University of Rome La Sapienza

Department of Computer, Control, and Management Engineering
Antonio Ruberti

Thesis For The Degree Of Doctor Of Philosophy

Developing Scenarios and Research Methodologies for Evaluating Human-Robot Interaction in Social Contexts

Thesis Advisor
Prof. Daniele Nardi

Correlatore
Prof. Andrea Marrella

Candidate
Lun Wang

XXXII cycle

For my family

Abstract

As a newly emerging field, Human-Robot Interaction (HRI) offers the unique opportunity to study the multi-discipline regarding robots, humans and interactions in social contexts. Understanding how people of different ages, gender and abilities best perceive from robot's behaviors remains an important challenge in the field. In this thesis, we aim to develop an HRI research approach that investigates *human behaviors toward robots* and *robot's behaviors perceived by humans* in social contexts. In this direction, the first part of this thesis introduces an overview of HRI, from its origins, and discusses the current state-of-the-art related to our research targets. The second part of the thesis presents research studies on *understanding human factors*, *developing evaluation techniques*, *designing task scenarios* and *conducting experimental studies*.

Specifically, we first introduce a confirmatory research study in a social context to investigate if user perception in HRI may be affected by human factors. Leveraging the standardized questionnaire Robot Social Attribute Scale (RoSAS), we analyze the quantitative data collected by the questionnaire and report the relevant findings (see Chapter 3).

Secondly, since questionnaires are among the most used evaluation technique in the field of social HRI, we develop an approach to design a new type of questionnaire as a task-driven evaluation technique for measuring user perception in social contexts. The approach consists of two steps. First, it relies on interviewing experts on HRI to understand which robot's behaviors can potentially affect the user perceptions during an HRI task. Then, it leverages a user survey to filter out those robot's behaviors that are not significantly relevant from the end user perspective. The results of the survey have allowed us to derive a final list of 17 behaviors to be captured in the questionnaire, which has been finally developed relying on a 5-point Likert-scale (see Chapter 4).

Thirdly, we employ the SciRoc challenge (that is a repeatable and general-purpose test method developed for HRI performance evaluation in a realistic social context) to introduce an exploratory research study to analyze the outcomes of the most sociable task scenario of SciRoc, namely "E4 - Take the elevator". The main novelty concerns the implementation of the research study in a realistic social robotic competition environment. We investigate robot's behaviors perceived by human by adopting the questionnaire developed and validate the reliability of the questionnaire as well (see Chapter 5).

Fourthly, we introduce a novel approach for designing HRI task scenarios in the context of the SciRoc challenge. The new approach consists of two steps, analyzing the elements of a task scenario and sketching out its layout (see Chapter 6). We conclude the second part of the thesis by introducing *qualitative research approaches for exploratory research study*, *quantitative research approach for confirmatory research study*, and *qualitative and quantitative (mixed-method) research approach* in the field of social HRI (see Chapter 7).

Finally, the third part summarizes the strengths and limitations of the research presented in this thesis, discussing potential future works in the field.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Human-robot interaction (HRI) is one of the most rapidly growing research fields in robotics and the most promising for the future of robotics technology [11]. As a newly emerging field, it has been gaining an increasing amount of interest by researchers and designers in the field of autonomous robotics [161], artificial intelligence (AI) [85, 168], natural language processing [59], social science, psychology science [83], ethology [66], as well as those in human-computer interaction (HCI) [29]. For instance, HCI offers a rich resource for research and design in HRI, as much has been learned in the last three decades about how people perceive and think about computer-based technologies [89]. However, HRI requires AI technologies for developing intelligent robots' functionalities such as recognition, understanding, communication, mapping and navigation. Natural language interaction has long been a topic of HRI research [87], with the additional challenge of spoken language understanding, often in noisy environments. Overall, HRI offers the unique opportunity to study multiple disciplines in a specific context (see Fig. 1.1).

Figure 1.1: HRI multidisciplinary introduced by Raviteja Burugu

Lasota et al. define HRI as a discipline that deals with collaboration, communication and cooperation between humans and robots [97]. Goodrich and Schultz define HRI as a field study dedicated to understanding, designing and evaluating robotic system for use by or with humans [65]. Therefore, HRI designers and researchers should have an *expanded* knowledge on *human factors*, *robotic system* and *interaction*. The three aspects mentioned are dependent, correlated and mutually influencing each other. For instance, Kanda et al. point out that affinity with the human is the important aspect of developing social robots and social interaction [83]; Bartneck et al. point out that embodiment makes social robots inherently different computing technologies [14].

For one thing, researchers of robotic systems aimed to develop both hardware and software in order to expand capability and functionality of robots (e.g., navigation, manipulation, perception, recognition, etc). In facts, robot's capabilities to interact with people in an entertaining, engaging, or seamless manner play an important role in social contexts [29], i.e., the robot's abilities to socially interact with humans and integrate into environments are crucial. For another thing, researchers of human features aimed to figure out the ideal set of human factors and behaviors that could supplement and augment a robot's social functionality. Hence, HRI researchers should understand limitations of robot's functionalities and preferences of human attributes and shape the interactions between humans and robots, i.e., they should make sure that the interaction modalities and the robot's functionalities developed fit the user individual attributes and user's needs.

When people interact, their behavior conveys affect alongside informational content. The behavioral messages are often unclear and open to multiple interpretations. Robots and humans should have mutual understandings of how behaviors ought to be interpreted and responded to emotionally [82]. For this reason, several HRI scholars addressed the development of robotic systems capable of recognizing human behavior and of adapting their behavior according to the models [142]. However, it is difficult to individualize human behaviors in social HRI contexts due to natural variations of human behaviors and numerous variations of environments.

To investigate human behaviors and their influences of human attributes and emotions, HRI researchers have developed several methods and metrics such as testing effectiveness of tasks [158], pshychophysiological human behaviors measurement [26] and behaviors measurement [16]. Despite few evaluation methods, approaches and metrics in this field, there have been a lot of studies concern perceived human behaviors and behaviors toward robots that investigate positive attitudes and negative attitudes [124].

To fill this gap, the first goal of this thesis is to study in-depth the large part of human factors, in particular, robot's behaviors perceived by humans and human behaviors toward robots, by proposing a specialized evaluation technique in this field.

The "standard location" to evaluate the performance of HRI is usually laboratory, for example, laboratories of universities. Because such evaluation study is easy to implement, and participants selected by HRI researchers are usually students. However, for a successful evaluation study, the environment and the user population should reflect realistically the application domain in order to guarantee natural interactions. To this end, we exploit robotic competitions which provide good opportunities for conducting evaluation studies in the field of HRI. Robotic competitions can implement a *realistic* environment and involve *real* participants emphasizing HRI focus. For this reason, we design and implement the first SciRoc challenge (Smart CIties RObotic challenges) [1] as a scientific competition which allows the measurement of performance and the robot's behaviors perceived by humans, i.e., SciRoc is a repeatable and general-purpose test method (benchmark) developed for HRI teaming performance evaluation investigating users' attitudes using HRI research methodology. In the context of the SciRoc challenge, we provided a new approach for designing task scenarios which comply with four principles: *suitability*, *usability*, *continuity* and *applicability.*

[1] https://sciroc.org/

Second, we concretely conduct two research studies: *exploratory research study (i.e., qualitative approach)* and *confirmatory research study (i.e., quantitative approach).* In the end, we analyze the findings concerning task performance, robot's behaviors perceived by humans, human attributes and human roles differences.

Hence, the second goal of this thesis is to introduce HRI research studies in robotic scientific competition contexts, from designing task scenarios to analyzing HRI findings, by emphasizing the characteristics and the advantages of the main novelties.

HRI researchers investigate the concerned research problems by conducting *exploratory research study* or *confirmatory research study.* However, social science and psychology suggest that qualitative and quantitative approaches can be used in conjunction to build and refine theory [55, 80, 181]. To this end, we propose a concise list of qualitative approach, quantitative approach and mixed-method approach and a clear mapping of the research procedure in the field of HRI.

Finally, the third goal of this thesis is to develop a HRI research methodology focused on human behaviors focuses, which researchers can easily to follow the procedures in order to acquire knowledges related to human factors, and improve the existing HRI research methodologies and approaches.

In this introduction chapter, we present a brief introduction of this thesis and *motivation* to address HRI in social contexts. We end this chapter by introducing the thesis organization and publications.

1.1 Motivation

- *Why we focused on human behaviors and robot's behaviors in social contexts ?*

Robots currently integrate into our everyday lives, but little is known about how they can behave socially. The idea is drawing on social psychological models of relationships between humans and humans, we look to examples of how people behave in such a social situation and model the robot's behavior off that [88].

However, Graaf et al. would like to postulate that robots themselves are not social, robots can only simulate social behavior or behave in such a manner perceived by human users as social [50]. Conversely, Nitsch and Glassen suggest humans may nevertheless behave differently towards such robots in social contexts, human behaviors on their perceived technological competence and their enthusiasm for technology [121]. The service robots are robust enough to be deployed in industrial contexts, but how these robots behave and interact with humans - act socially - remains largely unclear [15].

For these reasons, we aim to investigate human behaviors and robot's behaviors in social contexts by pushing the current state-of-the art in the field of *human factors* in HRI. We first classify *human factors* in the field of HRI into five different categories: human role, human attribute, human composition, human behavior and proxemics. Then, we discuss the positive and negative behaviors towards robots from three different perspectives: gender effect, age effect and cultural effect.

- *To further investigate human behaviors and robot's behaviors in social contexts, which research topics should be addressed?*

To further investigate human behaviors and robot's behaviors in social contexts, we should bring up discussions related to how we envision robots in the present and future society and ethical issues of HRI in social contexts. In fact, researchers should take a broad view of social behaviors studies and split up the research topics as follows: *technical developments*, *human factors and ethical issues*, *design and evaluation methodologies*, etc. This kind of classification is crucial to the field of HRI because it will ensure that new social robotic prototypes fit our social values and norms. In this thesis, we aim at exploring *human factors and ethical issues*, improving and developing *design and evaluation methodologies* by presenting several empirical studies.

- *To further improve and develop the research topic regarding design and evaluation methodologies, which aspects should be focused on?*

In the last decade, HRI researchers were often robot developers, which give an insight into the robot's functionalities more than social science and psychology. However, lacking of consensus on research paradigms and platforms means that HRI field is not yet in the phase that philosopher Thomas Kuhn would call "normal science" [64]. In this thesis, we aim at pushing the current state-of-the-art in HRI toward research methodologies by introducing (i) *evaluation methodologies and evaluation techniques*, subjective measurements, objective measurements and relevant techniques; (ii) *experimental settings*, participants, experimental environments and robot's autonomy; (iii) *task scenario*, taxonomies and frameworks for HRI scenarios; (iv) *research approaches*, qualitative and quantitative methodologies.

Our goal, in fact, is to improve the existing research methodologies and develop a new HRI research approach, which researches can easily to follow in order to acquire knowledge related to robot's behaviors perceived by human and human behaviors toward robots. To this end, we present our research milestones to reach our thesis goal.

1.2 Contributions

This section presents the main contributions of this thesis and highlights its specific research contributions:

1. Conducting an HRI confirmatory research study in a social context in order to investigate if user perception perceived may be affected by user's gender. Leveraging on Robot Social Attribute Scale (RoSAS) survey and on a statistical analysis, our results show that male users have more expectations in the robot's competences when the interaction becomes more elaborated. Interestingly, results also show the influence by changing the interaction modality with robot on user perception.

2. Developing an approach to design a new type of questionnaire as a task-driven evaluation technique for measuring user perception in social contexts. The approach consists of two steps. First, it relies on interviewing experts in HRI to understand which robots' behaviours can potentially affect the user perceptions during an HRI task. Then it leverages a user survey to filter out those robot's behaviours that are not significantly relevant from the end user perspective. We concretely enacted our approach over a specific scenario. The results of the survey have allowed us to derive a final list of 17 behaviours to be captured in the questionnaire, which has been finally developed relying on a 5-point Likert-scale.

3. Conducting an HRI exploratory research study by introducing an empirical study, it has the main novelty of having been devised and implemented in a realistic social robotic competition environment - the SciRoc challenge, where a representative sample of users were selected from the crowd by SciRoc organization, robots were configured to act autonomously, without the need of any external guidance. Specifically, our empirical study was performed over the most sociable scenario of the SciRoc challenge - "Take the elevator (E4)". We validate the reliability of the questionnaire developed and present the relevant results on *the relationship between the task performances evaluated by scoring and robot's behaviors perceived by participants.*

4. Conducting an HRI confirmatory research study by introducing an empirical study in the context of SciRoc (E4). Leveraging on the questionnaire developed in E4 and on statistical analysis, we found Gentlebots team and UC3M team make a difference on robot's behaviors perceived by users, meaning that task performance is not a factor that can influence robot's behaviors perceived by users. Interestingly, results also show the influences by *users' gender effect* and *users' role effect*: female users perceived the robot's behavior more positively than male users; users' role make a difference on behaviours highly related to spoken languages or dialogues.

5. Introducing a novel concept for designing HRI task scenarios based on a scientific robot competition. In particular, we describe design and implementation of an HRI task scenario including developing an approach for designing task scenarios. The new approach consists of two steps: (i) analyzing the elements of the task scenario; (ii) sketching out the layout of the task scenario.

6. Improving the existing HRI research qualitative and quantitative approaches, proposing a new HRI qualitative and quantitative (i.e., mixed-method [14]) research approach in social contexts. The HRI mixed-method research approach can be adapted by conducting research studies concerning human behaviors toward robots and robot's behaviors perceived by human.

1.3 Thesis organization and Publications

This thesis is organized in three main parts, the preliminaries Part I which introduces the related works concerning our thesis, Part II which describes our research studies on *human factors*, *developing an approach to design a new type of questionnaire*, *Empirical study*, *Designing Task scenario* and *Developing an HRI research approach*, and Part III which summarizes the thesis, opens the new research opportunities in the field of HRI and makes critical discussions.

1.3.1 Part I: Preliminaries

Chapter 2: Background and Related work

First, this chapter presents an overview of HRI, from its *origins* to its *application*, describes HRI issues occured in social contexts. Then, this chapter introduces the current state-of-the-art of the research questions related to our thesis. It presents the current research methodologies, relevant findings in social HRI, as well as discussions and limitations. The chapter presents an overview of social HRI and the research milestones to achieve final goal of our thesis through previous relevant literature.

1.3.2 Part II: HRI studies in a social context

Chapter 3: Human factors: Investigating User Perceptions of HRI in Social Contexts

The aim of this chapter is to present an HRI confirmatory research study in a social context to investigate if user perception perceived may be affected by human factor by introducing concepts of HRI confirmatory research study and HRI exploratory research study. Specially, we focused on three aspects of user perception: *warmth of robot*, *Competence of robot* and *discomfort of robot*. We determined four interactive modalities: *Funny modality*, *Junior modality*, *Senior modality* and *Foreign modality*. In order to collect users' feedback, we conducted a user study by adopting Robot Social Attribute Scale (RoSaS) [33] survey in the range of Maker Faire, due to its proven effectiveness when employed in social contexts. Results show human factor, in particular, users' gender has an influence on many aspects of the user perception. Furthermore, results also show the influence by changing the interaction modality with robot on user perception. A published paper related to this chapter is

- Lun Wang, Andrea Marrella, Daniele Nardi, "Investigating User Perceptions of HRI in Social Contexts", In 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 544-545, 2019 [179].

Chapter 4: Evaluation technique: Developing a Questionnaire to Evaluate Customers' Perception in the Smart City Robotic Challenge

The aim of this chapter is to develop an approach to design a new type of questionnaire as a task-driven evaluation technique for measuring user perception in social contexts. To achieve this object, a crucial step of our approach has consisted of identifying robot's behaviors through dedicated interviews and preliminary surveys performed with both experts and non-experts in the field of HRI, running two demos of the task episode. Based on such results, the concrete development of the questionnaire has been implemented. A published paper related to this chapter is

- Lun Wang, Luca Iocchi, Andrea Marrella, Daniele Nardi, "Developing a Questionnaire to Evaluate Customers' Perception in the Smart City Robotic Challenge", In 2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), pp. 1-6, 2019 [176].

Chapter 5: Experimental Study: HRI Users' studies in the context of the SciRoc Challenge

The aim of this chapter is to devise and implement an experimental study in a realistic social experimental setting context: realistic and dynamic social environment, representative sample of participants selected, robots configured to act autonomously, considerations of ethical issues, implementation of the robots' functionalities and reliability test of the questionnaire. The chapter introduces an HRI exploratory research study implemented through the SciRoc challenge, "take the elevator (E4)". A published paper related to this chapter is

- Lun Wang, Luca Iocchi, Andrea Marrella, Daniele Nardi, "HRI Users' Studies in the Context of the SciRoc Challenge: Some Insights on Gender-Based Differences", In Proceedings of the 8th International Conference on Human-Agent Interaction, pp. 287-289, 2020 [178].

Chapter 6: Task Scenario: Designing and Evaluating tasks based on the SciRoc challenge

The aim of this chapter is to design a task scenario in the context of the SciRoc challenge including developing a new approach for designing HRI task scenarios. The chapter describes the new approach which consists of two steps. First, it relies on analyzing the elements of task scenario based on the existing scenario frameworks. Then, it uses sketches to design the task scenario. Moreover, we show an example of HRI confirmatory research study in the context of the scientific robot competition based on the proposed task scenario. Journal paper related to this chapter is

- Lun Wang, Luca Iocchi, Andrea Marrella, Daniele Nardi, "Designing and Evaluating HRI teaming tasks based on the SciRoc challenge", (Journal publication under review) [177].

Chapter 7: Toward an HRI research approach

The aim of this chapter is to improve the existing HRI research approaches and frameworks and to introduce new concepts. The chapter describes improvements of qualitative approach for exploratory research study and advancements of quantitative approach for confirmatory research study. Moreover, the chapter describes a new HRI mixed-method research approach in social contexts aiming at researching human behaviors toward robots and robot's behaviors perceived by human in the field of social HRI. In the end, the chapter discusses contributions to the current state-of-the-art in the topic of research methodologies and approaches of HRI. A publication related to this chapter is in preparation.

