all? The lack of exploration of fundamental skills for social sensitivity supports this observation. Robots in the literature are still incapable of detecting ongoing norms or identifying that some correlations between contexts and human behaviors represent a norm. Moreover, robots are still incapable of detecting cues that let them predict that their actions might cause discomfort to people, for instance, by blocking the affordance space of an object.

Regarding individual social sensitivity skills, they still suffer from high computational requirements and accuracy

issues. Most works on group detection focus on standing conversational groups using 3rd person views, which implies that the biggest limitation of these methods is the assumption of perfect person detection. Works that consider uncertainty are computationally intensive, and all of these works are limited to using spatial information and velocities. Relying on a better synchronization of relevant features, like map semantic information, objects, gestures, and sound can potentially disambiguate difficult scenarios, or detect groups without the detection of all participants.

Of the analyzed works, the best algorithms for gaze detection are exceedingly computationally expensive for a mobile social robot. Others are unreliable at greater distances. None of the algorithms make explicit use of the scene context to improve estimation results. Efficient gaze detection from a moving robot still seems difficult to achieve, given image motion noise and occlusions. A possible route to lessen computational costs would be to explore prior information. For instance, object affordances and human pose information may provide valuable information to a robot estimating the human visual field of attention.

As for end-to-end methods like [27,84], they are application-specific. Even though they might learn to extract important social features from images and sound, these features lack interpretability. Moreover, these methods are computationally expensive and require significant amounts of training data.

Concerning the role-taking dimension of social sensitivity, it is still an underexplored topic. Existing works have identified that detecting people's reactions to technical failures of robots is easier than social norm violations, which remain a challenge. People's attitudes to norm violations can be ambiguous, since people may express laugher as a response to both error situations as well as norm compliant robot behavior. This data needs to be ecologically plausible for a robot to be able to receive feedback in the wild. Moreover, there is no relationship between human reactions and measurable quantities (either self-reported scales or physiological data) [118]. Finally, there seems to be a gap in receiving feedback related to physical discomfort related to an interaction. For instance, a socially sensible robot should be able to perceive whether a handshake is too tight or too loose from the person's reactions.

6 Social Insight

With social context data, the robot can reason about the scene and act accordingly. These understanding and decision skills correspond to Greenspan's social insight component. This component is composed of knowledge of social norms, scripts, and models. Here, we will address works that implicitly encode this information (Sect. 6.1) and those that explicitly do it through social norms (Sect. 6.2). Then, we identify several research gaps and propose research directions.

6.1 Implicitly Defined Social Comprehension

Yousuf et al. [137] modeled the problem of how a robot guide at a museum should approach a group of people to explain an exhibit. They based their model on previous proxemics and F-Formations, and define different approaching behaviors that depend on the number of persons looking at the exhibition and the robot. People's answers to a questionnaire reveal that they prefer the proposed system when compared to one that does not consider people's attention. Another work focuses on the interaction potential of approaching behaviors [79] for a holonomic robot. For an interaction to be successful, the robot must also be in a position where its sensors can capture people's information efficiently. Thus, they propose a solution that computes the engagement pose and maintains an appropriate distance to a human subject based on proxemics and the overall accuracy of the robot's sensors. In [126], the authors compute approaching areas taking into consideration proxemics, the human field of view, and social interactions. Then, they choose the center of the closest approaching area as the robot approach goal, with the robot facing the center of the interaction area (o-space for a group), or facing a single person. They further enhance their method in [125], being able to approach moving pedestrians (linear prediction of their movements) and groups gazing at objects. Other researchers [115] focused on the problem of a robot that approaches people to distribute flyers. Their work studies the approaching behaviors and whoom to interact with to maximize the number of distributed flyers. These works use proxemics and linear models to predict people's movements and act accordingly. A different approach is to use the social force model, as shown by [99]. They attempt to solve the problem of a human–robot duo approaching another person. A combination of forces draws the robot to the goal person while making it keep an appropriate distance from the accompanying person and avoiding objects and other people.

A different approach consists of learning the model that governs the scene's social norms through behavioral demonstration. In [96], the robot learns to approach one person through Inverse Reinforcement Learning. The state representation is a circular grid centered in the person, with a polar representation. The reward function is a linear combination of functions of state-action pairs. An expert controlled the robot remotely to approach the person, thus gathering the approach demonstrations. The robot can then use the learned reward function in two ways. The first way is to use it to solve the MDP, fitting a bézier curve to smooth the trajectory. The second way is to create a costmap with where each state has an associated Radial Basis Function weighted by learned reward

function weights. Dondrup and Hanheide [33] propose a distinct approach, also learned from demonstrations. Their trajectory planning method takes future navigation actions of robots and humans that move near each other into account. They propose a Qualitative Trajectory Calculus (QTC) witch consists of a spatial representation that encodes human–robot velocity interaction rules from demonstrations. Their training data consists of vectors with QTC states of humans and QTC states of the robot. With them, they create a conditional probability table to predict the appropriate robot action given a human observation. Predicted robot actions are then used to build velocity costmaps that limit trajectories sampled by a Dynamic Window Approach (DWA) local planner [42].

Researchers have also used Neural networks to tackle this problem. For instance, Yang and Peters train Long Short Term Memories (LSTM) on a semi-synthetic dataset to approach small groups of people. The authors of [48], they use a Generative Adversarial Network (GAN) and LSTMs to predict people's future trajectories given trajectory segments. Similarly, [135] generates approaching trajectories into free-standing conversational groups, given a training set of safe and socially acceptable paths.

6.2 Explicitly Defined Social Comprehension

None of the previous works explicitly defines social rules. The authors of [24] developed a framework for an explicit social rule execution for Petri Nets. Their work generates a Petri Net Plan that considers a set of social norms. Furthermore, they provide a formal definition of social norms for a robot. Porfirio and colleagues [89] developed an interaction design interface and a verification algorithm to test whether a human designed interaction scripts respect a set of previously encoded social norms. They model interactions with a state-machine like formulation (transition-system) and represent social norms using Linear Temporal Logic (LTL). Transitions between states occur when the robot detects human actions. The authors manually encoded social norms in LTL.

6.3 Research Gaps

In most works, the underlying algorithms (for navigation, for instance) implicitly encode the social rules. Thus, even though it is possible to tune some parameters, there is no explicit way to incorporate new norms. A social robot that follows a human-centered design must be able to perceive and incorporate social norms explicitly. Learning social norms through deep learning methods poses several application problems. While humans can make sense of them either after having them explained to them or through few observations, these methods require a prohibitive number of observations to learn models that encode the norms. There are also safety concerns about these methods. Even though a costmap based solution, as shown by [96] or training the robot in simulation [28] could reduce dangerous situations, the robot's behaviors can be unpredictable since the model's internal representation is often impossible to interpret. Thus, interpretable models like Carlucci et al.'s [24] and Porfirio et al.'s [89] may provide stronger safety guarantees. However, these do not learn from the data or demonstrations, thus requiring a human expert to design the interaction.