1.3.3 Part III: Conclusion

Chapter 8: Discussion and Conclusion

This chapter summarizes the contributions, achievements and novelties of this thesis, discusses limitations and opens new research opportunities in the field of social HRI.

Part I: Preliminaries

Chapter 2

Background and Related work

In this chapter, we present an overview of HRI, from its *origins* to its *applications*. First, this chapter describes development of HRI as a discipline and relevant applications in HRI. Then, this chapter introduces the current state-of-the-art of the research questions occurred in our thesis.

2.1 Background

2.1.1 Origins of HRI

The concept of *robot* has a long story in different cultural societies, as it appears in narrative stories and fictions. The term of *robot* has been appeared in Karel Capek's play Roussum's Universal robots in 1920. The *Three Laws of Robotics* [10] introduced by Isaac Asimov in 1941 can be considered as the first HRI guidelines for researchers. The Three Laws are:

1. *A robot may not injure a human being or, through inaction, allow a human being to come to harm.*

2. *A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.*

3. *A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.*

At the same time, Isaac Asimov pointed out the importance of human behavior toward robots *if the robots have three laws to protect men, is it too much to ask that men have a law or two to protect robots?* The middle of the twentieth century brought the first explorations of the connection between human intelligence and machines [153]. Around that time, the first robots were realized. The first industrial robot, the Unimate (see Fig. 2.1), was installed at General Motors' Inland Fisher Guide Plant in Ewing Township in 1961. People did consider how they would interact with the robot, but they were more concerned about the place robots would take among human workers [14]. The most commonly cited example of an early autonomous robot was Shakey (see Fig. 2.2) [94], it was developed from approximately 1966 through 1972 by Scientific research institute (SRI). There was almost no interaction between people and computers.

In the 1980s robotics was defined as the science which studies the intelligent connection between perception and action. Behavior-based robots for the first time have been appeared in this period.

Figure 2.1: Unimate developed by George Devol in 1954

Figure 2.2: Shakey developed by SRI in 1966

In [31], Brooks aimed at building cheap robots that can wander around human-inhabited space with no human intervention, advice, or control, there was no specific study concerning HRI for the development of architecture. In the 1990s, robotics researchers were focused on developing products with wide potential markets aimed at improving the quality of life [153]. In a daily environment, robots encounter humans, and the *interaction* among humans, robots and environment becomes essential. Thus, robotics researchers noticed the importance of HRI, e.g., Arkin et al. proposed the new approach of perceptual paradigm including the perceptual classes [9]. Hereafter, robots had various functions with the human-like appearance to interact with human in a social context, e.g., Asimo (see Fig. 2.3), a robot designed to interact socially with humans; Pepper (see Fig. 2.4), a social humanoid robot optimized in HRI; Tiago (see Fig. 2.5), a service robot designed to work in indoor environments with HRI features [131]. More recently, HRI research has been bolstered by the availability of reasonably priced commercial platforms that can be readily purchased by research laboratories. These platforms have expanded both the replicability and comparability of HRI research across labs, as well as the range of people who can engage in the discipline [14].

2.1.2 HRI applications

A successful HRI application requires "highly efficient" interactivity and robot's usability in the specific application domain. Every robot application appears to have some form of interaction. To further address the HRI research questions, we categorize robots into different domains, and discuss

Figure 2.3: Asimo developed by Honda from 2000 through 2018

Figure 2.4: Pepper developed by Aldebaran Robotics (now Softbank Robotics) in 2014

the relevant HRI problems of each application.

Industrial applications

Industrial robots, designed for performing operations quickly, repeatedly and accurately. They have a long heritage in the manufacturing industry, operate in relatively static environments and in large numbers [71]. To achieve an efficient HRI in an industrial domain, *safety* is the primary concern. *A safe interaction must be guaranteed to prevent harming humans having a direct contact with the moving robot* [172]. For instance, Lasota et al. [98] present a real-time safety system capable of allowing safe human-robot interaction at very low distances of separation, without the need for robot hardware modification or replacement. Additionally, to take full advantage skills of human skills, Bannat et al. [13] present a small slice of our concept, constituting a foundation for a multi-modal communication (see Fig. 2.6) between a human and a robot, which is flexible, robust and most appropriate for hybrid assembly. Intuitive interaction is another issue that may play a role in improving these human sensing capacities, this view is widely confirmed by many robot manufacturers [108].

Service applications

Service robots could comprise a single robot or a group of robots working together to achieve a common goal. For service robotics to move from research labs into wide use, development should be done in such a way that they provide a useful service with limited intervention by the user at a reasonable cost [84]. Service robots include tour guide robots, robotic vacuum cleaners, receptionist

Figure 2.5: Tiago developed by PAL robotics in 2016

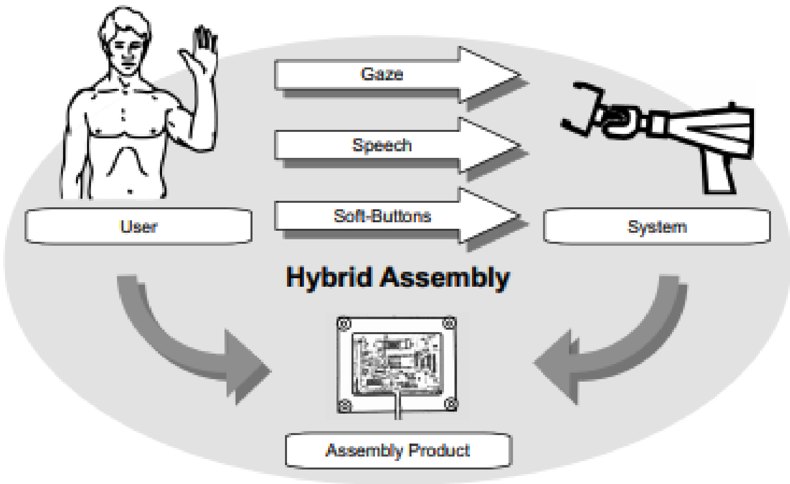


Figure 2.6: Multimodal HRI for industrial robot developed by Bannat et al.

robots, pet and toy robots, robots for healthcare, delivery robots, robots for education, etc. The tasks performed by those robots are repetitive, and they may need to interact with humans. HRI researchers address the situations when such robots operate in social contexts, and robots come into regular contact with humans. Social contexts are almost public environments, such as airports, shopping mall, public office, school, hospital, as well as more private environments, such as home, private location.

In a social context, the physical embodiment affects and impacts HRI. Therefore, embodied condition interaction has received a great deal of attention, demonstrating that knowledge is tied to perceptual, somatosensory, and motoric experience [38]. Furthermore, to seamlessly integrate into the human physical and social environment, robots must display appropriate proxemic interaction, which follow societal norms in establishing their physical and psychological distancing with people [53, 56, 109]. In [144], the authors report HRI studies in a social context, looking at group interaction, interpersonal interaction, and rhythmicity as salient factors that ought to be considered when designing for HRI. More recently, researchers found that *persuasion* and *emotion* interaction are fundamental aspects of how people engage and interact with each other in a social context [68, 73, 146].

2.2 Related work

2.2.1 Human factors

HRI research consists of at least two interrelated components: the human and the robot [14]. The research focus of HRI is on *human factors*, *robotic systems*, *interaction and interactivity*. Designers and engineers developed robot's appearance and capabilities based on human factors (i.e., users' preferences and the users' attitudes). Understanding how people of different ages, gender and abilities best learn from robots remains an important challenge that human factors should contribute to [152]. HRI researchers have classified human factors into four categories: human roles, human attributes, human composition and human behaviors (see Table 2.1). One common methodology of HRI community for human factors research is taking place an experimental study. Setting different experimental conditions, HRI researchers can measure how users react to the different experimental conditions.

Table 2.1: Human factors in HRI

Human factors	Classification
Human role	Supervisor, Operator, Collaborator,
	Cooperator, Bystander [130, 149]
Human attribute	Gender, Age,
	Culture, Education [69, 102, 180]
Human composition	Single, Group [134]
Human behavior	Human behavior toward robots,
	Robot's behavior perceived by human
Proxemics	Far, Close, Touching,
	Supporting, Invasive [175]

In the field of HCI, Nass et al. demonstrate systematically that users attributes such as age, gender and personality type are critical aspects of interactive computer interfaces [116, 117]. Reeves and Nass present the results of numerous psychological studies that lead to the conclusion that people treat computers, TV and new media as real people and places. To further research relevant findings of human factors in HRI, we review the most significant literature accordingly [137].

- *Gender effect* has been studied extensively and demonstrated in several HRI experiments [37, 52, 69, 101, 113, 147, 154, 159, 174], e.g., males show a significant positive attitude toward robots in healthcare with respect to females [95]; In [44], the authors reveal a significant trend between gender and the preferred robot approach direction in a service context.

- *Age effect* between different age groups is also significant; in particular, children and elders are susceptible to the impact of users' perception of HRI [125, 151, 169]. In [69], the authors report that older participants are less willing to use the robot than younger ones in an experiment conducted by Robocare robot; In [156], the authors emphasize the importance to seek mutual gaze and switch addressee often in conversational robot for children.

- *Cultural effect* exists in both positive and negative attitude towards robots [162, 183]. For instance, Li et al. conclude that the cultural background predicts people's positive attitudes towards social robots: people from countries that have a high exposure to industrial robots may have less positive attitudes towards social robots [100]; In [17, 18], the authors report that American users are the less negative towards robots, while Mexicans are the most negative, and Japanese participants do not show a particularly positive attitude towards robots. Furthermore, Lee and Šabanović suggest that culturally variable attitudes and preferences towards robots are not simply reducible to factors such as perceptions and acceptances, rather they relate to more specific social dynamics and norms [99].

To confirm the relevant findings concerning human factors in the field of social HRI, we are interesting in validating the research hypothesis as *User perception of a robot is influenced by user's gender* by conducting a HRI confirmatory research study in a social context. The details of this research study are shown in Chapter 3.

2.2.2 Evaluation techniques

Conducting an experimental study is a common research methodology in the field of HRI. Consequently, the need for developing an appropriate evaluation technique for such studies arises. Evaluation techniques vary in the stage at which they are commonly used and where they can be used. Some are more subjective than others and provide qualitative rather than quantitative measures. Some required more resources in terms of time, equipment and expertise than others [51]. Hence, under the premise of ensuring qualitative data, the evaluation technique should be suitable for experts and participants in terms of ease of use.

In this section, we firstly review the existing evaluation methodologies in the field of social HRI. In [26, 111, 155], the authors review previous literature and categorize evaluation methodologies applied in HRI into four main methods: (i) Self-Assessments or Self-Report Methodologies, (ii) Behavioural measurements, (iii) Psychophysiological measurements, (iv) Task performance metrics. Self-assessment and behavioural measures (i.e., subjective measurements) are the most commonly primary evaluation methodologies applied in HRI studies and research so far [24]. Psychophysiological measurements and task performance (i.e., objective measurements) do not have to rely on participants to report their intended behaviours or preferences. Steinfeld et al. [158] describe task metrics of social HRI as follows: navigation, perception, management, manipulation and social interaction. Tiberio et al. [165] suggest that HRI experimental evaluation method should combine objective and subjective measures.

Questionnaires and observation techniques are often required in experimental evaluation studies with experts or participants as measurement tools in the field of social HRI. Observational studies can rely on data collected in several different ways: notes and logs, video recording and robot logs [14]. The observer effect (also known as the Hawthorne effect) can affect social facilitation or social inhibition [78]. Comparing to observation techniques, questionnaires can be used to reach a wider participant group in a social context, it takes less time to administer, and it can be analyzed more rigorously [51].

We secondly review the exist questionnaires applied in the field of social HRI. The most highly cited questionnaire in the field of HRI is the *Godspeed questionnaire*, covering five aspects as anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety. The results have been distilled into five consistent questionnaires using semantic differential scales [16]. *Negative attitudes* *toward robots scale*, consists of fourteen questionnaire items, measuring humans' attitudes toward communication robots in daily life [123, 127]. *Robotic social attributes scale (RoSAS)*, consists of 18-item scale, is a validated scale to measure social perception of robots [33]. *Technology-specific expectation scale (TSES)* and *Technology-specific satisfaction scale (TSSS)* have been developed by Oliveira et al. to measure users' expectations before seeing and before interacting with a robot, and their satisfaction after the experience of interaction [5]. TSES and TSSS were developed for children to evaluate robot tutors. *Frankenstein Syndorome Questionnaire* is a psychological tool for measuring acceptance of humanoid robots, as well as expectations and anxieties [126]. *Multi dimensional robot attitude scale* investigate 12 dimensions that construct people's attitudes toward domestic robots [120]. Overall, there are very few questionnaires created for social HRI studies that meet validity and standardization criteria [21, 92].

To date, there is no clear method that covers subjective measurements and objective measurements of socially task-driven interaction in a social environment. To tackle this challenge, we realized a simple and repeatable approach to develop a questionnaire that is specifically tailored for contexts where the interaction happens. The details of this study are shown in Chapter 4.

2.2.3 Experimental evaluation settings

Setting an experimental evaluation study for HRI research is critical, because there are many experimental conditions (i.e., factors) that need to be considered. A successful user study in HRI requires to plan and design the experiment carefully [25].

Participants

The choice of participants in an experimental evaluation setting can influence the findings and the success of evaluation user studies. The *number of samples* should be chosen to match the experimental circumstance. Previous HRI users' studies have frequently few participants [21, 90, 113]. For instance, about 44% of user studies published in the proceedings HRI'17 involve fewer than 30 participants. In [96], the authors review the HRI publications of last 10 years, and claim the average number of participants per study was 49. The highest average number of participants is Social Definition at a 79 participant average, which is also the area with the largest number of authors, whereas many of the studies categorize in the Social Definition area are conducted online through Amazon Mechanical Turk, and therefore have a relatively large number of participants. Bethel and Murphy claim that determining the appropriate sample size appears to be a challenge in human studies in HRI [25]. Hence, larger and more physical representative samples are required when conducting experimental studies in social contexts.

User population effect can influence the findings and the success of evaluation user study as well. In [114], Mutz explains that the population-based experiments involve applying complex statistical models to estimate of convenience sample-based experimental treatment effects in order to estimate what they might be in the population as a whole. Samples of participants of previous HRI studies are often 'convenience' samples [19], whilst it is indeed convenient to use students to present a whole population. As we know from previous HRI research literature, human factors such as gender, age, education, etc., can influence users' perception of HRI. Hence, to avoid this problem, we need to emphasize the concept of "diversity" of representative samples when we recruit participants. Belhassein et al. suggest to *widen* the recruitment and randomly recruit people in social circumstance, because recruited participants should represent the target population [21]. There are several recruitment methods available, and they should be implemented for a successful study (e.g., flyers, participant pools, database, etc.) [25].

Environment

For a successful study, the environment should reflect realistically the application domain and the situations that would likely to be encountered so that participants respond in a natural manner [25]. The environment in which an experimental study is conducted can influence the findings and the success of the study. Baxter et al. [19] review the experiments of the published papers in the last three years (i.e., 2013, 2014 and 2015) of HRI conference, the majority of HRI studies have been conducted in laboratories. The laboratory environment is often used for research studies because it is cheaper than real or realistic environment. However, it is not easy to conduct the experimental study in a real environment due to privacy laws or local regulations. Thus, we need to mimic a real environment by setting up a *realistic* environment in social contexts.

Robot's autonomy

Beer et al. [20] categorize the levels of robot autonomy into 10 levels, from manual tele-operation to full autonomy (see Table 2.2). Specially, Tsiakas et. al [167] present a taxonomy regarding level of robot autonomy (see Fig. 2.7) in the field of socially assisted robots. Except for full autonomy level, all the others can be considered as *Wizard-of-Oz (WoZ)* methodologies in experimental studies.

Table 2.2: Level of robot autonomy [20]

Level of robot autonomy
1. Manual Teleoperation
2. Action Support
3. Assisted Teleoperation
4. Batch Processing
5. Decision Support
6. Shared Control with Human Initiative
7. Shared Control with Robot Initiative
8. Supervisory Control
9. Executive Control
10. Full Autonomy

Wizard-of-oZ (WoZ) refers to a person (usually the experimenter, or a confederate) remotely operating a robot, controlling any of a number of things, such as its movement, navigation, speech, gestures, etc [139]. WoZ is particularly suitable in situations in which technology of robot is in developing process. Baxter et al. review the interactive studies of published papers in the last three years (i.e., 2013, 2014 and 2015) of HRI conference, reporting only 40% of interactive studies were performed with autonomous robots, and 60% of interactive studies were performed with Woz methodologies [19]. Weiss claims that a WoZ controlled robot is serving more as a proxy for a

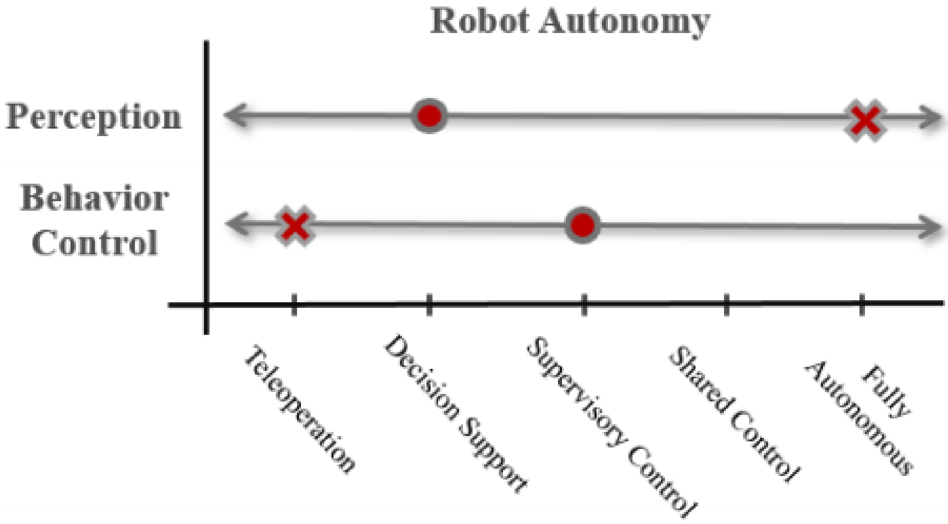


Figure 2.7: Levels of robot autonomy during perception and behavior control [167]
X represents adaptive social behavior generation for assistive robots; O represents supervised autonomy system for robot-enhanced therapy

human and less as an independent entity. Thus, WoZ is not really human-robot interaction so much as human-human interaction via a robot [182]. For this reason, we encourage HRI researchers to conduct experimental studies in social contexts through robots that act autonomously.