For social navigation-related algorithms, we also identify several research opportunities. The first one is that none of these methods adapt proxemics to the free space of the scene. Thus, if the scene is very cluttered, and the robot does not adapt its social costmap, it will not be able to navigate and approach people. The second research gap is related to the scene's semantic information. While the analyzed works do not consider it when the robot engages with people, this information is fundamental to plan and approach people without disturbing their interactions with the environment and each other. A possible way to address this issue is to explore objects' affordances and affordance spaces. With this social insight, a robot can, for instance, navigate without blocking the path of transient pedestrians in doorways and corridors.

7 Communication

The detected social context (Sect. 5) together with social insight (Sect. 6) allow social agents to understand the interaction and guide their communication behaviors. Here, we describe works that implement the skills to non-verbally communicate one's intentions and feelings (Sect. 7.1—referential communication), as well as communication strategies to guide the interaction toward one's goals (Sect. 7.2—social problem solving). We finalize the section by highlighting research gaps.

7.1 Referential Communication

Non-verbal communication skills are necessary to initiate a successful interaction. People with whom the robot intends to interact need to be aware of the robot's intentions, otherwise it risks being ignored. Thus, it is necessary to express one's intentions on time, especially when relying upon non-verbal behaviors. This observation is supported by [115] since the success of their flyer distributing robot depends on the timing of the robot's arm. Their best strategy was to have the robot approach the pedestrian and extend its arm nearby while gazing at the target person.

For a robot to initiate an interaction with people, it must be able to greet them in a socially acceptable way. The handshake is the most common greeting behavior in western civilization. There is some literature on the development of human–robot handshakes, even though most of it focuses

on the shaking motion. For instance, [57] studies the handshake motions between two human participants. They studied the velocity profile of human writs during handshake request and response and modeled a transfer function to generate the motion of the respondent based on the requester's movements. They implemented it on a robotic hand and performed a perception study with humans to test their method for several parameters. In one of their subsequent works [58], they adapt their model for small-sized robot arms. Later, they study the best arm and gaze movements for their robot to request a handshake [59]. Following this, they studied the timings and the lag between the start of a request of a handshake and the start of a response [85] [86]. In one of our past works [11], we implemented a handshake system on the Vizzy robot. We used information from the robot's Hall-effect-based tactile sensors [88] to control the robot's grip force with a PID controller and detect whether the handshake grasped a human hand or not with a K-Nearest Neighbors classifier with Dynamic Time Warping. People rated the handshake grip positively in terms of perceived enjoyment and safety. More recently, Mura and colleagues [83] implemented a human–robot handshake controller on a FRANKA robot arm with a custom silicon glove with pressure sensors. Their work focuses on stiffness and synchronization, and they use an EKF to learn human handshake sinusoidal motion parameters. They use hand pressure information as a control signal for arm stiffness control and hand closure control. Their results show that people positively evaluated the handshake and that people perceive distinct personality qualities with different motion controllers.

However, a social robot cannot be limited to handshake greetings, and individually modeling each behavior can become troublesome. A possible approach to have multiple greeting behaviors is to imitate humans. In [6], the authors propose and test two imitation learning algorithms: (i) Probabilistic Principal Component Analysis-Interaction Model, and (ii) Path Map-Interaction Model. They train their algorithms with motion capture data of two humans interacting. Later, they propose Interaction Primitives [7], an algorithm that learns the dependency between two agents' actions and follows the human action with the appropriate robot motion.

The previous algorithms require a motion capture of the humans' interactions, which still requires a considerable amount of time and extra equipment. A better option would be for the robot to learn these behaviors directly from cheaper sensors, which was proposed by Shu et al. [117]. From RGB-D data containing human–human interactions, they attempt to learn action possibilities that follow social norms (which they define as "social affordances") and perform real-time inference based on the learned interactions. They test the following behaviors with a Baxter robot: (i) handshake, (ii) hand wave, (iii) high five, (iv) pull up, and (v) hand over a cup.

7.2 Social Problem Solving

Qureshi and colleagues [94] use a Multimodal Deep Q-Network to make a robot learn when to use one of four behaviors: (i) wait; (ii) look toward a human; (iii) wave the hand; and (iv) handshake. The network takes grayscale and depth images and learns to choose one of the four actions. The robot receives a positive reward for a successful handshake (someone touches the robot's hand) and a negative reward for a negative one. In one of their recent works [95], they use an extra network to predict people's reactions (smile, eye contact, or smile) for each possible robot action. The reward function of the Q-net is computed based on the predicted reaction and the actual reaction of the person. In recent work, Porfirio and colleagues [90] used a state-machine-like formulation with Linear Temporal Logic (LTL) to update the interaction script from human feedback. They defined "interaction traces" as sequences of robot states and human actions. Through human–robot interaction, they ask humans to rate the robot's traces as positive $(+1)$, neutral (0), or negative (-1). They propose an adaptation algorithm the edits the script to maximize the score while complying with social norms encoded with LTL. A user study showed that the adapted model significantly improved user experience in the interaction.

7.3 Research Gaps

During the first encounter between a robot and a person, the robot does not have information about the person's culture. It can have priors related to its current location, but that is no longer a piece of strong information in an increasingly multicultural society. Thus, the robot must be able to switch between greeting models in real-time to match the subject's greeting. Finally, given the lack of works were the robot detects that it misbehaved, there are also no works were it automatically apologizes after getting negative feedback. Even though sometimes apologizing is not the best strategy [37], the robot must be aware of the human's dissatisfaction and employ the best recovery behavior for the situation. For instance, the robot could appologise to people after receiving negative social feedback, and attempt to explain why it failed. It is also important to study how the incongruency effect manifests in human–robot interaction and how distinct robot repair behaviors can lessen its effects.

8 Conclusions and Future Directions

In this survey, we covered the existing body of knowledge on robots that engage humans in first encounters and the necessary skills for perception, reasoning, and action. The current state-of-the-art still needs considerable improve-

ments to make graceful engagement between humans and robots a reality. We proposed a taxonomy based on Kendon's and Greenspan's models to analyze and categorize the surveyed works, covered methods used to open interaction with people in first encounters, and went through the state-of-the-art of individual skills needed do so.

We found that research works that implement robot architectures to approach people do not follow the same taxonomy. However, an analysis of the interaction stages of these works allowed us to classify their interaction stages under the same taxonomy, following Kendon's greeting model. This way, we could compare them and identify their gaps. None of the covered methods fully implements all stages of Kendon's model, and we could not find comparisons between them under the same conditions. Besides, although Kendon's greeting model results from human behavior observations, one may ask whether robots can learn to engage people in first encounters even more effectively than humans. Analyzed works used state-machines, thus assuming that the interaction stage was perfectly known. As stated by [54], such an approach might not be robust to errors. The exchange of social signals involved in the greeting to open the interaction is a tool for humans to keep track of the interaction stage [64], and even with them, human–human first encounters can fail, as reported by Schiffrin [110]. From this discussion, we highlight the following open questions:

- Which theories and methods can make a robot successfully open interaction with a human in a first encounter?
- How do models in the literature perform against each other and a fully implemented Kendon's model?
- Which techniques can we use to manage and track the interaction, with robustness to uncertain observations and imperfect models?
- Can data-driven methods improve human-designed interaction opening scripts for first encounters?

We believe that social context inference is a crucial topic for first encounters. Current technologies already detect significant information for autonomous robots. However, being able to integrate several reliable perceptual modules without using external computational power can be very difficult if these modules are computationally intensive, as is the case of several state-of-the-art deep learning methods [53,63,97,132]. A possible research direction would be to study how distinct perceptual algorithms could share features and information to improve the results. That is the approach used in OpenHeadPose [23] that leverages the knowledge of convolutional pose machines to estimate the head pose. These observations lead to the following open question:

- How can perceptual skills of social sensitivity be integrated in a robust and computationally efficient way?

Being able to define and learn social norms explicitly may improve their design and the system's explainability. Carlucci and colleagues [24] used Petri-nets to represent social norms explicitly, but they are hand-designed. A possible direction would be to explore behavior trees [29] as a representation of social norms. Some methods can learn behavior trees for robot control [14], and it may be possible to enhance them to learn social norms as behavior trees. As for navigation, the literature has not yet considered the need to adapt encoded social norms to react to environmental constraints. That is a relevant feature since there can be situations where the robot might need to violate social norms to engage with the target. Given these insights, we highlight the following open questions:

- How can a robot represent and learn social norms and scripts to open first encounters?
- How can the robot adapt norms and scripts to cope with dynamic navigation restrictions?

To be able to communicate its intentions to interact and comply with the social scripts in a first encounter, the robot should be able to adapt the salutation to match the one used by the interaction target. Moreover, we believe it would be interesting to teach new salutations to robots through human demonstrations. Thus, the following question arises:

- How can we develop nonverbal communication skills and strategies to open interaction during first encounters effectively?

Finally, we believe that social feedback should have more information than positive and negative values. Contextual information may give meaning to social feedback and reduce the search space of behaviors. Moreover, following the suggestion of the previous paragraph, a behavior tree could learn which recovery behavior might be appropriate after receiving negative feedback due to norm violations. The following open questions cover these problems:

- Which methods can perceive signals of social feedback?
- How can robots learn communication strategies to recover from failures during interaction-opening in first encounters?

Throughout our survey, we identified difficult problems that make interaction opening during first encounters an open challenge. Our analysis identified significant research gaps in all categories. We believe that strong multidisciplinary collaborations between the robotics, psychology, and sociology

communities are a powerful way to address these open challenges.