Although HRI research involving real-world autonomy is still limited in social contexts [38], there has been an increased interest over the past decade in developing and studying autonomous systems in the wild. To fulfill these requirements of participants, environments and robot's autonomous system for an experimental setting, we proposed the Smart City Robotics Challenge (the SciRoc challenge), a robotic competition that offered the possibility to concretely enact experimental studies to investigate users' attitudes in a specific HRI scenario involving *full autonomy robots*. The considered HRI task performed in the SciRoc challenge has the following features: i) realistic and dynamic social environment; ii) representative sample of users selected form the crowd by SciRoc organization; iii) robots configured to act autonomously, without the need of any external guidance. The details of this experimental study are shown in Chapter 5.

2.2.4 Task scenario

A scenario involves individuals, objects, and events referring to a situation or more precisely, an episode [187]. In a social context, a scenario refers to the activities performed by a person or a group of person in relation to a social robot or a group of social robots. HRI scenarios have been widely used for evaluating studies [188], as HRI scenarios could be used to help convey meaningful context to the interactions.

As we know, methods and techniques from the field of HCI could contribute to and learn from recent developments in the area of HRI [54], in particular, HCI methods that deal with scenarios provide a rich representation of activities from which cognitive and organizational perspectives can be developed [34]. First of all, we review the frameworks applied to HCI scenarios, e.g., Rosson and Carroll [143] present a framework for HCI scenarios (see Fig. 2.8), which analyze seven elements including setting, actors, task goals, plans, evaluation, actions and events. Benyon and Macaulay

[23] provide a more structured approach comparing to Carroll's scenario-based design.

Scenario Element	Definition	Examples
Setting	Situational details that motivate or explain goals, actions, and reactions of the actor(s)	Office within an accounting organization; state of work area, tools, etc., at start of narrative
Actors	Human(s) interacting with the computer or other setting elements; personal characteristics relevant to scenario	Accountant using a spreadsheet package for the first time
Task goals	Effects on the situation that motivate actions carried out by actors(s)	Need to compare budget data with values questioned in memo
Plans	Mental activity directed at converting a goal into a behavior	Opening the memo document will give access to memo information; resizing one window will make room for another
Evaluation	Mental activity directed at interpreting features of the situation	A window that is too large can be hiding the window underneath; dark borders indicate a window is active
Actions	Observable behavior	Opening memo document; resizing and repositioning windows
Events	External actions or reactions produced by the computer or other features of the setting; some of these may be hidden to the actor(s) but important to scenario.	Window selection feedback; auditory or haptic feedback from keyboard or mouse; updated appearance of windows

Figure 2.8: HCI scenarios defined by Rosson and Carroll [143]

Secondly, we review the frameworks that are applicable to various HRI scenarios. For instance, Onnasch and Roesler [130] present the HRI taxonomy framework which provides predefined categories to enable structured comparisons of different HRI scenario. The taxonomy framework includes three category clusters: Interaction context (dark grey), robot (medium grey) and team classification (light grey) (see Fig. 2.9). The interaction context classification describes the first layer of the hierarchical structure of the HRI and aims to explicate the specific domain context. The robot classification focuses on the robot's work context and design with the three variables of robot task specification, degree of robot autonomy, and robot morphology. The team classification to HRI scenarios characterizes the structure of interaction, aspects of composition and teamwork that are addressed with four variables: human role, team composition, communication channel and proximity (physical and temporal). In [7], Andre et al. present the *Environment Norms Autonomy Composition Task Embodiment Duration (ENACTED)* framework providing an organised structure around the definition of scenarios for human-agent and/or human-robot group interaction. Moreover, Agrigoroaie et al. [2] design a scenario diagram for experimental studies. Robins et al. [141] present a general methodological approach for scenario development in HRI research in the context of Robot-Children interaction (RCI). Zlotowski et al. [191] develop HRI scenarios based on human-

human studies as cost effective method to be used in the early phases of robot development. Malik et al. [103] develop on interactive scenarios concerning the suitability of measuring items in the Gross Motor Function Measure (GMFM) that is suitable to be applied to humanoid robot NAO.

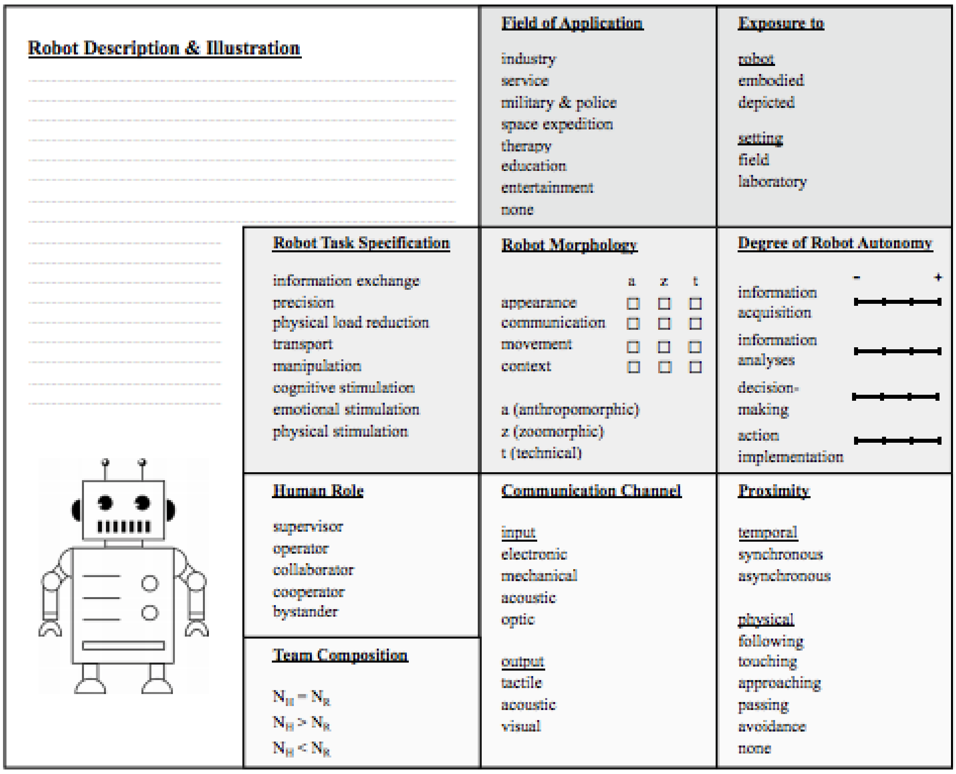


Figure 2.9: Taxonomy for HRI scenarios defined by Onnasch and Roesler [130]

Building upon methods and frameworks proposed, we develop a new approach for designing a repeatable task scenario in a robotic competition. To be more specific, we designed our task scenario, the most sociable episode of the SciRoc challenge "Take the elevator - E4", in which a robot must take an elevator of the shopping mall with customers to reach a service located in another floor. An approach to scenario design consists of two phases: (i) analyze concretely the elements of the HRI task scenario; (ii) sketch out the layout of the HRI task scenario. The details of this development are shown in Chapter 6.

2.2.5 Research methodology and research approach

The growing need for HRI research brings credibility and validity to scientific research, HRI researchers learn from relevant research methods and develop new research approach applicable to HRI.

A qualitative method allows researchers to understand qualitative of an interaction that are difficult to capture in numbers. It requires researchers to identify and interpret the underlying meaning or thematic patterns that they see in the social interaction. A qualitative method focuses on exploring ideas and formulating a theory in the context of social science and psychology, requires few participants, collects qualitative data through interview, focus groups, ethnography and literature review. Qualitative data is non-numerical meaning that it is difficult to analyze, it consists of texts, images and videos. Researchers analyze qualitative data by adopting content analysis, thematic analysis, text analysis and discourse analysis. For instance, Bickman and Rog [27] design an interactive model of qualitative methods (see Fig. 2.10), the model of research design have five components: goals of the experimental study, conceptual frameworks concerning literature and preliminary studies, research questions, methods for the experimental study and validity concerning the results and conclusions. However, the research studies on developing qualitative research methods for HRI are still underestimated.

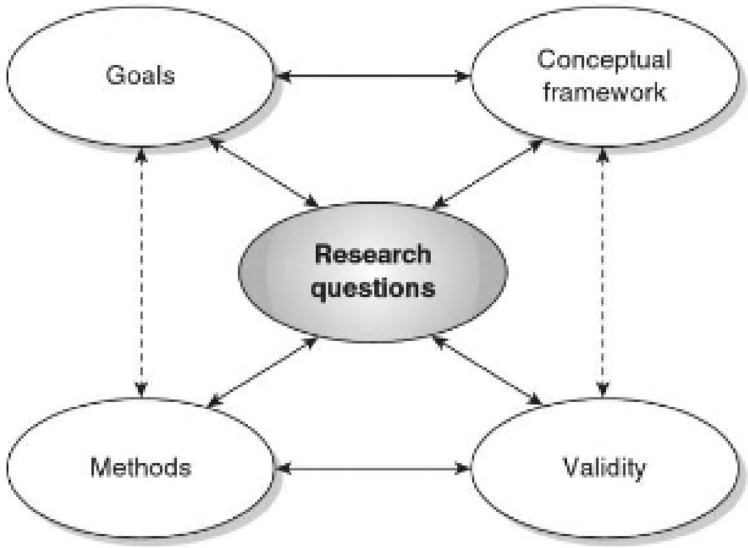


Figure 2.10: Social qualitative research approach developed by Bickman and Rog [27]

Quantitative methods, in contrast, often take the shape of surveys or controlled experiments and produce data that can be expressed numerically and analyzed statistically [14]. A quantitative method focuses on confirming a theory or hypothesis in the context of social science and psychology, requires large scale participants, collects quantitative data through surveys, observational cameras, etc. HRI researchers might arise from theoretical considerations, such as the expectation that people will treat robots as social or form the pragmatic need to test the usability of a certain robot feature or function [14]. Formulating a hypothesis is proposing a prediction of the possible outcome of an experiment. The aim of the experiment is to show that this prediction is correct. It is framed in terms of the independent and dependent variables, stating that a variation in the independent variable will cause a difference in the dependent variables [51]. Hence, along with formulating the research hypothesis, HRI research should establish correlation or causation between the variables. The establishment between correlation and causation can impact the adoption of data analysis method as well. Quantitative data is numerical meaning that it can be analyzed by statistical or mathematical methods. In this section, we list some of data analysis techniques often used in quantitative approach [40]:

- Analysis of variance (ANOVA): A parametric statistical test for determining whether the means of two or more groups on a single dependent variable differ significantly from each other by chance.

- Multivariate analysis of variance (MANOVA): A parametric statistical test used to determine whether the means of two or more groups on two or more dependent variables differ significantly from each other by chance.

- Maximum likelihood method: A method for finding estimates of the population parameters of a model which are most likely to give rise to the pattern of observations in the sample data.
- Parametric test: A statistic test based on the assumption that the population from which the samples are drawn is normally distributed.
- Correlation: An index of the strength and direction of the linear association between a quantitative criterion variable and a quantitative predictor variable controlling for the linear association between the predictor and one or more other quantitative predictor variables.
- Post-hoc test: A test for determining whether two groups differ when there are no strong grounds for expecting they will.

For instance, Hoffman and Zhao [72] develop an HRI quantitative approach (see Fig. 2.11). It begins with clarifying research questions, constructs, and hypotheses (Sections 2 and 3) and then moves to study design, which includes choosing variables and measures (Section 4), planning the study procedure (Section 5), and sampling participants (Section 6). It then goes on to cover data collection (Section 7), some commonly used experimental statistical methods and their appropriateness in various contexts (Section 8), as well as recommendations for reporting and discussing results (Section 9).

However, measuring behaviors in the field of HRI is not easy: not only one has to deal with a different, embodied technology compared to computers, but also their application poses different technical and social challenges for research. It is important to be precise about the methodological approaches used in HRI studies, but at the same time HRI research needs to be aware that there is no ‘once-and-for-all’ solution applicable across HRI [42]. For instance, Broadbent et al. [30] describe an HRI research approach, which aims to develop and test healthcare robots for older people. The research approach which consists a series of steps: literature review, identifying needs, implementing robots, evaluation studies by adopting qualitative approaches or quantitative approaches, measurement of outcomes, risks and benefits, providing good quality of life, usability studies.

Hence, we summarize some differences, advantages and disadvantages between qualitative approach and quantitative approach in the field of HRI as follows:

- Qualitative approach is used for exploratory research study *vs* Quantitative approach is used for confirmatory research study
- Qualitative approach is subjective *vs* Quantitative approach is objective
- Qualitative approach requires few participants *vs* Quantitative approach requires large scale samples
- Qualitative approach collects data through interview, focus groups, ethnography and literature review *vs* Quantitative approach collects data through surveys, cameras, etc.
- Qualitative approach analyzes data through content analysis, thematic analysis, text analysis and discourse analysis *vs* Quantitative approach analyzes data through statistical and mathematical methods.

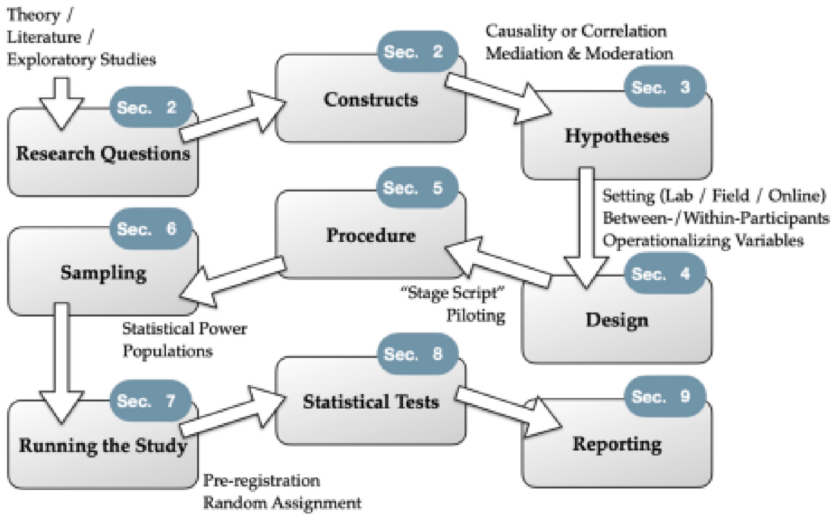


Figure 2.11: HRI research framework developed by Hoffman and Zhao [72]

- Qualitative approach is easy to implement *vs* Quantitative approach is difficult to implement, specially, in the context of social HRI.

To improve the existing research approaches and methodologies in the field of HRI, we propose a new HRI qualitative and quantitative (i.e., mixed-method) research approach aiming at contributing more reliable research results. The details of this development are shown in Chapter 7.

Part II: HRI studies in a social context

Chapter 3

Human factors: Investigating User Perceptions of HRI in Social Contexts

3.1 Introduction

While analyzing the present and future roles of HRI in social contexts, we shall be aware of the underlying principles and the influences by human factors. HRI researchers suggest that, from the perspective of the user, interaction with an artificial entity is similar to interaction with fellow humans. Explanations for this treatment of robots in a social way assume that due to our social nature, humans will use their interaction routines also when confronted with artificial entities [93]. In particular, there is considerable interpersonal variation factor with respect to whether or not artificial communication partners are treated as social actors [57]. To explore the influences of human behaviors by human factors in HRI, we consider previous users studies which aim to discover and characterize possibly influencing factors of human attitude toward robots. For instance, Vanzo et al. [171] were interested in investigating influence of human collaborative attitudes toward robots by user's gender, user's height and proxemics (i.e., human factor); the authors have conducted a quantitative HRI confirmatory research study. De Graaf et al. have conducted a qualitative HRI exploratory research study and the authors were interested in exploring the relation between an interaction with a robot and peoples' attitudes and emotion towards robots and such influences by user's gender and user's nationality (i.e., human factors) [49]. By analyzing the literature, HRI researchers usually investigate human factors by conducting two kinds of research study:

- **Exploration research study**: It *explores* relevant findings without expectation. It can be considered as a qualitative research methodology.
- **Confirmatory research study**: It *validates* a research hypothesis. It can be considered as a quantitative research methodology, as it allows researchers to confirm hypotheses through data analysis obtained by experiments.

The findings obtained by previous HRI studies concerning human factors are shown in Section 2.2.1. In this chapter, we are interested in validating the hypothesis on *gender effect* as primary goal. Since interaction modality has been demonstrated as an influence factor in HRI [160], we are interesting in validating the hypothesis on *interaction modality effect* as the secondary goal. To sum up, we propose an *HRI confirmatory research study* by conducting an empirical experiment in a real context to validate the hypotheses on *gender effect* and *interaction modality effect*.

3.2 Methods and Materials

3.2.1 Defining research question

To form a research question, we extend the relevant findings in the field of social HRI concerning *gender effect* and *interaction modality*, and propose the hypotheses as follows:

- Hypothesis 1 (H1): User perception of a robot is influenced by user's gender in a social context.
- Hypothesis 2 (H2): User perception of a robot is influenced by changing the interaction modality with the robot.

3.2.2 Determining *dependent factors* and *independent factors*

We design our research study by first carefully selecting the *dependent factors* to be measured and the *independent factors* to manipulate for producing many conditions for comparison.

New data confirm two universal dimensions of social cognition: warmth and competence. Promoting survival, these dimensions provide fundamental social structural answers about competition and status [58]. Moreover, to facilitate interactions between human and robot, we divided robot's behaviors perceived by human into two categories: comfort behaviours of robot and discomfort behaviors of robot. Hence, user perception on robot's behaviors can be represented by three aspects: *warmth of robot*, *competence of robot* and *discomfort of robot*. We identified 18 dependent factors to be considered, reflecting the user perceptions that can be captured during an interaction with a robot. According to RoSAS [33], any dependent factor can be associated with one of the three aspects of user perception. Specifically, factors *happy, feeling, social, organic, compassionate, emotional* with warmth of robot; factors *capable, responsive, interactive, reliable, competent, knowledge* with competence of robot; factors *scary, strange, awkward, dangerous, awful, aggressive* with discomfort of robot.