Funding Information This work was funded by Fundação para a Ciência e a Tecnologia through grant SFRH/BD/133098/2017 and supported by the LARSyS - FCT Project UIDB/50009/2020.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Abelson RP (1981) Psychological status of the script concept. Am Psychol 36(7):715–729. https://doi.org/10.1037/0003-066x.36.7.715
2. Ahmad M, Mubin O, Orlando J (2017) A systematic review of adaptivity in human–robot interaction. Multimodal Technol Interact 1(3):14. https://doi.org/10.3390/mti1030014
3. Ahmad MI (2018) An emotion and memory model for social robots: A long-term interaction. PhD thesis, Western Sydney University (Australia)
4. Alletto S, Serra G, Calderara S, Cucchiara R (2015) Understanding social relationships in egocentric vision. Pattern Recognit 48(12):4082–4096. https://doi.org/10.1016/j.patcog.2015.06.006
5. Ambady N, Skowronski JJ (2008) First impressions. Guilford Press, New York
6. Amor HB, Vogt D, Ewerton M, Berger E, Jung B, Peters J (2013) Learning responsive robot behavior by imitation. In: 2013 IEEE/RSJ international conference on intelligent robots and systems, IEEE, https://doi.org/10.1109/iros.2013.6696819
7. Amor HB, Neumann G, Kamthe S, Kroemer O, Peters J (2014) Interaction primitives for human-robot cooperation tasks. In: 2014 IEEE international conference on robotics and automation (ICRA), IEEE, https://doi.org/10.1109/icra.2014.6907265
8. Amos B, Ludwiczuk B, Satyanarayanan M (2016) Openface: A general-purpose face recognition library with mobile applications. Tech. rep., CMU-CS-16-118, CMU School of Computer Science
9. Argyle M (1988) Bodily communication, 2nd edn. Methuen Publishing, London
10. Asch SE (1946) Forming impressions of personality. J Abnorm Soc Psychol 41(3):258–290. https://doi.org/10.1037/h0055756
11. Avelino J, Paulino T, Cardoso C, Nunes R, Moreno P, Bernardino A (2018) Towards natural handshakes for social robots: human-aware hand grasps using tactile sensors. Paladyn J Behav Robotics 9(1):221–234. https://doi.org/10.1515/pjbr-2018-0017
12. Avelino J, Gonçalves A, Ventura R, Garcia-Marques L, Bernardino A (2020) Collecting social signals in constructive and destructive events during human-robot collaborative tasks. In: Companion of the 2020 ACM/IEEE international conference on human–robot interaction, association for computing machinery, New York, NY, USA, HRI '20, p 107–109
13. Baltrusaitis T, Zadeh A, Lim YC, Morency LP (2018) OpenFace 2.0: Facial behavior analysis toolkit. In: 2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018), IEEE, https://doi.org/10.1109/fg.2018.00019
14. Banerjee B (2018) Autonomous acquisition of behavior trees for robot control. In: 2018 IEEE/RSJ international conference on intelligent robots and systems (IROS), IEEE, https://doi.org/10.1109/iros.2018.8594083
15. Banerjee S, Silva A, Chernova S (2018) Robot classification of human interruptibility and a study of its effects. ACM Transn Human–Robot interact 7(2):1–35. https://doi.org/10.1145/3277902
16. Bazzani L, Cristani M, Tosato D, Farenzena M, Paggetti G, Menegaz G, Murino V (2012) Social interactions by visual focus of attention in a three-dimensional environment. Expert Syst 30(2):115–127. https://doi.org/10.1111/j.1468-0394.2012.00622.x
17. Bicchieri C (2005) The grammar of society: the nature and dynamics of social norms. Cambridge University Press, Cambridge
18. Bracken CC, Jeffres LW, Neuendorf KA (2004) Criticism or praise? the impact of verbal versus text-only computer feedback on social presence, intrinsic motivation, and recall. CyberPsychol Behav 7(3):349–357. https://doi.org/10.1089/1094931041291358
19. vom Brocke J, Simons A, Niehaves B, Riemer K, Plattfaut R, Cleven A (2009) Reconstructing the giant: On the importance of rigour in documenting the literature search process. In: ECIS 2009 proceedings
20. vom Brocke J, Simons A, Riemer K, Niehaves B, Plattfaut R, Cleven A (2015) Standing on the shoulders of giants: Challenges and recommendations of literature search in information systems research. Commun Assoc Inf Syst. https://doi.org/10.17705/1cais.03709
21. Broekens J (2007) Emotion and reinforcement: affective facial expressions facilitate robot learning. In: Huang TS, Nijholt A, Pantic M, Pentland A (eds) Artif Intell Human Comput. Springer, Berlin, Heidelberg, pp 113–132
22. Brščić D, Ikeda T, Kanda T (2017) Do you need help? a robot providing information to people who behave atypically. IEEE Trans Robot 33(2):500–506. https://doi.org/10.1109/tro.2016.2645206
23. Cao Y, Canévet O, Odobez JM (2018) Leveraging convolutional pose machines for fast and accurate head pose estimation. In: 2018 IEEE/RSJ international conference on intelligent robots and systems (IROS), pp 1089–1094, https://doi.org/10.1109/IROS.2018.8594223
24. Carlucci FM, Nardi L, Iocchi L, Nardi D (2015) Explicit representation of social norms for social robots. In: 2015 IEEE/RSJ international conference on intelligent robots and systems (IROS), IEEE, https://doi.org/10.1109/iros.2015.7353970
25. Cavell T (1990) Social adjustment, social performance, and social skills: a tri-component model of social competence. J Clin Child Adolesc Psychol 19(2):111–122. https://doi.org/10.1207/s15374424jccp1902_2
26. Charalampous K, Kostavelis I, Gasteratos A (2017) Recent trends in social aware robot navigation: a survey. Robot Auton Syst 93:85–104. https://doi.org/10.1016/j.robot.2017.03.002
27. Chen CY, Grauman K (2016) Subjects and their objects: localizing interactees for a person-centric view of importance. Int J Comput Vis 126(2–4):292–313. https://doi.org/10.1007/s11263-016-0958-6
28. Chen YF, Everett M, Liu M, How JP (2017) Socially aware motion planning with deep reinforcement learning. In: 2017 IEEE/RSJ international conference on intelligent robots and systems (IROS), IEEE, https://doi.org/10.1109/iros.2017.8202312
29. Colledanchise M (2018) Behavior trees in robotics and Al. CRC Press, Boca Raton. https://doi.org/10.1201/9780429489105
30. Correia F, Guerra C, Mascarenhas S, Melo FS, Paiva A (2018) Exploring the impact of fault justification in human–robot trust. In: Proceedings of the 17th international conference on autonomous agents and multi agent systems, international foundation for autonomous agents and multiagent systems, Richland, SC, AAMAS '18, p 507–513