Since the value of dependent factors is "dependent" on changes made to the independent factors, we identified two of such factors: *gender* and *interaction modality*, in particular, gender assumes two values *male* and *female*. Concerning the interaction modality, we identified four different values for each factor:

- Funny modality (F): Robot tells a funny joke to the user in Italian language;
- Junior modality (J): Robot asks a short and easy question in Italian language, and the user selects one of the available answers on the tablet. Finally, the robot provides a comment on the user's answer;
- Senior modality (S): Robot asks a long and non-trivial question in Italian language, and the user selects one of the available answers on the tablet. Finally, the robot provides a comment on the user's answer.
- Foreign modality (E): Like the Junior modality, but questions and feedbacks are provided in English.

All the above interaction modalities are enacted by the robot with the aid of gestures, head-pose, gaze pattern, images shown on the tablet and voice (see Fig. 3.1).

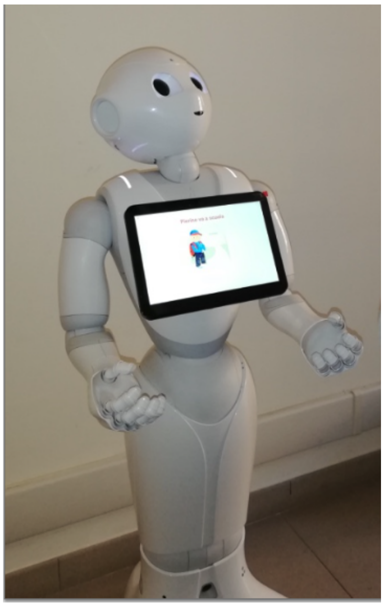


Figure 3.1: Funny modality interaction implemented with Pepper

3.2.3 Selecting robots

Pepper is a robot ([1]), developed by SoftBank Robotics, created with the goal of achieving these interaction modalities.

Pepper is an industrially produced humanoid robot launched in June 2014 that was first created for B2B needs and later adapted for B2C purposes. The robot is capable of exhibiting body language, perceiving and interacting with its surroundings, and moving around. It can also analyze people's expressions and voice tones, using the latest advances and proprietary algorithms in voice and emotion recognition to spark interactions. The robot is equipped with features and high-level interfaces (i.e., tablet, etc.) for multimodal communication with the humans around it [132]. Moreover, Pepper robot becomes a well recognised example of social robotics moving into public social contexts [1, 118]. As we expect to study the interaction between one robot and people, we decide to conduct our experimental study by employing one Pepper robot.

3.2.4 Implementing interaction modalities

The interaction with the robot happened in a face-to-face fashion. Users select one of the four available interaction modalities, and the robot activated the routine associated to selected modality. The diagram of the four available interaction modalities is shown in Fig 3.2. The duration of each single interaction is around 2 minutes. Users are free to complete an interaction or leave it anytime. When a single interaction expired, the users could leave the "demo area" or start a new interaction with the robot. Since we conduct our experimental study in a public social context, tactile user interaction (through the tablet) is preferred over speech due to the noise occurred.

[1] https://www.softbankrobotics.com/emea/en/pepper

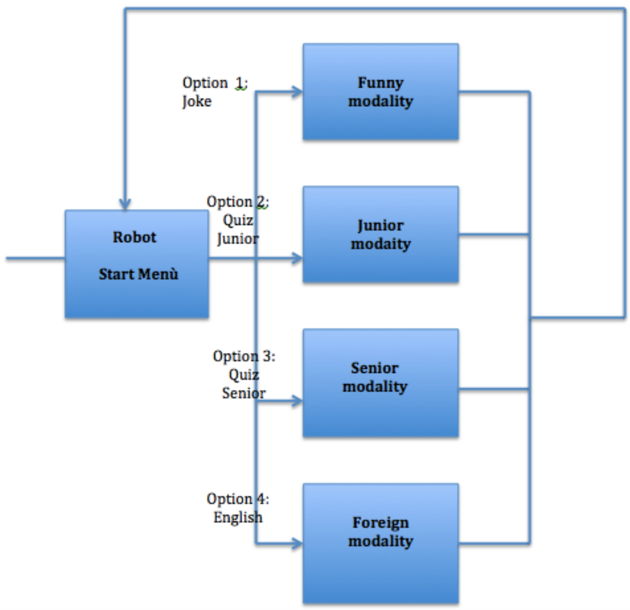


Figure 3.2: Diagram of the four interaction modalities

3.2.5 Selecting experimental contexts

We performed our user study in the range of Maker Faire ([2]), an event created by Make magazine to "celebrate arts, crafts, engineering, science projects and the Do-It-Yourself (DIY) mindset". Maker Faire events have been held in Europe, America, Asia and Africa since 2006. The Maker Faire event where we conducted the experiment was held in Rome from 12th October to 14th October, 2018 (see Fig. 3.3). The Maker Faire event in Rome aims at showcasing recent innovative technological solutions. In order to increase the attention of the participants, we placed the robot in the corridor of the venue.

Figure 3.3: Maker Faire event 2018 in Rome

3.2.6 Selecting participants

In social contexts, users are male or female, and usually have a broad range of age, nationality, cultural background, etc. We employed the robot in a social context, and expect the natural and spontaneous interaction between the robot and people. Therefore, any user participating to the Maker Faire event 2018 in Rome was representative for our experimental study. We use the term *subject* or *participant* to refer to a person taking part in an experiment [170].

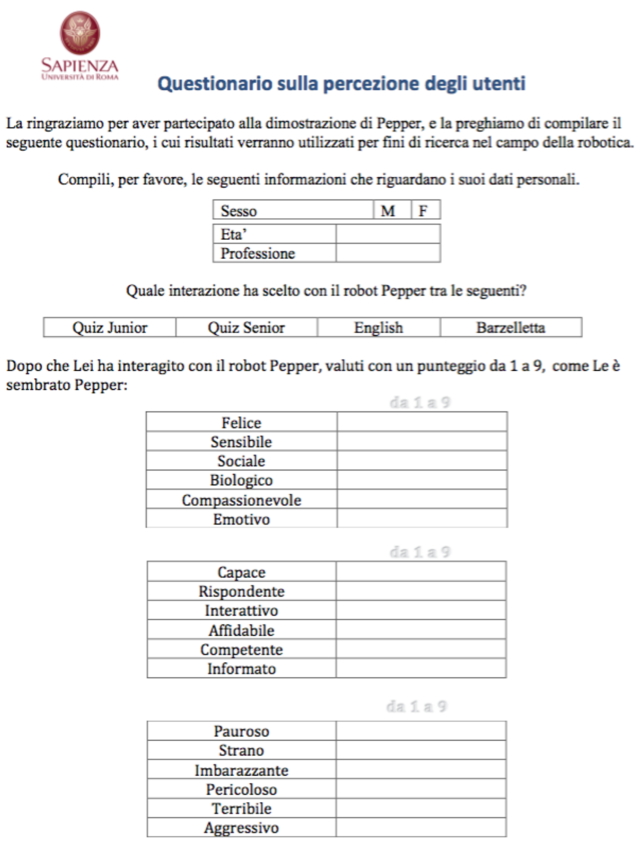
[2] https://makerfairerome.eu/it/edizioni/2018-it/

3.2.7 Selecting experimental designs

We perform our experimental study using the *between-subject* design, i.e., each user was assigned randomly to a different experimental condition, consisting of a single interaction or multiple interactions with the robot. Between-subject designs are typically cleaner because participants are exposed to only one experimental condition and typically do not experience practice effects or learn from other task condition [25]. Data analyses of between-subject design are statistically simple to perform as long as random assignment is achieved across groups [36]. Hence, we adopt t-test and ANOVA for data analysis.

3.2.8 Selecting evaluation techniques

We focus on investigating three aspects of user perception: *warmth of robot*, *competence of robot* and *discomfort of robot*. In order to collect users' feedback, we conduct a user study by adopting Robot Social Attribute Scale (RoSAS) survey, due to its proven effectiveness when employed in social contexts. We translate RoSAS survey in Italian language (see Fig. 3.4). Once a user completed a (single or multiple) interaction with the robot, we collect her/his feedback through the RoSAS survey. Users have to fill general information such as age, gender, profession. Then they provide a score from 1 to 9 to any of the 18 dependent factors to be measured.


SAPIENZA UNIVERSITÀ DI ROMA

Questionario sulla percezione degli utenti

La ringraziamo per aver partecipato alla dimostrazione di Pepper, e la preghiamo di compilare il seguente questionario, i cui risultati verranno utilizzati per fini di ricerca nel campo della robotica.

Compili, per favore, le seguenti informazioni che riguardano i suoi dati personali.

Sesso	M	F
Eta'		
Professione		

Quale interazione ha scelto con il robot Pepper tra le seguenti?

Quiz Junior	Quiz Senior	English	Barzelletta

Dopo che Lei ha interagito con il robot Pepper, valuti con un punteggio da 1 a 9, come Le è sembrato Pepper:

	da 1 a 9
Felice	
Sensibile	
Sociale	
Biologico	
Compassionevole	
Emotivo	

	da 1 a 9
Capace	
Rispondente	
Interattivo	
Affidabile	
Competente	
Informato	

	da 1 a 9
Pauroso	
Strano	
Imbarazzante	
Pericoloso	
Terribile	
Aggressivo	

Figure 3.4: RoSAS survey in Italian language

3.3 Data Analysis and Results

3.3.1 Data collection

145 users participated to the experimental study spontaneously. We collected 125 valid answers to the survey. Since only one user has selected the "foreign" interaction modality, we decided to exclude such modality from the analysis and focus on the other three modalities. Gender distribution was as follow: 44% of male users, 53,6% of female users and 2,4% of users that did not declare their gender (see Fig. 3.5). Users have declared their interaction modalities. Distribution of users in single and multi interaction modalities are shown in 3.6. The first four elements of the distribution table concern single interaction modality, meaning each user have selected only one interaction modality. The six elements in the middle of the distribution table concern plural interaction modality, meaning each user have selected two interaction modalities. The five elements at the bottom of the distribution table concern multi interaction modality, meaning each user have selected three or four interaction modalities.

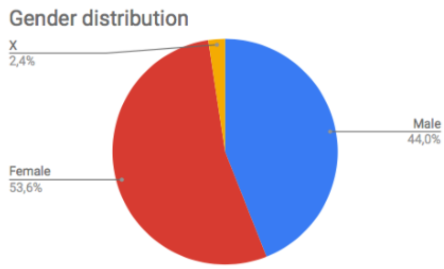


Figure 3.5: Distribution of user's gender

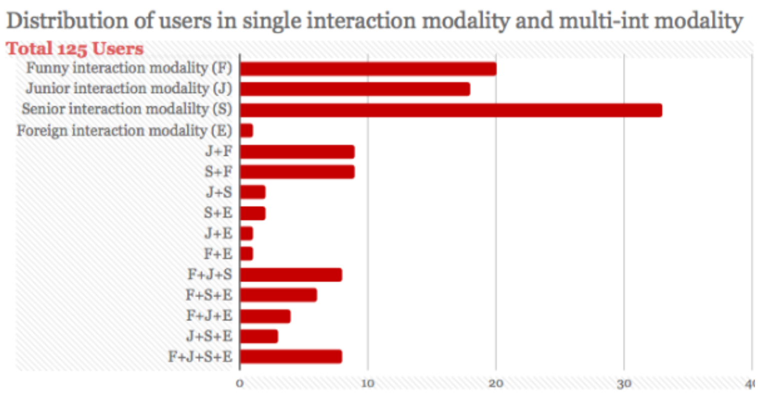


Figure 3.6: Distribution of users in single and multi-interaction modalities

3.3.2 Analyzing impact of gender on user perception

We analyzed gender information of each user that performed exclusively a single interaction with the robot. Then, we used t-test, a type of inferential statistic used to determine if there is a significant difference between the means of two groups, to check how male and female users perceived the interaction with the robot in the different interaction modalities. Significant results (i.e., $p<0.05$) have been found in any of the interaction modalities. For example, female users found the robot more *biologic*, *compassionate* and *capable* during the "Funny" interaction modality ($p_{organic}=0.039$, P

$_{compassionate}$ =0.0228, p $_{capable}$ =0.035), and more *responsive*, *interactive*, *reliable*, *competent*, and *knowledgeable* (all aspects related to the competence perceived of the robot) in the "Senior" interaction modality (p $_{responsive}$ =0.033, p $_{interactive}$ =0.0309, p $_{reliable}$ =0.0129, p $_{competent}$ =0.0125, p $_{knowledgeable}$ =0.0354). It is worth to notice that, even when the results are not statistically significant, the mean of the answers provided by female users is always higher than for male users, whose answers have often a larger variance.

We conclude that male users have more expectations in the competence of the robot when the interaction becomes more elaborated (i.e., during the "Senior" interaction modality). This confirms our first experimental hypothesis.

3.3.3 Analyzing impact of different interaction modalities on user perception

First of all, we analyzed only users that performed exclusively a single interaction with the robot. For any of the 18 dependent factors (e.g., "biologic", etc.), we used ANOVA to check if there were statistically significant differences in the obtained scores related to the factor by varying the interaction modality (i.e., "Junior" vs "Funny" vs "Senior"). Only in one case – "scary" – we found a significant difference ($p<0.05$) between the collected scores (p $_{scary}$ =0.0193). To precisely identify the source of the difference, we used t-test to compare the scores obtained for the "scary" factor evaluating interaction modalities in pairs. In this way, we identified that users perceived a greater feeling of discomfort (in particular, of scary) when the interaction with robot was more elaborated, such as in the "Senior" and "Funny" modalities, while this discomfort disappeared when the interaction happened exclusively in the "Junior" modality.

Secondly, we checked with ANOVA if performing multiple interactions (that always included the "Senior" and "Funny" modalities) could reduce the feeling of discomfort behaviour got during single interactions, but no statistical evidence has been captured. Hence, this allows us to conclude that our second experimental hypothesis is satisfied only for one dependent factor (i.e., *scary*) during the "Senior" and "Funny" interaction modalities.

3.4 Discussion and Conclusion

The aim of this study is to discover gender differences in real social context interacting with an autonomous robot.

We present an HRI confirmatory research study in the real social context of Maker Faire 2018 in Rome and validate the relevant results concerning *gender effect* and *interaction modality effect*. Leveraging on Robot Social Attribute Scale (RoSAS) survey and on a statistical analysis, our results show male users will be more likely to accept robot in a public social environment if they know that the competence of the robot was elaborated, in contrast, female users will be more likely to accept robot in a public social environment when they perceived "less" complicated of the competence of the robot. Hence, the findings can benefit a more precise interaction and design model with which it could fit individual needs in social contexts.

Moreover, our results show performing multiple interactions could reduce the feeling of discomfort got during single interactions. This finding can benefit social robot researchers to deal with users' negative feelings toward robots in real social contexts. Avoiding fears and anxiety toward robot is a specific theme that should be addressed in near future.

This study confirms the importance of *human factors* in the field of social HRI. Conducting experimental studies in researching *human factors* can be challenging. Careful planning and design can make the experience more positive and successful [25]. Through the use of an appropriate evaluation technique, research findings obtained can provide more validity and credibility. In Chapter 4, we address the specific problem on developing an appropriate evaluation technique in social contexts.

Chapter 4

Evaluation technique: Developing a Questionnaire to evaluate customers' perception in the SciRoc challenge

4.1 Introduction

In the field of HRI, *query techniques* are used to collect details of the user's perception about the interaction with a robot. The advantage of such methods is that they get the user's viewpoint directly and may reveal issues that have not been considered by the designer [51]. Two main types of query techniques are often used in HRI experiments: *interviews* and *questionnaires*. Interviews are particularly effective for collecting information about user preferences, impressions and attitudes. On the negative side, they are time consuming. Consequently, they are not suitable as evaluation techniques in social contexts because they usually reach a limited number of users. In contrast, questionnaires can be used to reach a wider participant group as they take less time to administer, and the answers can be analyzed more rigorously. Since the evaluator/researcher is not directly involved in the completion of the questionnaire, it is crucial that it is well designed. Developing a valid questionnaire can take a considerable amount of time and the absence of standardisation makes it difficult to compare the results with other studies [148].

To date, with the context of existing questionnaires employed in HRI (see in Section 2.2.2), there is no clear method that covers the breadth and depth of the social task-driven interaction experience with robots.

With the purpose of testing and evaluating human behaviours toward robot and robot's behavior perceived by human in the field of HRI, we realise a simple and repeatable approach to the development of a questionnaire that is specifically tailored for the task-driven context. To achieve this objective, we focus on the most social episodes of the SciRoc challenge - "Take the elevator (E4)". E4 is the collaborative task which the robot must take an elevator crowded with customers to reach a service located in another floor. In E4, the robot is able to enter/exit the elevator at the right floor in the presence of people nearby and /or inside. To perform the task, the robot can interact with the customers in spoken language. The robot is not supposed to push buttons, and it can ask the people around to do it (see Fig. 4.1). According to the rules for E4, the robot will encounter two persons while moving towards the elevator. Once arrived in front of the elevator, the robot will encounter one to three persons. All such people will take the elevator with the robot.

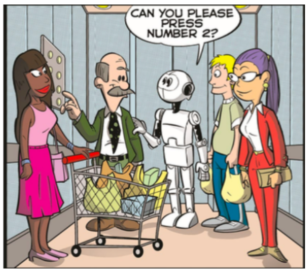


Figure 4.1: E4: Take the elevator

To develop an effective questionnaire that is able to properly capture users' perceptions after their interaction with the robot in E4, it is necessary to understand which robot's behaviors may positively or negatively affect users perception. Since there may be discrepancies between subjective assessment and objective assessment, a number of studies of HRI have combined both measurements [145], and our questionnaire follows the same principles. To this end, a crucial step of our approach has consisted of identifying such behaviors through dedicated interviews and preliminary surveys performed with both experts and non-experts in the HRI field, running two demos of the task. In particular, we leverage the scenario presented in E4 to develop a *new type of questionnaire* that focuses on three perspectives, which describe interaction experience with robots in social contexts: *Social behavior of robot*, *Proxemics between human and robot*, *Collaboration behavior of robot*.