31. Crick NR, Dodge KA (1994) A review and reformulation of social information-processing mechanisms in children's social adjustment. Psychol Bull 115(1):74
32. Cristani M, Bazzani L, Paggetti G, Fossati A, Tosato D, Bue AD, Menegaz G, Murino V (2011) Social interaction discovery by statistical analysis of F-formations. In: Procedings of the British machine vision conference 2011, British Machine Vision Association, https://doi.org/10.5244/c.25.23
33. Dondrup C, Hanheide M (2016) Qualitative constraints for human-aware robot navigation using velocity costmaps. In: 25th IEEE international symposium on robot and human interactive communication (RO-MAN). IEEE. https://doi.org/10.1109/roman.2016.7745177
34. Drew P, Raymond G, Weinberg D (2006) Talk and interaction in social research methods. SAGE Publications Ltd, California. https://doi.org/10.4135/9781849209991
35. DuBois DL, Felner RD (1996) The quadripartite model of social competence: theory and applications to clinical intervention. Cognitive therapy with children and adolescents: a casebook for clinical practice. The Guilford Press, New York
36. Ekman P, Friesen WV (1969) The repertoire of nonverbal behavior: categories, origins, usage, and coding. Semiotica. https://doi.org/10.1515/semi.1969.1.1.49
37. Engelhardt S, Hansson E, Leite I (2017) Better faulty than sorry: investigating social recovery strategies to minimize the impact of failure in human–robot interaction. In: WCIHAI@ IVA, pp 19–27
38. Ethofer T, Stegmaier S, Koch K, Reinl M, Kreifelts B, Schwarz L, Erb M, Scheffler K, Wildgruber D (2020) Are you laughing at me? neural correlates of social intent attribution to auditory and visual laughter. Hum Brain Mapp 41(2):353–361. https://doi.org/10.1002/hbm.24806
39. Fehr E, Fischbacher U (2004) Social norms and human cooperation. Trends Cogn Sci 8(4):185–190. https://doi.org/10.1016/j.tics.2004.02.007
40. Fong T, Nourbakhsh I, Dautenhahn K (2003) A survey of socially interactive robots. Robot Auton Syst 42(3):143–166 socially Interactive Robots
41. Foster ME, Alami R, Gestranius O, Lemon O, Niemelä M, Odobez JM, Pandey AK (2016) The mummer project: Engaging human-robot interaction in real-world public spaces. In: Agah A, Cabibihan JJ, Howard AM, Salichs MA, He H (eds) Social Robotics. Springer International Publishing, Cham, pp 753–763
42. Fox D, Burgard W, Thrun S (1997) The dynamic window approach to collision avoidance. IEEE Robot Autom Magn 4(1):23–33
43. Garfinkel H (1967) Studies in ethnomethodology. Prentice-Hall, Englewood Cliffs, NJ
44. Gelfand MJ, Raver JL, Nishii L, Leslie LM, Lun J, Lim BC, Duan L, Almaliach A, Ang S, Arnadottir J, Aycan Z, Boehnke K, Boski P, Cabecinhas R, Chan D, Chhokar J, D'Amato A, Ferrer M, Fischlmayr IC, Fischer R, Fülöp M, Georgas J, Kashima ES, Kashima Y, Kim K, Lempereur A, Marquez P, Othman R, Overlaet B, Panagiotopoulou P, Peltzer K, Perez-Florizno LR, Ponomarenko L, Realo A, Schei V, Schmitt M, Smith PB, Soomro N, Szabo E, Taveesin N, Toyama M, Van de Vliert E, Vohra N, Ward C, Yamaguchi S (2011) Differences between tight and loose cultures: a 33-nation study. Science 332(6033):1100–1104. https://doi.org/10.1126/science.1197754
45. Goldfried MR, D'Zurilla TJ (1969) A behavioral-analytic model for assessing competence. Current topics in clinical and community psychology. Elsevier, Armsterdam, pp 151–196. https://doi.org/10.1016/b978-1-4831-9972-6.50009-3
46. Gordon G, Spaulding S, Westlund JK, Lee JJ, Plummer L, Martinez M, Das M, Breazeal C (2016) Affective personalization of a social robot tutor for children's second language skills. In: Thirtieth AAAI conference on artificial intelligence
47. Greenspan S (1981) Defining childhood social competence: a proposed working model. Adv Spec Educ 3:1–39
48. Gupta A, Johnson J, Fei-Fei L, Savarese S, Alahi A (2018) Social GAN: socially acceptable trajectories with generative adversarial networks. In: IEEE conference on computer vision and pattern recognition (CVPR), CONF
49. Hall ET (1966) The hidden dimension, vol 609. Doubleday, Garden City, NY
50. Hastie R (1980) Person memory: the cognitive basis of social perception. Lawrence Erlbaum Associates, New Jersey
51. Hastie R, Kumar PA (1979) Person memory: personality traits as organizing principles in memory for behaviors. J Pers Soc Psychol 37(1):25–38. https://doi.org/10.1037/0022-3514.37.1.25
52. Hayes N (2000) Foundations of psychology. 3rd edn. Cengage learning EMEA
53. He W, Motlicek P, Odobez JM (2018) Deep neural networks for multiple speaker detection and localization. In: 2018 IEEE international conference on robotics and automation (ICRA), IEEE, https://doi.org/10.1109/icra.2018.8461267
54. Heenan B, Greenberg S, Aghel-Manesh S, Sharlin E (2014) Designing social greetings in human robot interaction. In: Proceedings of the 2014 conference on designing interactive systems—DIS' 14, ACM Press, https://doi.org/10.1145/2598510.2598513
55. Ishiguro H, Ono T, Imai M, Maeda T, Kanda T, Nakatsu R (2001) Robovie: an interactive humanoid robot. Ind Robot Int J 28:498–504
56. Jerónimo R, Garcia-Marques L, Ferreira MB, Macrae CN (2015) When expectancies harm comprehension: encoding flexibility in impression formation. J Exp Soc Psychol 61:110–119. https://doi.org/10.1016/j.jesp.2015.07.007
57. Jindai M, Watanabe T (2007) Development of a handshake robot system based on a handshake approaching motion model. In: IEEE/ASME international conference on advanced intelligent mechatronics. IEEE. https://doi.org/10.1109/aim.2007.4412423
58. Jindai M, Watanabe T (2010) A small-size handshake robot system based on a handshake approaching motion model with a voice greeting. In: 2010 IEEE/ASME international conference on advanced intelligent mechatronics, IEEE, https://doi.org/10.1109/aim.2010.5695738
59. Jindai M, Watanabe T (2011) Development of a handshake request motion model based on analysis of handshake motion between humans. In: 2011 IEEE/ASME international conference on advanced intelligent mechatronics (AIM), IEEE, https://doi.org/10.1109/aim.2011.6026975
60. Kanda T, Ishiguro H (2017) Human–robot interaction in social robotics. CRC Press, Boca Raton. https://doi.org/10.1201/b13004
61. Kanda T, Ishiguro H, Imai M, Ono T (2004) Development and evaluation of interactive humanoid robots. Proc IEEE 92(11):1839–1850
62. Kato Y, Kanda T, Ishiguro H (2015) May i help you?—design of human-like polite approaching behavior. In: 2015 10th ACM/IEEE international conference on human–robot interaction (HRI), IEEE, pp 35–42
63. Kellnhofer P, Recasens A, Stent S, Matusik W, Torralba A (2019) Gaze360: physically unconstrained gaze estimation in the wild. In: IEEE international conference on computer vision (ICCV)
64. Kendon A (1991) Conducting interaction: patterns of behavior in focused encounters (Studies in Interactional Sociolinguistics). Cambridge University Press, Cambridge
65. Kenny DA (2004) PERSON: a general model of interpersonal perception. Pers Soc Psychol Rev 8(3):265–280. https://doi.org/10.1207/s15327957pspr0803_3
66. Knapp ML, Hall JA, Horgan TG (2013) Nonverbal communication in human interaction. Cengage Learning, Boston