4.2 Approach

4.2.1 Overview of the Approach

The proposed approach to the questionnaire development is based on two steps, as shown in 4.2.

Step 1

First, we implemented and recorded two complete demos of E4 in the corridor of our department. Then, we relied on dedicated interviews with researchers in robotics to understand which robot's behaviours can potentially affect the users' perceptions during a task of HRI in a social environment. The results of the first step have allowed us to identify the preliminary robot's behaviors.

Step 2

Second, leveraging a user survey performed with non-experts in HRI, we validated which of the previously selected behaviours are concretely relevant from the user perspective, filtering out those ones considered as secondary. The results of this second step have allowed us to derive the final list of behaviours to capture in the questionnaire.

4.2.2 Implementing of two demos for E4

To capture the potential interesting behaviors of the robot during E4, we implemented, executed and recorded (through a video camera) two complete demos of E4 employing two different types of robot. We mimicked the scenario of E4 and performed the task in the corridor of our department.

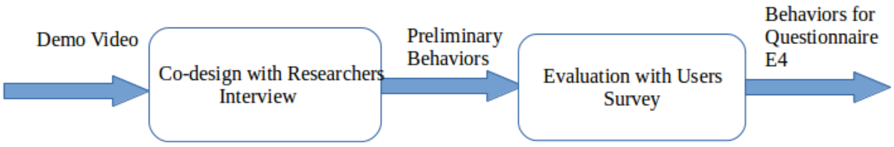


Figure 4.2: Overview of the Approach

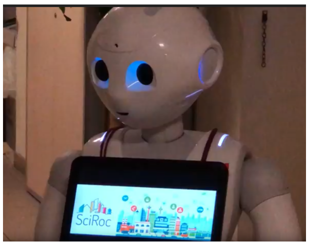


Figure 4.3: Pepper, Social robot selected to perform the two demos

For the first demo, we chose Pepper (see Fig. 4.3) as social robot, which can interact with users through spoken language or, alternatively, with a tablet attached to the robot. The tablet displays images and allows for tactile interaction.

For the second demo, we chose MARRtino ([1]) as social robot (see Fig. 4.4). MARRtino is a ROS-based low-cost differential drive robot platform that can be assembled in many shapes. We assembled MARRino with a tablet, a microphone and a loudspeaker. The tablet displays images, detects faces, and communicates with users.

We performed both demos using Wizard-of-Oz method [139]. The video recordings of the demos are available online, at: `https://bit.ly/2I8zBxU` and `https://bit.ly/2VhGvbT`. It is worth to notice that our decision to employ two robots having different shapes (Pepper with human-like features, MARRtino with a machine-like shape) has been targeted to avoid that the shape of the robot could influence the spectrum of behaviours perceived by the users during a HRI task.

4.2.3 Determining the preliminary robot's behaviors through interviews

Recording demos of E4 was useful to show the concrete working of the episode to several users. In this direction, we realized dedicated interviews with five researchers of the robotic laboratory at Sapienza University of Rome, showing them the video demos of E4. The interviews were targeted to investigate which of the behaviours covered by the existing query methods in HRI are able to potentially affect the users' perceptions during E4. To this end, we grouped such behaviours according to the three perspectives discussed at the end of introduction, and we expanded and analyzed them one by one (see Fig. 4.5).

- *Social behavior of robot*. For this perspective, we relied on RoSAS, which was developed exclusively to measure the social behavior of a robot in an HRI task. RoSAS considers the following behaviors:
  - *Warmth of robot*, Happy, Feeling, Social, Organic Compassionate, Emotional.

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

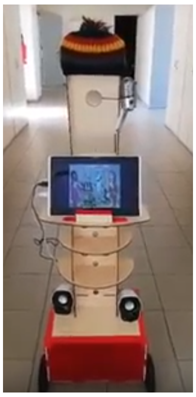


Figure 4.4: MARRtino, Social robot selected to perform the two demos

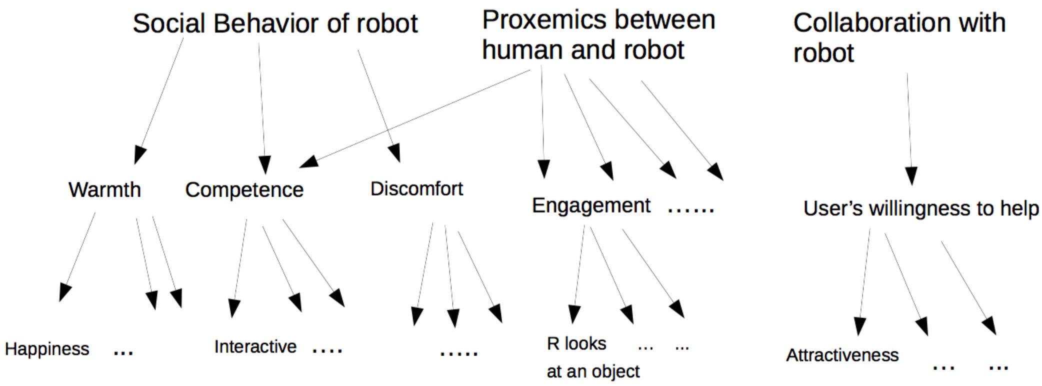


Figure 4.5: Analyzing the preliminary robot's behavior through interview

- *Competence of robot*, Capable, Responsive, Interactive, Reliable, Competent, Knowledgeable.
- *Discomfort of robot*, Scary, Strange, Awkward, Dangerous, Awful, Aggressive.

The researches involved in the interview evaluated all the behaviors covered by RoSAS as potentially relevant (except Organic, which was filtered out) for measuring users' perception in E4.

- *Proxemics between human and robot.* Proxemics is the study of how a robot uses the available space considering the physical and psychological distancing from humans [109]. People might perceive robots that do not show appropriate distance as threatening to their social environments and work practices [133]. On the other hand, carefully designed proxemics behaviors in robots might foster closer relationships with humans, enabling widespread acceptance of robots. While this issue is well investigated in the literature (e.g., see [164, 171]), in E4 we decided to analyze proxemics through customers' perceptions. In this direction, the researchers involved in the interview considered the following behaviours, covered by [16] and [105], as potentially relevant in E4, except the *competence of robot*, which was already discussed in the previous perspective:
  - *Engagement*, robot looks at an object, robot looks at user's eyes, robot looks at user's

face, users look at robot's face, users look at robot's eyes, users pay attention to the conversation with robot, users understand well the meaning of the conversation. These behaviors can be considered as *objective assessments*.

  - *Anthropomorphism of robot*, natural, human-like, conscious, lifelike, moving elegantly.
  - *Likeability or robot*, like, friendly, kind, pleasant, nice. Moreover, researchers suggested the insertion of an additional behaviour: politeness of the robot.
  - *Technology adoption by customers*, adaptability, ease of use, usefulness, trust.

- *Collaboration with robot*, on the basis of previous studies on collaborative attitudes of humans towards robots (see [12, 171, 184]), the following behaviours were considered as potentially relevant in E4 by the researchers involved in the interview:

  - *Users' willingness to help*, attractiveness, enjoyment, endearment, symbiotic relationship, reciprocity relationship, collaborativeness.

  Conversely, researchers decided to exclude the behaviors related to the avoidance of users to help, e.g., unconcerned, authoritative, handful, hateful.

4.2.4 Evaluating with non-experts

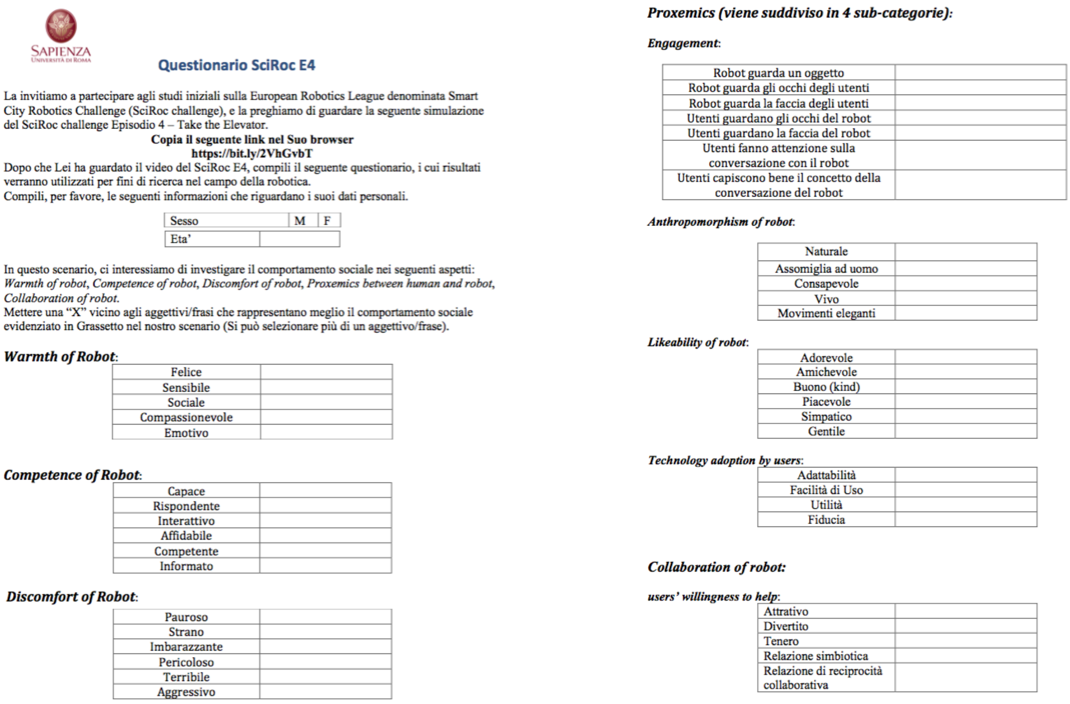
After the identification of preliminary robot's behaviours through the interviews, we performed a survey with 100 users having no previous experience with robots (i.e., non-experts). Such users were invited to watch the video recordings of the two demos: 51 of them watched the first demo, the other 49 watched the second demo. Users were selected randomly among the undergraduate students at Sapienza University of Rome. Once a specific user completed watching the video recording of a demo, s/he was asked through the survey to indicate which robot's behaviours s/he perceived as relevant during the recorded HRI task. The survey has been implemented in Italian language (See Fig. 4.6). In a nutshell, for any of the behaviours identified the user could decide to select (or neglect) it, e.g., a user could consider the interaction as Social but not Knowledgeable. We finally collected 100 answers to the survey.

4.3 Data Analysis and Results

4.3.1 Data Analysis

100 users participated our research study as non-experts. The gender distribution was as follows: 74% of male users, 24% of female users and 2% of users that did not declare their gender (see Fig. 4.7). The average age of users was 21,05 years old. The analysis of the data collected with the survey has been enacted following a three-steps approach.

- The first step: for each demo, we assessed how many times a behaviour was selected by the users. Specifically, if N is the amount of performed surveys for a specific demo and K is the number of times that a behaviour B is selected, we decided to consider B as relevant for the final questionnaire if $K/N > 0,5$, i.e., if more than half of the users selected B in the survey. Conversely, we immediately discarded all behaviours for which $K/N < 0,2$. For example, if



SAPIENZA UNIVERSITÀ DI ROMA

Questionario SciRoc E4

La invitiamo a partecipare agli studi iniziali sulla European Robotics League denominata Smart City Robotics Challenge (SciRoc challenge), e la preghiamo di guardare la seguente simulazione del SciRoc challenge Episodio 4 – Take the Elevator.

Copia il seguente link nel Suo browser
https://bit.ly/2VhGvbT

Dopo che Lei ha guardato il video del SciRoc E4, compili il seguente questionario, i cui risultati verranno utilizzati per fini di ricerca nel campo della robotica.
Compili, per favore, le seguenti informazioni che riguardano i suoi dati personali.

Sesso	M	F
Eta'		

In questo scenario, ci interessiamo di investigare il comportamento sociale nei seguenti aspetti: *Warmth of robot*, *Competence of robot*, *Discomfort of robot*, *Proxemics between human and robot*, *Collaboration of robot*.
Mettere una "X" vicino agli aggettivi/frasi che rappresentano meglio il comportamento sociale evidenziato in Grassetto nel nostro scenario (Si può selezionare più di un aggettivo/frase).

Warmth of Robot:

Felice	
Sensibile	
Sociale	
Compassionevole	
Emotivo	

Competence of Robot:

Capace	
Rispondente	
Interattivo	
Affidabile	
Competente	
Informato	

Discomfort of Robot:

Pauroso	
Strano	
Imbarazzante	
Pericoloso	
Terribile	
Aggressivo	

Proxemics (viene suddiviso in 4 sub-categorie):

Engagement:

Robot guarda un oggetto	
Robot guarda gli occhi degli utenti	
Robot guarda la faccia degli utenti	
Utenti guardano gli occhi del robot	
Utenti guardano la faccia del robot	
Utenti fanno attenzione sulla conversazione con il robot	
Utenti capiscono bene il concetto della conversazione del robot	

Anthropomorphism of robot:

Naturale	
Assomiglia ad uomo	
Consapevole	
Vivo	
Movimenti eleganti	

Likeability of robot:

Adorevole	
Amichevole	
Buono (kind)	
Piacevole	
Simpatico	
Gentile	

Technology adoption by users:

Adattabilità	
Facilità di Uso	
Utilità	
Fiducia	

Collaboration of robot:

users' willingness to help:

Attrativo	
Divertito	
Tenero	
Relazione simbiotica	
Relazione di reciprocità collaborativa	

Figure 4.6: Survey for non-experts evaluation

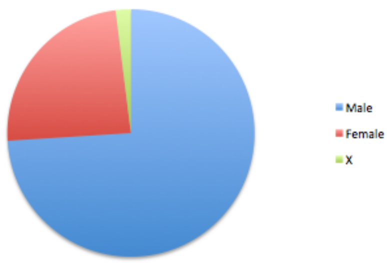


Figure 4.7: Gender distribution

we consider the first demo, the Social behaviour of the robot was selected K = 44 times out of N = 51 surveys (see Fig. 4.8). This means that more than 86% (i.e., 44/51 = 0,86) of users perceived the robot as *Social*. On the other hand, only 1 user perceived the robot as *Dangerous*, meaning that only 1,9% (i.e., 1/51 = 0.019) of users selected this behaviour as relevant. Consequently, it was filtered out by the questionnaire.

- The second step: for all those behaviours for which $0,2 \leq K/N \leq 0,5$,we performed a statistical test to investigate if it could be considered for the questionnaire or not. Specifically, we leveraged the binomial test [32], a statistical procedure that compares the observed frequencies of the two categories of a dichotomous variable to the frequencies that are expected under a binomial distribution with a specified probability parameter. The null hypothesis for this kind of test, which is also the one we wanted to confirm, is that the observed results do not differ significantly from what is expected. To be properly run, a binomial test requires

TABLE I
SOCIAL BEHAVIOR OF ROBOT (DEMO 1)

	k	1st step	2rd step(P-value)	3rd step
Warmth of robot				
Happy	23	-	0,576 ✓	✓
Feeling	2	✗	-	-
Social	44	✓	-	✓
Compassionate	1	✗	-	-
Emotional	1	✗	-	-
Competence of robot				
Capable	11	-	-	0,000 ✗
Responsive	26	✓	-	✓
Interactive	39	✓	-	✓
Reliable	2	✗	-	-
Competent	3	✗	-	-
Knowledgeable	17	-	0,024 ✗	-
Discomfort of robot				
Scary	2	✗	-	-
Strange	26	✓	-	✓
Awkward	13	-	0,001 ✗	-
Dangerous	1	✗	-	-
Awful	1	✗	-	-
Aggressive	0	✗	-	-

TABLE II
SOCIAL BEHAVIOR OF ROBOT (DEMO 2)

	k	1st step	2rd step(P-value)	3rd step
Warmth of robot				
Happy	12	-	0,000 ✗	-
Feeling	5	✗	-	-
Social	42	✓	-	✓
Compassionate	1	✗	-	-
Emotional	1	✗	-	-
Competence of robot				
Capable	18	-	0,085 ✓	✓
Responsive	30	✓	-	✓
Interactive	31	✓	-	✓
Reliable	5	✗	-	-
Competent	5	✗	-	-
Knowledgeable	12	-	0,000 ✗	-
Discomfort of robot				
Scary	0	✗	-	-
Strange	33	✓	-	✓
Awkward	16	-	0,021 ✗	-
Dangerous	0	✗	-	-
Awful	0	✗	-	-
Aggressive	0	✗	-	-

Figure 4.8: Social behavior of Robot

three ingredients: N, K, and a probability parameter, that in our case can be set to 0,5. For example, if we consider the first demo, 23 users perceived the robot as *Happy*, i.e., the observed proportion was K/N = 23/51 = 0,45. Our null hypothesis states that this proportion should be 0,5 for the entire population of N = 51 surveys. The outcome of the binomial test is the p-value, which in this case is 0,576, i.e., there is 57,6% of chance of selecting 23 times the behavior *Happy* out of N = 51 surveys performed with respective users. If this chance had been smaller than 5% ($p < 0{,}05$), the null hypothesis would have been rejected. In this example, the null hypothesis is confirmed, meaning that *Happy* can be considered as a relevant behaviour for the questionnaire. Conversely, on the other hand, if we consider the second demo (see Fig. 4.8), we found that 16 users perceived the robot as *Awkward* out of N = 49 surveys, i.e., the observed proportion was K/N = 16/49 = 0,32. In this case, the computed p-value was 0,021, meaning that there is 2,1% of possibilities that *Awkward* is selected 16 times in 49 experiments. Since $p < 0{,}05$, the null hypothesis is rejected. Consequently, this behaviour was filtered out by the questionnaire.

It is worth to notice that a behaviour that successfully passed the first or second step of the approach in one demo but not in the other was considered as sufficiently relevant for our purposes, meaning that it was inserted in the final questionnaire. This was the case of Robot looks at user's face, which did not pass the first and second step in the second demo, but passed the first step in the first demo (see Fig. 4.9).