67. Knox WB, Stone P (2009) Interactively shaping agents via human reinforcement. In: Proceedings of the fifth international conference on Knowledge capture—K-CAP'09, ACM Press, https://doi.org/10.1145/1597735.1597738
68. Kontogiorgos D, Pereira A, Sahindal B, van Waveren S, Gustafson J (2020) Behavioural responses to robot conversational failures. In: Proceedings of the 2020 ACM/IEEE international conference on human–robot interaction, ACM, https://doi.org/10.1145/3319502.3374782
69. Lazzeri N, Mazzei D, Cominelli L, Cisternino A, Rossi DD (2018) Designing the mind of a social robot. Appl Sci 8(2):302. https://doi.org/10.3390/app8020302
70. Li J (2015) The benefit of being physically present: a survey of experimental works comparing copresent robots, telepresent robots and virtual agents. Int J Human–Comput Stud 77:23–37
71. Linder T, Arras KO (2014) Multi-model hypothesis tracking of groups of people in RGB-D data. In: 17th International conference on information fusion (FUSION), pp 1–7
72. Loftin R, Peng B, MacGlashan J, Littman ML, Taylor ME, Huang J, Roberts DL (2015) Learning behaviors via human-delivered discrete feedback: modeling implicit feedback strategies to speed up learning. Auton Agents Multi-Agent Syst 30(1):30–59. https://doi.org/10.1007/s10458-015-9283-7
73. MacGlashan J, Ho MK, Loftin R, Peng B, Wang G, Roberts DL, Taylor ME, Littman ML (2017) Interactive learning from policy-dependent human feedback. In: Precup D, Teh YW (eds) Proceedings of the 34th international conference on machine learning, PMLR, International Convention Centre, Sydney, Australia, Proceedings of Machine Learning Research, vol 70, pp 2285–2294
74. Malle BF, Bello P, Scheutz M (2019) Requirements for an artificial agent with norm competence. In: Proceedings of the 2019 AAAI/ACM conference on AI, Ethics, and Society—AIES '19, ACM Press, https://doi.org/10.1145/3306618.3314252
75. Martins GS, Santos L, Dias J (2018) User-adaptive interaction in social robots: a survey focusing on non-physical interaction. Int J Soc Robot 11(1):185–205. https://doi.org/10.1007/s12369-018-0485-4
76. Massé B, Ba S, Horaud R (2018) Tracking gaze and visual focus of attention of people involved in social interaction. IEEE Trans Pattern Anal Mach Intell 40(11):2711–2724. https://doi.org/10.1109/TPAMI.2017.2782819
77. Massé B, Lathuilière S, Mesejo P, Horaud R (2019) Extended gaze following: Detecting objects in videos beyond the camera field of view. In: 2019 14th IEEE international conference on automatic face gesture recognition (FG 2019), pp 1–8, https://doi.org/10.1109/FG.2019.8756555
78. Mavridis N (2015) A review of verbal and non-verbal human–robot interactive communication. Robot Auton Syst 63:22–35. https://doi.org/10.1016/j.robot.2014.09.031
79. Mead R, Matarić MJ (2016) Autonomous human–robot proxemics: socially aware navigation based on interaction potential. Auton Robots 41(5):1189–1201. https://doi.org/10.1007/s10514-016-9572-2
80. Mirnig N, Stollnberger G, Miksch M, Stadler S, Giuliani M, Tscheligi M (2017) To err is robot: how humans assess and act toward an erroneous social robot. Front Robot AI. https://doi.org/10.3389/frobt.2017.00021
81. Mitsunaga N, Smith C, Kanda T, Ishiguro H, Hagita N (2008) Adapting robot behavior for human–robot interaction. IEEE Trans Robot 24(4):911–916. https://doi.org/10.1109/tro.2008.926867
82. Moreno P, Nunes R, Figueiredo R, Ferreira R, Bernardino A, Santos-Victor J, Beira R, Vargas L, Aragão D, Aragão M (2016) Vizzy: a humanoid on wheels for assistive robotics. In: Robot 2015: second Iberian robotics conference, Springer, pp 17–28
83. Mura D, Knoop E, Catalano MG, Grioli G, Bächer M, Bicchi A (2020) On the role of stiffness and synchronization in human–robot handshaking. Int J Robot Res. https://doi.org/10.1177/0278364920903792
84. Nigam A, Riek LD (2015) Social context perception for mobile robots. In: 2015 IEEE/RSJ international conference on intelligent robots and systems (IROS), IEEE, https://doi.org/10.1109/iros.2015.7353883
85. Ota S, Jindai M, Fukuta T, Watanabe T (2014) A handshake response motion model during active approach to a human. In: IEEE/SICE international symposium on system integration. IEEE. https://doi.org/10.1109/sii.2014.7028056
86. Ota S, Jindai M, Sasaki T, Ikemoto Y (2015) Handshake response motion model with approaching of human based on an analysis of human handshake motions. In: 2015 7th international congress on ultra modern telecommunications and control systems and workshops (ICUMT). IEEE. https://doi.org/10.1109/icumt.2015.7382396
87. Paetzel M, Perugia G, Castellano G (2020) The persistence of first impressions: The effect of repeated interactions on the perception of a social robot. In: Proceedings of the 2020 ACM/IEEE international Conference on human–robot interaction, association for computing machinery, New York, NY, USA, HRI '20, pp 73–82
88. Paulino T, Ribeiro P, Neto M, Cardoso S, Schmitz A, Santos-Victor J, Bernardino A, Jamone L (2017) Low-cost 3-axis soft tactile sensors for the human-friendly robot vizzy. In: 2017 IEEE international conference on robotics and automation (ICRA), IEEE, https://doi.org/10.1109/icra.2017.7989118
89. Porfirio D, Sauppé A, Albarghouthi A, Mutlu B (2018) Authoring and verifying human-robot interactions. In: The 31st annual ACM symposium on user interface software and technology—UIST'18, ACM Press, https://doi.org/10.1145/3242587.3242634
90. Porfirio D, Sauppé A, Albarghouthi A, Mutlu B (2020) Transforming robot programs based on social context. In: Proceedings of the 2020 CHI conference on human factors in computing systems, Association for Computing Machinery, New York, NY, USA, CHI '20, pp 1–12, https://doi.org/10.1145/3313831.3376355
91. Portugal D, Santos L, Alvito P, Dias J, Samaras G, Christodoulou E (2015) Socialrobot: an interactive mobile robot for elderly home care. In: 2015 IEEE/SICE international symposium on system integration (SII), pp 811–816
92. Powers A, Kiesler S, Fussell S, Torrey C (2007) Comparing a computer agent with a humanoid robot. In: Proceedings of the ACM/IEEE international conference on human–robot interaction, ACM, New York, NY, USA, HRI '07, pp 145–152
93. von der Pütten AM, Krämer NC, Gratch J, Kang SH (2010) "it doesn't matter what you are!" explaining social effects of agents and avatars. Comput Hum Behav 26(6):1641–1650. https://doi.org/10.1016/j.chb.2010.06.012
94. Qureshi AH, Nakamura Y, Yoshikawa Y, Ishiguro H (2016) Robot gains social intelligence through multimodal deep reinforcement learning. In: 2016 IEEE-RAS 16th international conference on humanoid robots (Humanoids), IEEE, https://doi.org/10.1109/humanoids.2016.7803357
95. Qureshi AH, Nakamura Y, Yoshikawa Y, Ishiguro H (2018) Intrinsically motivated reinforcement learning for human–robot interaction in the real-world. Neural Netw 107:23–33. https://doi.org/10.1016/j.neunet.2018.03.014
96. Ramirez OAI, Khambhaita H, Chatila R, Chetouani M, Alami R (2016) Robots learning how and where to approach people. In: 25th IEEE international symposium on robot and human interactive communication (RO-MAN). IEEE. https://doi.org/10.1109/roman.2016.7745154
97. Ravanelli M, Parcollet T, Bengio Y (2019) The Pytorch–Kaldi speech recognition toolkit. In: In Proc of ICASSP