- The third step: we used the *Pearson correlation* to investigate if it was possible to merge two different behaviours in a single one. In a nutshell, a *Pearson correlation* is a measure of the strength of the linear relationship between two dichotomous variables. The outcome is a number r that varies between -1 and 1, reflecting the extent to which two variables are linearly related. A value of r equal to -1/1 indicates a perfect negative/positive linear relationship between the two investigated variables. On the other hand, if r is equal to 0, there is no linear relationship between two variables. In our case, for all those behaviours who "survived" from the first two steps of the approach, we decided to apply the *Pearson correlation* and check if they were strongly (positively or negatively) correlated, i.e., with $r > 0{,}5$ or $r < 0{,}5$, or not. Just in one case we found a strong positive correlation. Analyzing the results of the surveys related to the first demo, under the *Proxemics between human and robot* perspective (see Fig. 4.9) we found that behaviours *Users pay attention to conversation with robot and Users understand well the meaning of conversation* were strongly related in a positive way ($r = 0{,}704$). However, since this same correlation was not verified also in the second demo ($r > 0$ but $r < 0{,}5$, see Fig. 4.9), we finally could not merge the two behaviours in a single one.

The results of the surveys analyzed through our three-steps approach are shown in Tables I, II, III, IV, V and VI (see Fig. 4.8, 4.9, 4.10). Results are categorized per demo and according to the considered perspective.

4.3.2 Implementing the questionnaire

The results of the survey allowed us to identify the 17 behaviors to include in the questionnaire, which have been selected by researchers (i.e, experts) and users (i.e., non-experts) for E4 task scenario. We implemented the questions of the questionnaire. The questions were organized as follows, according to the perspective they belong to:

Social Behavior of robot

- Have you perceived happiness of the robot?
- Have you perceived sociability of the robot?
- Have you perceived capability of the robot?
- Have you perceived responsiveness of the robot?
- Have you perceived interactiveness of the robot?
- Have you perceived strangeness of the robot?

Proxemics between human and robot

- Did robot look at users' face during the conversation between user and the robot?
- Did user look at the robot's face during the conversation between user and the robot?
- Have users paid attention to the conversation with the robot?
- Have users understood well the meaning of conversation?
- Have you perceived consciousness of the robot?
- Have you perceived friendliness of the robot?
- Have you perceived politeness of the robot?
- Have you perceived adaptability of the robot?
- Have you perceived ease of use with the robot?

Collaboration with robot

- Have you perceived enjoyment of the robot?
- Have you perceived collaborativeness of the robot?

Any question can be answered selecting one item from a 5-point Likert scale [46]. If the robot's behavior involved in a specific question reflects a positive attitude, answers are of kind: *Absolutely No=1*, *No=2*, *Neutral=3*, *Yes=4*, *Absolutely Yes=5*. Conversely, in case of a negative attitude, e.g., for a negative behavior like *strange*, the scores assigned to the answers are turned over, i.e., *Absolutely No=5*, *No=4*, *Neutral=3*, *Yes=2*, *Absolutely Yes=1*. In the end, we added the personal information to the questionnaire as well.

The questionnaire has been developed on basis of the task scenario of E4, it was adopted to evaluate the performance of robots in E4 at the first SciRoc challenge. This questionnaire has been provided to the teams in advance. Teams could address the HRI studies concerning the robot's behaviors appeared in the questionnaire, and improve the performance of robots in E4. However, we believe that it stands as a novel contribution to the evaluation of performance of robots in social and smart city contexts.

4.4 Discussion and Conclusion

We present an approach to design a new type of questionnaire as a task-driven evaluation technique for measuring robot's behaviors perceived by users in social contexts. We consider our approach to design the questionnaire as a first step towards systematic evaluation of robots performance in social settings that specifically address HRI issues. The approach consists of two steps: interview with experts and survey with non-experts.

The aim of this study is to arise awareness of the issues of evaluation techniques in HRI research study. The approach describes the development process of the evaluation technique addressing human behaviors toward robots and robot's behaviors perceived by human in social contexts. The questionnaire was developed to measure the extent to which individuals report user perception of robot's behaviors. The texts, structure and scale of the questionnaire are created in the phase of the questionnaire development. In order to avoid the problems concerning individual differences, subjective measurement (i.e., Social behaviors of robot, etc.) and objective measurement (i.e., Robot looks at your face, etc.) are both included in the questionnaire E4.

However, the future work of the questionnaire is to validate the robustness, the replicability of our approach and the reliability of the developed questionnaire. The validation study is needed to be conducted by administering it to a larger and more diverse set of users during the enactment of the episode E4, e.g., Putten et al. [135] proposed a validation study for the questionnaire proposed in the field of HRI. We should furthermore propose an exploratory study in a realistic social environment, by exploiting the questionnaire developed, to validate its reliability. In Chapter 5, we present an HRI exploration research study in the realistic social context of the SciRoc challenge to validate the questionnaire proposed and to explore the relevant results concerning *human factors*.

TABLE III

PROXEMICS BETWEEN HUMAN AND ROBOT (DEMO 1)

	k	1st stp	2rd stp(p)	3rd stp
Engagement				
Robot looks at an object	11	-	0,000 ✗	-
Robot looks at user's eyes	9	✗	-	-
Robot looks at user's face	33	✓	-	✓
Users look at robot's eyes	11	-	0,000 ✗	-
Users look at robot's face	22	-	0,401 ✓	✓
Users pay attent. to conversa.	18	-	0,049 ✓	✓
Users understand well conversa.	19	-	0,092 ✓	✗
Anthropomorphism of robot				
Natural	11	-	0,000 ✗	-
Human-like	11	-	0,000 ✗	-
Conscious	12	-	0,000 ✗	-
Lifelike	11	-	0,000 ✗	-
Moving elegantly	10	✗	-	-
Likeability of robot				
Like	10	✗	-	-
Friendly	27	✓	-	✓
Kind	16	-	0,011 ✗	-
Pleasant	6	✗	-	-
Nice	16	-	0,011 ✗	-
Politeness	32	✓	-	✓
Technology adoption				
Adaptability	8	✗	-	-
Ease of Use	23	-	0,576 ✓	✓
Useful	14	-	0,002 ✗	-
Trust	11	-	0,000 ✗	-

TABLE IV

PROXEMICS BETWEEN HUMAN AND ROBOT (DEMO 2)

	k	1st stp	2rd stp(p)	3rd stp
Engagement				
Robot looks at an object	5	✗	-	-
Robot looks at user's eyes	1	✗	-	
Robot looks at user's face	12	-	0,000 ✗	-
Users look at robot's eyes	1	✗	-	-
Users look at robot's face	17	-	0,044 ✓	✓
Users pay attent. to conversa.	22	-	0,568 ✓	✓
Users understand well conversa.	23	-	0,775 ✓	✓
Anthropomorphism of robot				
Natural	4	✗	-	-
Human-like	2	✗	-	-
Conscious	32	✓	-	✓
Lifelike	5	✗	-	-
Moving elegantly	3	✗	-	-
Likeability of robot				
Like	2	✗	-	-
Friendly	23	-	0,775 ✓	✓
Kind	15	-	0,009 ✗	-
Pleasant	9	✗	-	-
Nice	8	✗	-	-
Politeness	24	-	1,000 ✓	✓
Technology adoption				
Adaptability	18	-	0,085 ✓	✓
Ease of Use	25	✓	-	✓
Useful	16	-	0,021 ✗	-
Trust	6	✗	-	-

Figure 4.9: Proxemics between human and robot

TABLE V
COLLABORATION WITH ROBOT (DEMO 1)

	k	1st step	2rd step (p-value)
Users' willingness to help			
Attractiveness	9	✗	-
Enjoyment	7	✗	-
Endearment	16	-	0,011 ✗
Symbiotic Relationship	7	✗	-
Reciprocity Relationship	10	✗	-
Collaborativeness	27	✓	-

TABLE VI
COLLABORATION WITH ROBOT (DEMO 2)

	k	1st step	2rd step(p)	3rd step
Users' willingness to help				
Attractiveness	9	✗	-	-
Enjoyment	18	-	0,085 ✓	✓
Endearment	2	✗	-	-
Symbiotic Relationship	6	✗	-	-
Reciprocity Relationship	9	✗	-	-
Collaborativeness	27	✓	-	✓

Figure 4.10: Collaboration with robot

Chapter 5

Experimental Study: HRI Users' studies in the context of the SciRoc Challenge

5.1 Introduction

An exploratory research study, which differs from confirmatory research studies, means as a first attempt to verify the possibility of identifying factors that could be useful for the development of HRI [74], i.e., it seeks what factors might be important and which outcomes are possible [14]. HRI researchers consider exploratory research studies as qualitative research methodologies, because they can help researchers to explore underlying findings in the field of HRI. For instance, De Graaf et al. [48] propose an exploratory research study that investigates the acceptance factors of social robots, Mutlu et al. [112] explore how cooperation versus competition in a game shaped people's perceptions of the social robot ASIMO.

The robotic competitions provide good opportunities for conducting exploratory research studies in the field of HRI, because among the principles that characterize robotic competitions, replicability is considered fundamental to allow for rigorous comparison of results and thus affects the processes and products of scientific research [6]. Yanco et al. introduce HRI studies with the goal of identifying areas for improvement that would lead to better HRI design and overall robot performance in the DARPA Robotics Challenge Trials and Finals [128, 189]. In social contexts, an exploratory research study has been conducted in the context of ICRA 2008 challenge evaluated by experts and non-experts [185]. Even though Robocup@home [77] emphasized the importance of HRI as a primary capability, there was no specific HRI studies conducted in the context of the robotic competition.

In this chapter, we present an exploratory research study by conducting an experiment in the context of the SciRoc challenge, a repeatable and general-purpose test method developed for HRI performance evaluation, aiming at investigating robot's behaviors perceived by human and human behaviors toward robot. Our research study has as the main novelty the fact of having been devised and implemented in (i) *a realistic and dynamic social environment*; (ii) through a *representative samples of participants* selected by the SciRoc organization; (iii) employing robots configured to *act autonomously*, without the need of any external guidance. To perform our research study, we focused on "Take the elevator (E4)" that contains several elements for social HRI.

5.2 SciRoc challenge

SciRoc is a EU-H2020 funded project supporting the European Robotics League (ERL) and whose purpose is to bring ERL tournaments in the context of smart cities. This adds a new challenge to ERL, which has been pursued through the organisation of two SciRoc challenges in 2019 and 2021. A key novelty of the SciRoc project is the introduction of robots in social and smart cities, i.e., the set of communications, connections, actions that exist in scenarios where both robots and the city infrastructures are involved [166].

The objective of the first SciRoc as scientific competition is to provide a task benchmark to teams, which allows for the measurement of performance, and to develop software and HRI solutions. Task benchmark means measuring performance using a methodology that is repeatable, reproducible, accurate and evaluating on task-specific criteria that look at the success and quality of task execution.

The first SciRoc challenge has been held in the shopping mall of Milton Keynes (UK)[1], where a set of networked devices provided static and dynamic information from a number of heterogeneous data sources, e.g., location of shops, current availability of items, audio/visual inputs from CCTV cameras, crowd density sensor information, and many others, from 16 to 22 September 2019. In addition, robots interacted with MK customers, accomplishing tasks of different nature in three different scenarios: assisting customers, providing professional services and supporting during emergency situations. Five episodes (see Fig.5.1) were selected and enacted during the challenge:

- Delivery coffee shop order (E3), the robot assisted customers in a coffee shop to take care of customers, by taking orders and bringing objects to and from customers' tables.
- Take the elevator (E4), the robot took the elevator crowded with customers to reach a service located in another floor.
- Shipping pick and pack (E7), the robot is located in one of the booths of the mall, and on the shelves there are some goodies that are displayed for sale to customers. Customers can place orders through a tablet. The robot must move and collect the requested packages for the customer, placing them in a box.
- Through the door (E10), this episode is based on a door assembly comprising that the robot had to operate, and a surrounding test area where the robot could move.
- Fast delivery of emergency pills (E12), the aerial robot attended an emergency situation in which a first-aid kit needed to be delivered to a customer. The robot was able to fly autonomously to the customer location as fast as possible.

Figure 5.1: Five episodes of 1st SciRoc: E3; E4; E7; E10 and E12.

[1] https://www.centremk.com/

The competition arena was set up in MK Hall using a straight square truss system, the episode specific arenas were located around the main central area.

5.3 Methods and Materials

5.3.1 participants

The main theme of the competition was the interaction between robots and smart cities, or more in general, to showcase to the general public how robots can coexist in a public scenario. For these reasons, some episodes (i.e., Deliver coffee shop orders (E3) and Take the elevator (E4)) required direct interactions between the robot and humans. Involving directly the customers of the shopping mall was infeasible for multiple reasons (e.g., health and safety regulations), therefore the SciRoc organization preselected representative samples (i.e., participants) with no background in robotics (see Fig. 5.2). The participants were uniformly distributed in age and had a variety of personal and work backgrounds. Specially, 40 participants have been selected by the SciRoc organization for our research study.

Figure 5.2: Participants involved in our research study

5.3.2 Realistic environment

To create a social realistic environment, we implemented the competition area of the SciRoc challenge in the Centre:MK shopping mall (see Fig. 5.3). The episode-specific arenas were located around the main central area. The aim of E4 is to evaluate the interaction between the robot and the customers in a restricted space such as an elevator. To achieve this we recreated a mock-up lift inside the arena, complete with movable doors (see Fig. 5.4). In addition, the elevator has been equipped with a video camera showing to the audience the episode in progress (see in Fig. 5.5).

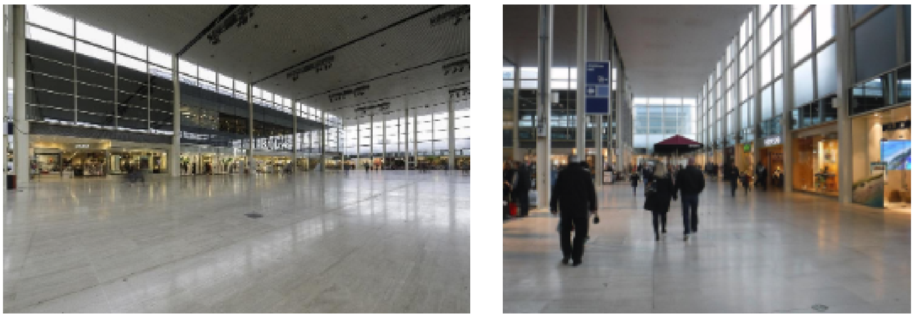


Figure 5.3: Competition Arena

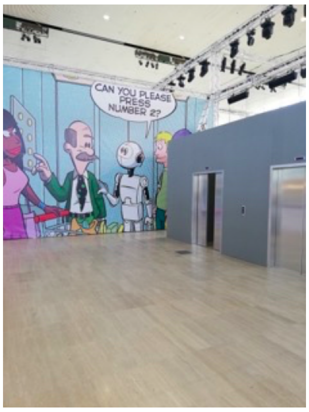


Figure 5.4: Competition Arena E4

5.3.3 Ethical consideration

HRI research on people's interactions with and social reactions towards robots is necessary to shape the *ethical*, *societal* and *legal perspectives*, and facilitates the design of responsible robotics and the successful introduction of robots into our society [47]. In the emerging literature devoted to "ethical robots", HRI researchers proposed some ethical guidelines and consideration. For instance, Riek and Howard have proposed principles for an HRI code of ethics in four aspects: (i) *human dignity considerations*, (ii) *design considerations*, (iii) *legal considerations*, (iv) *social considerations* [138]. Specially, the authors proposed a social principle concerning *Wizard-of-Oz* methodology, and suggested Wizard-of-Oz should be employed as judiciously and carefully as possible, and should aim to avoid Turing deceptions.

On the basis of the existing ethical principles, we proposed our ethical considerations and introduced them in the context of the SciRoc challenge as follows:

- The robot cannot hit a human.
- The robot cannot hit or damage the furniture or object.
- Team members cannot give instructions to the robot during the task performance

If one of the aforementioned situations occurs, task performance of E4 must be stopped immediately.

5.3.4 Robots

Five teams have entered to "Take the elevator-(E4)", employing five different robot variants (see Fig. 5.6):

Figure 5.5: Competition Arena E4: a view from the video camera

- UC3M, a robotics laboratory from University Carlos III of Madrid, the team has deployed TIAGo robot produced by PAL robotics; in addition to the basic platform, the robot had a manipulator[2].

- Gentlebots[3], a team of researchers in robotics from the Rey Juan Carlos University and the University of Leon, the team has deployed TIAGo robot by PAL robotics, in addition to the basic platform, the robot had one tablet and one microphone in front, and the status of the robot was shown on the tablet.

- HEARTS, the healthcare engineering and assistive robotics technology and services (HEARTS) team is based in the Bristol Robotics Laboratory, a collaboration between the University of the West of England (UWE) and the University of Bristol, the team has brought a Pepper robot produced by Softbank.

- eNTiTy, a team of the R & D department of NTT Disruption in Spain, the team has deployed TIAGo robot produced by PAL robotics, in addition to the basic platform, the robot had a manipulator and a signal light stuck on the head, the signal light can change the color in according to the speech recognition status.

- LASR[4], an AI group team from the University of Leeds, the team has deployed TIAGo robot produced by PAL robotics, in addition to the basic platform, the robot had a manipulator.

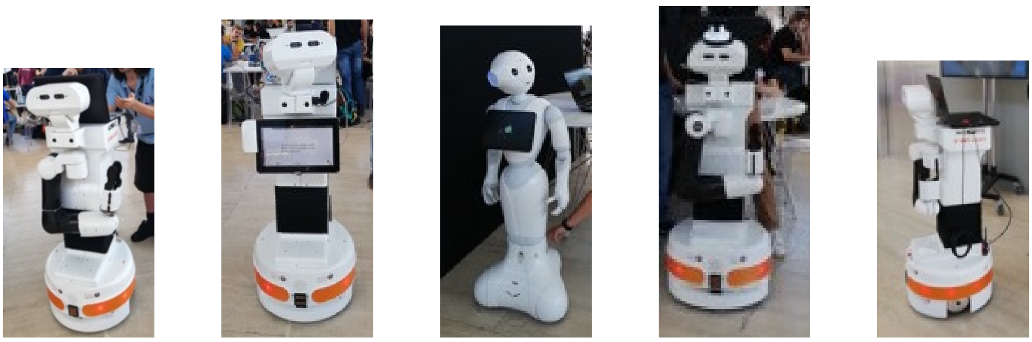


Figure 5.6: Robots employed during E4, developed respectively by UC3M, Gentlebots, HEARTS, eNTITY and LASR teams.

[2] https://tiago.pal-robotics.com/
[3] http://www.gentlebots.robotica.gsyc.es/
[4] https://sensiblerobots.leeds.ac.uk/lasr/

5.3.5 Robot's functionalities

Team members have implemented the robot's functionalities independently. According to the rulebook E4, main functionalities tested in this episode are *Navigation respecting proxemics when the robot interacts with customers*, *Spoken Dialogue with customers*, *People detection* and *Object detection.* Robots were required to act autonomously in E4, without the need of any external guidance.

For instance, LASR team made use of the ROS package move base ([5]) with proprietary PAL planners for both global and local planning. LASR team has implemented 3D obstacle avoidance by incorporating a Voxel Costmap layer which uses the depth sensor to create obstacles in the 2D map for planning as well. The LASR team software stack is shown in Fig. 5.7.

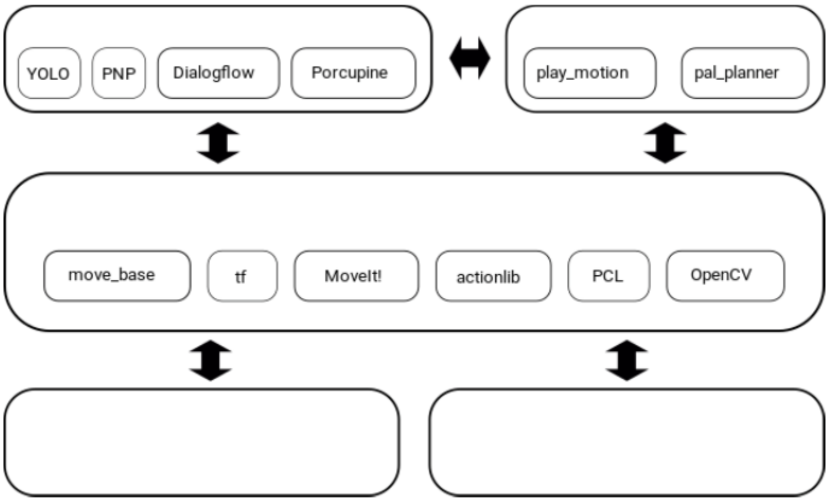


Figure 5.7: Software stack implemented by LASR

Gentlebots team designed a cognitive architecture for the robot (see Fig.5.8), and used ROSPlan [35] and BICA [39] to implement the architecture. ROSPlan is a planning framework, whereas BICA is a toolbox to create software architectures for robots. Tier 1 starts the instances and predicates of the problem to be solved. When a state machine at Tier 1 establishes a goal, the planner at Tier 2 creates a plan using the content of his knowledge base. Tier 3 contains the implementation of the actions defined in the PDDL domain. Tier 4 contains skills that can be activated from actions. This level includes perceptual, attention, dialogue, and manipulation modules, among others. Knowledge Graph stores the information relevant to the operation of the robot.

5.3.6 Experimental design

The SciRoc challenge lasted 4 days. In total, 10 runs have been scheduled in E4: 9 runs have been scheduled in the first 3 days, and the final run has been scheduled in the last day. In each general run, all the five teams have performed in randomized order. The duration of the task was around 10 minutes (i.e., short-term interaction). In the final run, four teams which had higher scoring have performed in randomized order (see Table 5.1).

The evaluation experiment relied on the *mixed-model factorial* design, which includes both

[5]http://wiki.ros.org/move$_{b}ase$

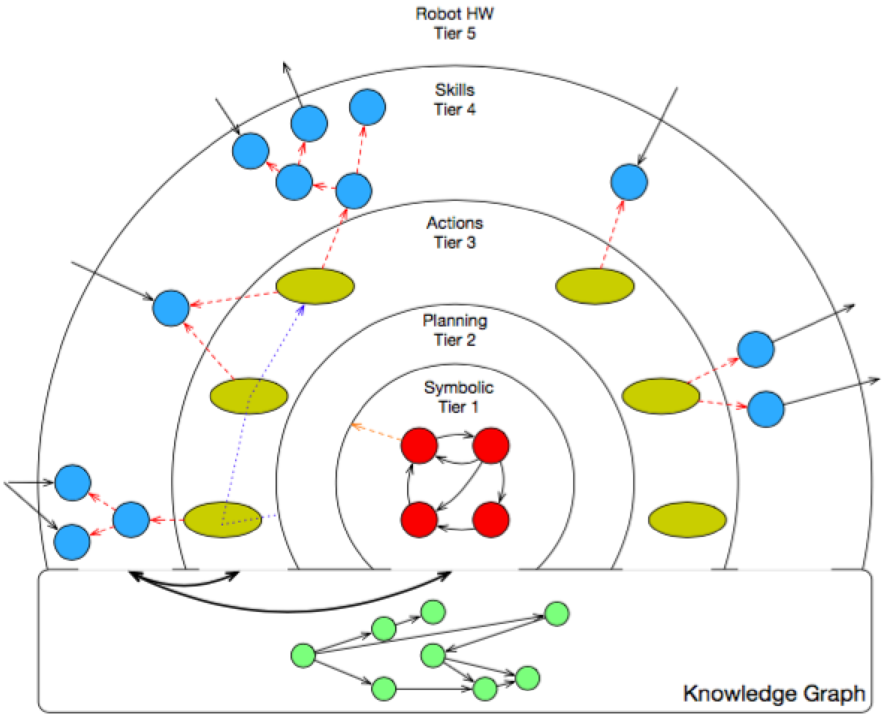


Figure 5.8: Cognitive architecture implemented by Gentlebots

between-subjects and within-subjects design component [25]. Specially, the study involved a total of 40 users. The same 4 users participated only in one of the 10 runs. In each run, five (or four) different teams/robots (with-in subject) performed the test according to the run schedule. User's gender (between subject) was declared by users, before the starting of each run. User's role (between subject) was assigned by E4 referee (1 user was assigned to role A, 1 user was assigned to Role B, 2 users were assigned to role C), before the starting of each run.

Table 5.1: Run schedule

	Trial 1	**Trial 2**	**Trial 3**	**Trial 4**	**Trial 5**
Run 1	UC3M	eNTiTy	LASR	HEARTS	Gentlebots
Run 2	UC3M	eNTiTy	LASR	Gentlebots	HEARTS
Run 3	LASR	eNTiTy	UC3M	Gentlebots	HEARTS
Run 4	UC3M	Gentlebots	eNTiTy	LASR	HEARTS
Run 5	eNTiTy	LASR	UC3M	HEARTS	Gentlebots
Run 6	LASR	eNTiTy	UC3M	HEARTS	Gentlebots
Run 7	Gentlebots	HEARTS	UC3M	eNTiTy	LASR
Run 8	HEARTS	LASR	UC3M	Gentlebots	eNTiTy
Run 9	eNTiTy	HEARTS	LASR	UC3M	Gentlebots
Final Run	HEARTS	LASR	eNTiTy	Gentlebots	

Table 5.2: Distribution of participants by genders per run

		Male Users	Female Users
Day 1	Run 1	2	2
	Run 2	2	2
	Run 3	3	1
Day 2	Run 4	3	1
	Run 5	4	0
	Run 6	1	3
Day 3	Run 7	1	3
	Run 8	2	2
	Run 9	2	2
Day 4	Final Run	3	1

5.3.7 Evaluation technique

At the end of any run, the participating users filled a dedicated questionnaire built ad-hoc for this episode [176]. The questionnaire has been thought to specifically keep track of 17 behavioural aspects related to: (i) social behavior of robot, (ii) proxemics between human and robot, and (iii) collaboration with robot (see Chapter 4). The scores assigned in the scale range from: Absolutely No=1 to Absolutely Yes=5. If compared with the original questionnaire, we decided to convert the only negative behavior, "Perceived Strangeness", into its "positive version", i.e., "Perceived Naturalness ". The questionnaire of E4 adopted by our experiment is shown in Fig. 5.9.

5.4 Data Analysis and Results

5.4.1 Data Collection

40 users have participated 10 runs in E4. Since we knew that the diversity of users is fundamental in an experiment, in our case, the users were selected by the SciRoc organization committee randomly. They were diversified as workers, secretaries, employees, students, retirees and etc. Gender distribution was as follow: 23 male users and 17 female users. Age distribution was as follows: 3 users from 18-19 years old, 11 users from 20-29 years old, 16 users from 30-39 years old, 6 users from 40-49 years old, 3 users from 50-59 years old, 1 user over 60 years old (See Fig. 5.10). The distribution of gender per run is shown in Table 5.2. We collected 196 questionnaires overall, of which 78 were considered incomplete (i.e., not filled at all or not filled completed because of a failed test in a run, which was a circumstance happened especially in the first day of the competition), 118 questionnaires were considered completed.

5.4.2 Reliability of the questionnaire

The importance of measuring the accuracy and consistency of research instruments (especially questionnaires) known as validity and reliability [28]. Validity is often not well addressed in course questionnaire design as there are no straightforward tests that can be applied to an individual instrument [86]. Reliability refers to the degree to which the results obtained by a measurement and procedure can be replicated [129]. Cronbach's alpha is a test reliability technique that requires only a single test administration to provide a unique estimate of the reliability for a given test [63].

We calculated the Cronbach's alpha coefficient (α) for all of the three macro categories of the questionnaire, obtaining the following results: α of *Social Behavior of robot* = 0.907; α of *Proxemics between human and robot* = 0.921; α of *Collaboration with robot* = 0.83. According to [60], which discusses cut-off values for reliability indices, values of α coefficient greater than 0.8 indicate a reliability of the adopted scale among very good and excellent.

5.4.3 Analyzing the relationship between task performances evaluated by scoring and robot's behaviors perceived by participants

Task performances evaluated by scoring

The scoring in a regular test, which is specified in terms of task [186], mainly reflects the key features of robotic system or task performance. To measure the task performance of E4 properly, we drew up the scores which together test the integration of functional abilities. The scoring of E4 was determined by two sets: achievements score and penalties score. According to the principles of HRI ethical consideration (see Section 5.3.3), we determined the disqualifying behaviors. If one of the disqualifying behaviors occurred, the task performance was stopped and any score achieved so far was canceled.

Achievements Score

- The robot properly deals with Customer A (avoidance, no interaction).
- The robot properly deals with Customer B (interaction).
- The robot enters the elevator.
- The robot declares the target floor to Customer C.
- The robot exits the elevator at the proper floor.
- The robot reaches the finish area.
- Above 80% of positive users' perceptions (>=4) over total valid answers of questionnaire.

Penalties Score

- Robot requires a customer to move away to avoid a collision.
- A customer instructs the robot to move away from one location.
- The robot acts customers' requests wrongly.
- The robot obstructs the way to the customers.
- Above 80% of negative users' perception (<=2) over total valid answers of questionnaire.

Disqualifying behaviours

- The robot hits a human.
- The robot hits and damages the furniture and/or objects.
- Team members give instructions to the robot during the task performance.

Each item represents one score, we added the results of the questionnaire to the scoring. In each general run, the aggregate score has been determined by the third highest score according to the ERL system. The top 4 teams in the ranking were qualified for the final. The final ranking for assigning the first, second and third place was determined by the performance in the Final. The E4 Score Sheet is shown in Table 5.11.

Regression analysis

We were interested in exploring the relationship between the scoring and the results of the questionnaire. In particular, we were interested in exploring if robot's behaviors could be predictors of scoring. Since participating teams of E4 have been penalized few times and never been disqualified, we approximated the achievements score as the overall score to check our exploratory research question.

We first subtracted the scores obtained from the questionnaire from the achievements scores (LASR and Entity got scores from the questionnaire), and then conducted *Multi Linear Regression study* for data analysis. We calculated the R^2 value of regression for all the three macro categories of the questionnaire, obtaining the following results: R^2 of Social Behavior of robot = 0,572; R^2 of Proxemics between human and robot = 0,431; R^2 of Collaboration with robot = 0,515. According to [136], R^2 represents the goodness of fit the model which cut-off value of R^2 is 0,5. Hence, the macro category *Proxemics between human and robot* of the questionnaire cannot be further analyzed in our analysis study.

For the following items related to *Perceived Interactiveness* and *Perceived Collaborativenss*, we found significant values as follows: *Perceived Interactiveness* (p=0,03 B=0,78 t-value=3,33) and *Perceived Collaborativenss* (p=0,00 B=0,842 t-value=4,402), meaning these behaviors are significant predictors of the performance scoring of E4. The beta coefficient is the degree of change in the outcome variable for every 1-unit of change in the predictor variable. It can be positive or negative. In our case, we obtained two positive predictors: *Perceived Interactiveness* and *Perceived Collaborativenss*, i.e., for every 1 unit of change positively in users' perception on interactiveness of robot's behavior, performance scoring will increase 0.78 as the degree of change; for every 1 unit of change positively in users' perception on collaborativeness of robot's behavior, performance scoring will increase by the beta value 0.842. These results support the relationship between the scoring of E4 and the results of questionnaire E4, i.e., the subjective evaluation of HRI by participants can indeed reasonably approximated based on objective scoring. It is worth to notice that similar finding has been revealed in [107].

5.5 Discussion and Conclusion

The aim of this study is to investigate the relationship between the performance scoring obtained in the competition context and robot's behaviors perceived by users measured through the dedicated questionnaire.

In this chapter, we present an exploratory research study in the context of the SciRoc challenge by introducing an experiment in Episode 4 - "Take the elevator". The results of data analysis confirms the relationship between the scoring of task performance and the results of questionnaire E4 evaluated by participants.

When we first start our intention to bring an exploratory research study in a scientific robotic competition to the heart of a major public space, our ambition was to stimulate fact based discussion by the public about the future roles of robots in their social context, to demonstrate that HRI studies are at the core of robotic research. This concept of conducting exploratory research studies can reproduce in several different competition contexts. This concept can improve the *quality* of research studies in the field of social HRI because of the particularities of features aforementioned. We encourage HRI researchers to reproduce empirical studies in the other social task scenario contexts to test the validity of our research methodology. In Chapter 6, we focus on developing a new approach for designing an HRI task scenario in a scientific robot competition.

SciRoc

Questionnaire E4 Take the Elevator

Thanks a lot for your participating in the performance of the task Episode 4 - Take the Elevator.

Please fill out the following questionnaire, the results of which will be used for research purposes in the field of robotics.

Please fill in the following information regarding your personal data.

Gender	M	F
Age		

After you have interacted or observed the robotic behaviors in E4, please select one from each 5-point likert-scale. **If the behavior is not possible to evaluate, just leave the blank.**

Social Behavior of robot

	Absolutely Yes (5)	Yes (4)	Neutral (3)	No (2)	Absolutely No (1)
Have you perceived happiness of the robot?					
Have you perceived sociability of the robot?					
Have you perceived capability of the robot?					
Have you perceived responsiveness of the robot?					
Have you perceived interactiveness of the robot?					
Have you perceived naturaless of the robot?					

Proxemics between human and robot

	Absolutely Yes (5)	Yes (4)	Neutral (3)	No (2)	Absolutely No (1)
Did robot look at users' face during the conversation between user and the robot?					
Did user look at the robot's face during the conversation between user and the robot?					
Have users paid attention to the conversation with the robot?					
Have users understood well the meaning of conversation?					
Have you perceived consciousness of the robot?					
Have you perceived friendliness of the robot?					
Have you perceived politeness of the robot?					
Have you perceived adaptability of the robot?					
Have you perceived ease of use with the robot?					

Collaboration with robot

	Absolutely Yes (5)	Yes (4)	Neutral (3)	No (2)	Absolutely No (1)
Have you perceived enjoyment of the robot?					
Have you perceived collaborativeness of the robot?					

Figure 5.9: Questionnaire E4 utilized in our research study

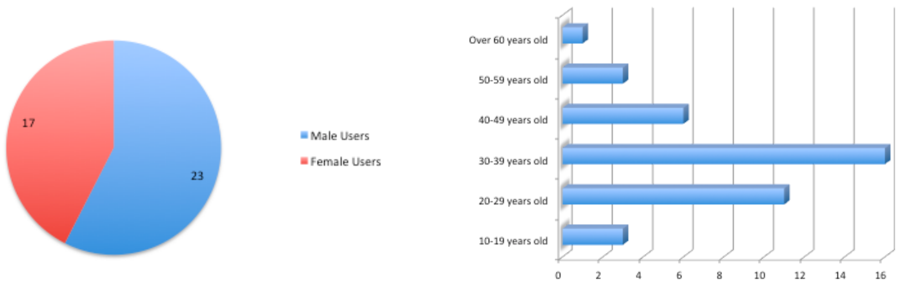


Figure 5.10: Gender distribution and Age distribution

GENTLEBOTS			ENTITY			LASR			HEARTS			UC3M		
Run	A	P	Run	A	P	Run	A	P	Run	A	P	Run	A	P
1	3	1	1	1	0	1	1	2	1	0	0	1	0	0
2	1	0	2	6	0	2	1	0	2	0	0	2	0	0
3	5	0	3	3	2	3	2	0	3	1	0	3	0	0
4	4	0	4	5	0	4	3	0	4	1	0	4	0	0
5	1	0	5	5	0	5	1	0	5	5	0	5	3	0
6	5	0	6	0	0	6	6	0	6	1	0	6	1	0
7	5	0	7	1	0	7	2	0	7	3	0	7	1	0
8	4	0	8	5	0	8	6	0	8	5	0	8	2	1
9	5	0	9	4	0	9	1	0	9	1	0	9	2	0
Final1°-2°	4	0	Final 1°-2°	2	0	Final 3°-4°	7	0	Final3°-4°	5	0	-	-	-

Figure 5.11: Score Sheet

Chapter 6

Task Scenario: Designing and evaluating HRI teaming tasks based on the SciRoc challenge

6.1 Introduction

Robots coexist with human beings in a real social context and provide the needed and the favorable services accordingly. There were still a lot of limitations in the performance of current robots in a real social context so far. To address the human behaviors toward robots and the robot's behaviors perceived by human in a social context, it is necessary to provide a task scenario that mimics the real interaction in social contexts and takes into account the user's requirements and the functions that the robot can provide [3].