98. Reeves B, Nass CI (1996) The media equation: how people treat computers, television, and new media like real people and places. Cambridge University Press, Cambridge
99. Repiso E, Garrell A, Sanfeliu A (2018) Robot approaching and engaging people in a human–robot companion framework. In: 2018 IEEE/RSJ international conference on intelligent robots and systems (IROS), IEEE, https://doi.org/10.1109/iros.2018.8594149
100. Riggio RE, Friedman HS (1986) Impression formation: the role of expressive behavior. J Pers Soc Psychol 50(2):421–427. https://doi.org/10.1037/0022-3514.50.2.421
101. Rios-Martinez J, Spalanzani A, Laugier C (2014) From proxemics theory to socially-aware navigation: a survey. Int J Soc Robot 7(2):137–153. https://doi.org/10.1007/s12369-014-0251-1
102. Ritschel H, Baur T, André E (2017) Adapting a robot's linguistic style based on socially-aware reinforcement learning. In: 2017 26th IEEE international symposium on robot and human interactive communication (RO-MAN), pp 378–384, https://doi.org/10.1109/ROMAN.2017.8172330
103. Rossi S, Ferland F, Tapus A (2017) User profiling and behavioral adaptation for HRI: a survey. Pattern Recognit Lett 99:3–12. https://doi.org/10.1016/j.patrec.2017.06.002
104. Rubio F, Valero F, Llopis-Albert C (2019) A review of mobile robots: concepts, methods, theoretical framework, and applications. Int J Adv Robot Syst 16(2):1729881419839,596. https://doi.org/10.1177/1729881419839596
105. Saad E, Broekens J, Neerincx MA, Hindriks KV (2019) Enthusiastic robots make better contact. In: 2019 IEEE/RSJ international conference on intelligent robots and systems (IROS), IEEE
106. Satake S, Kanda T, Glas DF, Imai M, Ishiguro H, Hagita N (2010) How to approach humans? strategies for social robots to initiate interaction. J Robot Soc Jpn 28(3):327–337. https://doi.org/10.7210/jrsj.28.327
107. Satake S, Kanda T, Glas DF, Imai M, Ishiguro H, Hagita N (2013) A robot that approaches pedestrians. IEEE Trans Robot 29(2):508–524. https://doi.org/10.1109/tro.2012.2226387
108. Saunderson S, Nejat G (2019) How robots influence humans: a survey of nonverbal communication in social human–robot interaction. Int J Soc Robot 11(4):575–608. https://doi.org/10.1007/s12369-019-00523-0
109. Schank RC, Abelson RP (1977) Scripts, plans, goals and understanding: an inquiry into human knowledge structures. Scripts, plans, goals and understanding: a inquiry into human knowledge structures., Lawrence Erlbaum, Oxford, England
110. Schiffrin D (1977) Opening encounters. Am Sociol Rev 42(5):679. https://doi.org/10.2307/2094858
111. Setti F, Cristani M (2019) Evaluating the group detection performance: the GRODE metrics. IEEE Trans Pattern Anal Mach Intell 41(3):566–580. https://doi.org/10.1109/tpami.2018.2806970
112. Setti F, Lanz O, Ferrario R, Murino V, Cristani M (2013) Multi-scale F-formation discovery for group detection. In: 2013 IEEE International conference on image processing, IEEE, https://doi.org/10.1109/icip.2013.6738732
113. Setti F, Russell C, Bassetti C, Cristani M (2015) F-formation detection: individuating free-standing conversational groups in images. PLoS ONE 10(5):e0123,783. https://doi.org/10.1371/journal.pone.0123783
114. Sheikhi S, Odobez JM (2015) Combining dynamic head pose-gaze mapping with the robot conversational state for attention recognition in human-robot interactions. Pattern Recognit Lett 66:81–90. https://doi.org/10.1016/j.patrec.2014.10.002
115. Shi C, Satake S, Kanda T, Ishiguro H (2017) A robot that distributes flyers to pedestrians in a shopping mall. Int J Soc Robot 10(4):421–437. https://doi.org/10.1007/s12369-017-0442-7
116. Shinozawa K, Naya F, Yamato J, Kogure K (2005) Differences in effect of robot and screen agent recommendations on human decision-making. Int J Human–Comput Stud 62(2):267–279
117. Shu T, Gao X, Ryoo MS, Zhu SC (2017) Learning social affordance grammar from videos: tansferring human interactions to human–obot interactions. In: 2017 IEEE international conference on robotics and automation (ICRA), IEEE, https://doi.org/10.1109/icra.2017.7989197
118. Sirithunge C, Jayasekara AGBP, Chandima DP (2019) Proactive robots with the perception of nonverbal human behavior: a review. IEEE Access 7:77,308–77,327. https://doi.org/10.1109/access.2019.2921986
119. Srull TK (1981) Person memory: some tests of associative storage and retrieval models. J Exp Psychol Hum Learn Mem 7(6):440–463. https://doi.org/10.1037/0278-7393.7.6.440
120. Srull TK, Lichtenstein M, Rothbart M (1985) Associative storage and retrieval processes in person memory. J Exp Psychol Learn Mem Cogn 11(2):316–345. https://doi.org/10.1037/0278-7393.11.2.316
121. Subramanyam R (2013) Art of reading a journal article: methodically and effectively. J Oral Maxillofac Pathol 17(1):65. https://doi.org/10.4103/0973-029x.110733
122. Sunnafrank M, Ramirez A (2004) At first sight: persistent relational effects of get-acquainted conversations. J Soc Pers Relatsh 21(3):361–379. https://doi.org/10.1177/0265407504042837
123. Triebel R, Arras K, Alami R, Beyer L, Breuers S, Chatila R, Chetouani M, Cremers D, Evers V, Fiore M, Hung H, Ramírez OAI, Joosse M, Khambhaita H, Kucner T, Leibe B, Lilienthal AJ, Linder T, Lohse M, Magnusson M, Okal B, Palmieri L, Rafi U, van Rooij M, Zhang L (2016) SPENCER: a socially aware service robot for passenger guidance and help in busy airports. Springer tracts in advanced robotics. Springer International Publishing, Cham, pp 607–622. https://doi.org/10.1007/978-3-319-27702-8_40
124. Trung P, Giuliani M, Miksch M, Stollnberger G, Stadler S, Mirnig N, Tscheligi M (2017) Head and shoulders: automatic error detection in human–robot interaction. In: Proceedings of the 19th ACM international conference on multimodal interaction—ICMI 2017, ACM Press, https://doi.org/10.1145/3136755.3136785
125. Truong X, Ngo T (2018) "to approach humans?": a unified framework for approaching pose prediction and socially aware robot navigation. IEEE Trans Cogn Dev Syst 10(3):557–572. https://doi.org/10.1109/TCDS.2017.2751963
126. Truong XT, Ngo TD (2016) Dynamic social zone based mobile robot navigation for human comfortable safety in social environments. Int J Soc Robot 8(5):663–684. https://doi.org/10.1007/s12369-016-0352-0
127. Tsiakas K, Abujelala M, Makedon F (2018) Task engagement as personalization feedback for socially-assistive robots and cognitive training. Technologies 6(2):49. https://doi.org/10.3390/technologies6020049
128. Vascon S, Mequanint EZ, Cristani M, Hung H, Pelillo M, Murino V (2015) A game-theoretic probabilistic approach for detecting conversational groups. In: Computer vision—ACCV 2014, Springer International Publishing pp 658–675, https://doi.org/10.1007/978-3-319-16814-2_43
129. Ventura R, Basiri M, Mateus A, Garcia J, Miraldo P, Santos P, Lima P (2016) A domestic assistive robot developed through robotic competitions. In: IJCAI 2016 workshop on autonomous mobile service robots, New York, USA
130. Weber K, Ritschel H, Aslan I, Lingenfelser F, André E (2018) How to shape the humor of a robot—social behavior adaptation based on reinforcement learning. In: Proceedings of the 2018 on international conference on multimodal interaction—ICMI '18, ACM Press, https://doi.org/10.1145/3242969.3242976

131. Webster J, Watson RT (2002) Analyzing the past to prepare for the future: writing a literature review. MIS quarterly pp 13–23
132. Wei SE, Ramakrishna V, Kanade T, Sheikh Y (2016) Convolutional pose machines. In: CVPR
133. Wood E, Baltruaitis T, Zhang X, Sugano Y, Robinson P, Bulling A (2015) Rendering of eyes for eye-shape registration and gaze estimation. In: 2015 IEEE international conference on computer vision (ICCV), IEEE, https://doi.org/10.1109/iccv.2015.428
134. Xu J, Howard A (2018) The impact of first impressions on human–robot trust during problem-solving scenarios. In: 2018 27th IEEE international symposium on robot and human interactive communication (RO-MAN), pp 435–441
135. Yang F, Peters C (2019) Appgan: generative adversarial networks for generating robot approach behaviors into small groups of people. In: ROMAN'19
136. Yoshioka G, Sakamoto T, Takeuchi Y (2018) Polite approach to engrossing person based on two-dimensional attitude of interaction with other. In: 27th IEEE international symposium on robot and human interactive communication (RO-MAN). IEEE. https://doi.org/10.1109/roman.2018.8525786
137. Yousuf MA, Kobayashi Y, Kuno Y, Yamazaki A, Yamazaki K, (2013) How to move towards visitors: a model for museum guide robots to initiate conversation. In: IEEE RO-MAN. IEEE. https://doi.org/10.1109/roman.2013.6628543
138. Zaraki A, Pieroni M, Rossi DD, Mazzei D, Garofalo R, Cominelli L, Dehkordi MB (2017) Design and evaluation of a unique social perception system for human–robot interaction. IEEE Trans Cogn Dev Syst 9(4):341–355. https://doi.org/10.1109/tcds.2016.2598423
139. Zhang L, Hung H (2016) Beyond F-formations: determining social involvement in free standing conversing groups from static images. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR), IEEE, https://doi.org/10.1109/cvpr.2016.123
140. Zhao M, Li D, Wu Z, Li S, Zhang X, Ye L, Zhou G, Guan D (2019) Stepped warm-up–the progressive interaction approach for human–robot interaction in public. In: Design, user experience, and usability. User experience in advanced technological environments. HCII 2019, Springer International Publishing, pp 309–327, https://doi.org/10.1007/978-3-030-23541-3_23

João Avelino is a Ph.D. Candidate at the Institute for Systems and Robotics at Instituto Superior Técnico, University of Lisbon. He has a background in Electrical and Computer engineering, majoring in Systems, Decision, Control & Robotics. He has research interests in Computer Vision, Machine Learning, Robotics, and Human-robot interaction. His main research focuses on socially aware engagement behaviors for mobile robots and the detection of social cues for human-robot interaction.

Leonel Garcia-Marques was supervised by David L. Hamilton and the topic was person memory and the incongruency effect. More recently, he has explored the interaction between memory and several other social and cognitive topics. His main research interests are focused in the interaction between memory and topics like learning, impression formation, stereotypes, and group processes. He is also interested in exploring the relationships between judgment and decision-making and learning. Finally, he is also concerned about philosophy of science and epistemology of statistics. He has published nearly 100 articles in national and international outlets. Some of them in some of the most prestigious journals of Cognitive and Social Psychology and received several research awards and coordinated FCT-funded projects. He has currently served as an associate editor for the Personality and Social Psychology Bulletin and he has also served before as an associate editor and editor-in-chief for the European Journal of Social Psychology. Now, he is working as a Head of the Department and Principal coordinator of CICPSI.

Rodrigo Ventura (Assistant Professor) received the Licenciatura (1996), M.Sc. (2000), and PhD degree (2008), in ECE from Instituto Superior Técnico (IST), Lisbon, Portugal. He is a (tenured) Assistant Professor at IST, and a member of Institute for Systems and Robotics (ISR-Lisboa). He has published more than 130 publications in international journals and conferences, on various topics intersecting Robotics and Artificial Intelligence. He is also co-inventor of several national and international patents on innovative solutions for robotic systems. He is a founding member of the Biologically-Inspired Cognitive Architecture society. He participated in several international and national research projects, including as Principal Investigator of national projects on field and service robots. Broadly, his research is focused on the intersection between Robotics and Artificial Intelligence, with particular interest in human-robot interaction, mobile manipulation, biologically inspired cognitive architectures, and decision making under uncertainty. This research is driven by applications in space robotics for both orbital and planetary environments, urban search and rescue robotics, aerial robots, and social service robots.

Alexandre Bernardino (PhD 2004) is a tenured Associate Professor at the Dept. of Electrical and Computer Engineering and Senior Researcher at the Computer and Robot Vision Laboratory of the Institute for Systems and Robotics at IST, the faculty of engineering of Lisbon University. He published 60+/170+ research papers in peer-reviewed international journals/conferences in the field of robotics, machine learning vision and cognitive systems, with more than 2500 citations (Scopus, h-index 21). He has graduated 13 PhD students and 80+ MSc students. He participated in 20+ nationals and international research projects, being the principal investigator in 5 of them. He is a Senior Member of the IEEE and the chaired the IEEE Portugal Robotics and Automation Chapter during 2015–19. His main research interests focus on the application of computer vision, machine learning, cognitive science and control theory to advanced robotics and automation systems.