Such viewpoints are found in the field of HRI research, where researchers are typically addressing robots in a particular service robotics task scenario, e.g. robots as assistants, where social interaction with people is necessarily part of the research agenda [43]. For instance, Glas et al. [62] implemented the task scenarios to showcase typical functionalities used in the field of social HRI.

To provide an appropriate design of task scenarios in a social context, we summarize some principles that were set out by previous HRI researchers as follows: *suitability*, task scenarios should be constructed on suitability that will be executed by the robot [104]; *usability*, task scenarios of HRI should be used to help convey meaningful context of the interactions to the evaluators, so as to induce proper feedback on robots [188]; *continuously*, task scenarios rely on a continuous interpretation of the sensor data to generate expressions [79]; *applicability*, task scenarios should be applied to novel applications.

Based on the aforementioned principles for task scenario design, we develop a new approach for designing task scenarios in the field of social HRI, and implement the new approach in the SciRoc challenge to design the episode "Take the elevator - E4". The new approach consists of two steps. First, it relies on analyzing the elements of the task scenario. Then, it sketches out the layout of the task scenario.

"Take the elevator - E4" is the most social episode of the SciRoc challenge, in which robot must take an elevator of the shopping mall with customers to reach a service located in another floor. We end this chapter by showing an example of HRI confirmatory research study leveraging on the task scenario proposed.

6.2 Approach for designing task scenarios

6.2.1 Analyzing the elements of E4 task scenario

Referring the framework proposed by [7] of task scenarios analysis in social contexts, we analyzed the elements of E4 task scenario as follows:

- *Environment*, the central aim of E4 is to conduct the task performance in a restricted space such as an elevator. To achieve a *realistic dynamic social environment*, we recreated a mock-up lift inside the arena, complete with movable doors.
- *Autonomy*, E4 requires a fully autonomy robot to enact the task performance. Based on the experience matured on autonomous system from previous competitions, the system requires the integration of a large set of abilities and technologies including HRI abilities, navigation, reasoning, planning, behavior control, object recognition, etc. [186]. The technical committee of E4 has released the rulebook in advance in order to leave necessary time for participated teams to develop fully autonomy robots.
- *Composition*, Parashar and colleagues [134] divided team composition of human-robot teaming interaction into five categories: (i) single human to single robot; (ii) single human to multiple-homogeneous robots; (iii) single Human to multi-heterogeneous robots; (iv) multi-homogeneous humans to single/multi-homo/multi-hetero robots; (v) multi-heterogeneous humans to single/multi-Homo/multi-Hetero robots. To create a *realistic situation* that humans encounter in a shopping mall, we employed multi-heterogeneous humans (i.e., 4 participants) to single robot, and each user was assigned a specific *role*. Each role is assigned to a participant who represents a target population: **role A** represents people who are not friendly with robots; **role B** represents people who are friendly with robots; **role C** represents people who communicate and interact with robots in a public space. Specifically, **role A** stands in a predefined location not far from the elevator. S/he can observe the robot but it is not interested in interacting with it, **role B** actively moves towards the robot willing to interact it, **role C** takes the elevator with the robot and gets off before/with/after the robot, see Fig. 6.1.
- *Task*, E4 is a collaborative task; the robot must take an elevator of shopping mall crowded with customers (i.e., role C) to reach a service located in another floor. The robot can enter/exit the elevator at the right floor in the presence of people nearby and/or inside and can interact with the customers in spoken language. The robot is not supposed to push buttons, but it can ask the customers around to do it. To be more specific, the robot encounters two customers while moving towards the elevator (i.e., role A and role B). Once arrived in front of the elevator, the robot encounters two additional customers (i.e., role C), both customers take the elevator with the robot. At this point, the robot interacts randomly with one of them asking to push the button for the floor it wants to reach. The two persons that take the elevator are instructed to reach a specific floor, which can be different (or the same) from the one assigned to the robot.

- *Embodiment*, E4 requires the participating teams to employ the physical humanoid robot or physical non-humanoid robot, virtual agents are banned to access in E4.
- *Duration*, the task duration of E4 is around 10 minutes (i.e., short-term interaction).

Figure 6.1: E4 Role A; Role B; Role C

6.2.2 Sketching out the layout of E4 task scenario

Once we analyzed the elements of E4 task scenario, we sketch out the layout of E4 task scenario.

Encounter Situation (Phase 1)

The robot enters in the competition arena and continues the path toward Zone A and Zone B (see Fig. 6.2); role A and role B are deployed in Zone A or Zone B randomly, and they can shift the zones during the run. The expected behavior of the robot is not to interact with role A and to interact with role B. (i.e., when the robot detects a customer who is not interested in interacting, the robot should just avoid him/her and proceed without any attempt to communicate; when the robot detects a person who is interested in interacting, the robot should stop moving and communicate with him/her accordingly.)

Entering/exiting the elevator (Phase 2)

Once the robot passed Zone B, it moves toward Zone C. Zone C is occupied by two standing customers (Role C), it should place itself at a proper location outside the elevator depending on the location of other customers. Elevators doors are operated by the Referee after the robot signal that it has reached its desired location outside the elevator. When the elevator door opens, the robot has to wait until all the customers around enter the door, then it can move and enter the door occupying a proper space in the elevator cabin. The robot is not able to press the button, hence, it has to face one customer, declare its target exit floor, and asks for help to press the button. The target exit floor of the robot has been communicated through MK data hubs. When the door opens, the referee declares the current floor, the robot is allowed to ask to one customer (while facing him/her) which is the current floor. If the current floor is the destination one for the robot, it must exit, otherwise, it must stay and the elevator continues to "go up" to the next target floor.

Each customer has been given his target floor, and he/she should exit accordingly, the customers' target exit floors are unknown to the robot. In case the robot has to exit with a customer, it must negotiate with the customer in the elevator who is going first. The door may open several times before reaching the robot's target exit floor.

Moving to exit (Phase 3)

The robot moves from the elevator to the finish area (exit).

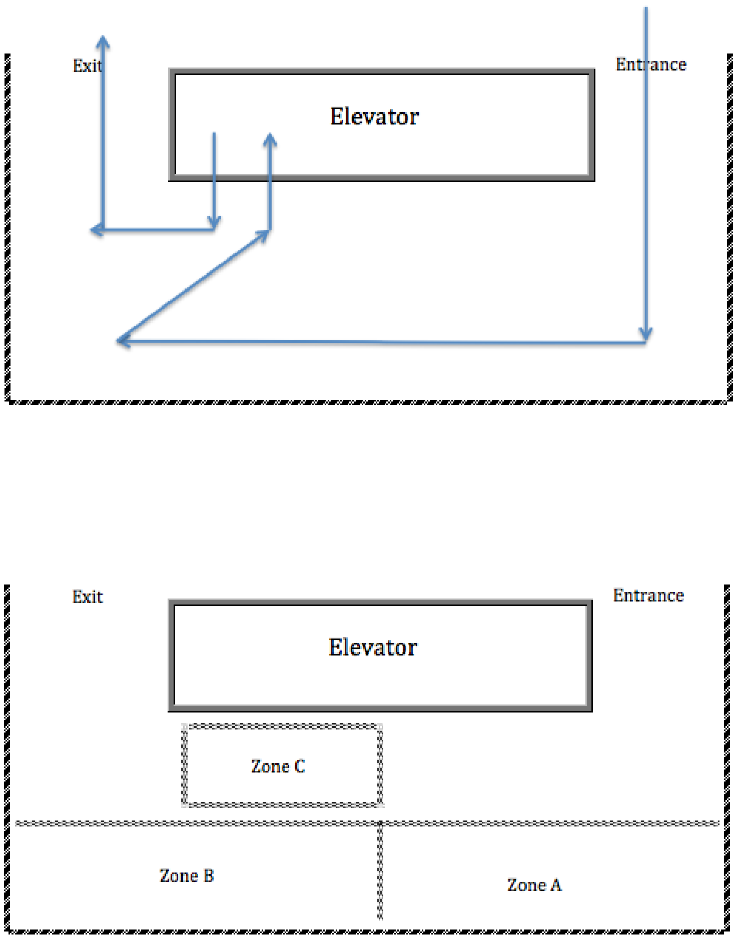


Figure 6.2: Layout of E4 task scenario

6.3 HRI confirmatory research study in E4

6.3.1 Defining research questions in E4

To address the issues concerning *gender effect* and *users' role effect*, we conduct an HRI confirmatory research study by introducing the experiment studies in E4 task scenario. *Gender effect* has been studied by several HRI researchers (see in Chapter 2), whereas *users' role effect* has been studied few times by HRI researchers. For instance, Koulouri et al. [91] present an experiment study in HRI changing the combination of gender and role of users (i.e., instructor or follower), finding that male users/instructors and female "robots" followers are associated with the fastest and most accurate completion of the navigation tasks. Hüttenrauch and Severinson-Eklundh [76] investigate the willingness to help in HRI when the users are bystanders.

In this chapter, we focus on *human factors* aiming at validating the research hypotheses in following aspects: *Robots' behaviors perceived by users*, *gender effect*, *users' role effect*. To be more specific, we are interesting in validating the research hypotheses as follows:

- The robot's behavior perceived by users is influenced by users' gender.
- The robot's behavior perceived by users is influenced by users' role.
- There is a difference on robot's behaviors perceived by users among different robots developed by different participating teams.

6.3.2 Determining *dependent factors* and *independent factors* in E4

We designed our HRI confirmatory research study by first carefully selecting the *dependent factors* to be measured and the *independent factors* to manipulate for producing many conditions for comparison. In E4, robot's behaviors perceived by human can be represented by three aspects: *social behavior of robot*, *proxemics between human and robot* and *collaboration behavior of robot*. We identified 17 dependent factors to be considered, reflecting robot's behaviors in E4 (see Chapter 4). We identified two of independent factors: *user's gender* and *user's role*, in particular, user's gender assumes two values *male* and *female*, users' role assumes three values *role A*, *role B* and *role C*.

6.3.3 Experiment settings in E4

The experiment setting in E4 is the same explained in Chapter 5. The novelty of the experiment setting in E4 is that we set out the ethical criteria by analyzing ethical and social concerns. It is in the interest of HRI researchers to take ownership of HRI ethics issues and to make attention to those issues a routine aspect of their everyday work. A culture of ethical awareness and sophistication within the HRI community will, thus, advantage the cause of HRI research, development, and marketing [138].

6.4 Data Analysis and Results

6.4.1 Analyzing robot's behaviors perceived by users among different robots developed by different participating teams

We first completed the missing data using mean imputation method, and then conducted Repeated Measures ANOVA (RMANOVA) to check how users perceived the robot's behaviors among different robots developed by different participating teams. We found statistically significant differences on all the robot behaviors (i.e., p values of all the robot behaviors are less than 0,05). In the end, we conducted pairwise comparison to check how users perceived differently between two participated teams. The results of significant values and mean difference values of pairwise comparison are shown in Tables 6.1, 6.2 and 6.3.

Table 6.1: Pairwise Comparison Between Participated Teams in Social Behavior

Social Behavior				
	Team (i)	Team (j)	Mean difference (i-j)	P-value
Perceived Happiness	eNTiTy	UC3M	0,897	**0,000**
	HEARTS	UC3M	0,897	**0,000**
	eNTiTy	LASR	0,573	0,029
	eNTiTy	Gentlebots	1,269	**0,000**
	HEARTS	LASR	0,573	0,014
	LASR	Gentlebots	0,688	0,010
	HEARTS	Gentlebots	1,260	**0,000**
Perceived Sociability	eNTiTy	UC3M	0,783	**0,001**
	HEARTS	UC3M	0,639	**0,000**
	eNTiTy	Gentlebots	1,210	**0,000**
	HEARTS	Gentlebots	1,066	**0,000**
Perceived Capability	eNTiTy	UC3M	0,463	0,041
Perceived Responsiveness	HEARTS	UC3M	0,417	0,035
	UC3M	Gentlebots	0,479	0,028
	eNTiTy	Gentlebots	0,813	**0,000**
	LASR	Gentlebots	0,631	0,044
	HEARTS	Gentlebots	0,896	**0,000**
Perceived Interactiveness	HEARTS	UC3M	0,480	0,016
	HEARTS	Gentlebots	0,733	**0,000**
***Perceived Naturalness**	eNTiTy	UC3M	0,306	0,037
	HEARTS	UC3M	0,835	**0,000**
	UC3M	Gentlebots	0,352	0,048
	HEARTS	eNTiTy	0,529	0,048
	eNTiTy	Gentlebots	0,658	**0,000**
	HEARTS	LASR	0,528	**0,007**
	LASR	Gentlebots	0,659	**0,007**
	HEARTS	Gentlebots	1,188	**0,000**

For the items related to *Social Behavior of the robot*, *Proxemics between human and robot*, *Collaboration with robot*, we found remarkable and statistically significant differences in pairs comparing with UC3M team and Gentlebots team, meaning that users perceived less sociable, less suitable in proxemics and less collaborative of robot's behaviors of UC3M team and Gentlebots team. Hence, we confirmed our research hypothesis and conclude that UC3M team and Gentlebots team make a difference on robot's behaviors perceived by users.

This was an interesting finding because of Gentlebots team had done very well in performing task E4, meaning that robot's behaviors perceived by users were not total depending on task performances. This finding should be noticed by HRI researchers and robot engineers. Social HRI shall have its particular importance in the new era of developing social robot technology.

6.4.2 Analyzing gender effect

We first completed the missing data using mean imputation method, and then conducted Mixed-ANOVA to check how male and female users perceived differently the robot behavior. We found no *interaction effect* between with-in subject factor (i.e., teams) and between-subject factor (i.e., gender), meaning that the impact of between-subject factor does not depend on the level of with-in

Table 6.2: Pairwise Comparison Between Participated Teams in Proxemics

Proxemics between human and robot				
	Team (i)	Team (j)	Mean difference (i-j)	P-value
Did robot look at your face ...	eNTiTy	UC3M	0,900	**0,000**
	LASR	UC3M	0,550	**0,003**
	HEARTS	UC3M	0,792	**0,000**
	eNTiTy	Gentlebots	1,160	**0,000**
	LASR	Gentlebots	0,810	**0,003**
	Hearts	Gentlebots	1,052	**0,000**
Did you look at the robot's face ...	eNTiTy	UC3M	0,992	**0,000**
	LASR	UC3M	0,900	**0,000**
	HEARTS	UC3M	0,700	**0,000**
	Gentlebots	UC3M	0,658	**0,000**
Have you paid attention to con ...	eNTiTy	UC3M	0,633	**0,000**
	LASR	UC3M	0,400	**0,000**
	HEARTS	UC3M	0,371	**0,005**
	eNTiTy	Gentlebots	0,500	**0,003**
Have you understood well con ...	eNTiTy	UC3M	0,538	**0,001**
	UC3M	Gentlebots	0,409	0,034
	eNTiTy	HEARTS	0,468	**0,001**
	eNTiTy	Gentlebots	0,947	**0,000**
	LASR	Gentlebots	0,758	**0,001**
	HEARTS	Gentlebots	0,479	0,011
Perceived Consciousness	eNTiTy	UC3M	0,639	**0,000**
	LASR	UC3M	0,447	0,018
	HEARTS	UC3M	0,792	**0,000**
	eNTiTy	Gentlebots	0,660	**0,001**
	HEARTS	Gentlebots	0,813	**0,000**
Perceived Friendliness	eNTiTy	UC3M	1,454	**0,000**
	LASR	UC3M	1,000	**0,000**
	HEARTS	UC3M	1,450	**0,000**
	eNTiTy	Gentlebots	1,575	**0,000**
	LASR	Gentlebots	1,120	**0,000**
	HEARTS	LASR	0,450	0,050
	HEARTS	Gentlebots	1,570	**0,000**
Perceived Politeness	eNTiTy	UC3M	1,050	**0,000**
	HEARTS	UC3M	1,042	**0,000**
	UC3M	Gentlebots	0,548	**0,001**
	eNTiTy	LASR	0,556	0,019
	eNTiTy	Gentlebots	1,598	**0,000**
	HEARTS	LASR	0,548	0,010
	LASR	Gentlebots	1,041	**0,000**
	HEARTS	Gentlebots	1,589	**0,000**
Perceived Adaptability	eNTiTy	UC3M	0,800	**0,000**
	LASR	UC3M	0,551	**0,002**
	HEARTS	UC3M	0,867	**0,000**
	eNTiTy	Gentlebots	0,598	0,011
	HEARTS	Gentlebots	0,664	**0,000**
Perceived Ease of use	eNTiTy	UC3M	0,696	**0,000**
	LASR	UC3M	0,603	**0,000**
	HEARTS	UC3M	0,655	**0,000**
	eNTiTy	Gentlebots	0,706	**0,001**
	LASR	Gentlebots	0,612	0,020
	HEARTS	Gentlebots	0,664	**0,001**

Table 6.3: Pairwise Comparison Between Participated Teams in Collaboration

Collaboration with robot				
	Team (i)	Team (j)	Mean difference (i-j)	P-value
Perceived Enjoyment	eNTiTy	UC3M	1,000	**0,000**
	LASR	UC3M	0,879	**0,000**
	HEARTS	UC3M	1,262	**0,000**
	eNTiTy	Gentlebots	0,845	**0,000**
	LASR	Gentlebots	0,724	**0,002**
	HEARTS	Gentlebots	1,107	**0,000**
Perceived Collaborativeness	eNTiTy	UC3M	1,038	**0,000**
	LASR	UC3M	0,990	**0,000**
	HEARTS	UC3M	1,089	**0,000**
	eNTiTy	Gentlebots	0,768	**0,000**
	LASR	Gentlebots	0,720	**0,002**
	HEARTS	Gentlebots	0,819	**0,000**

subject factor. However, we found highly significant difference of *main effect* among the with-in subject factor, meaning that the overall effect over with-in subjective effects is statistically significant, and this finding has been further discussed (see Subsection 6.4.1).

For the following items related to *Social Behavior* of the robot: *Perceived Responsiveness (p=0.02)*, *Perceived Interactiveness (p=0.03)* and *Perceived Naturalness (p=0.019)*, we found significant differences between female and male users, meaning that female users perceived the robot's behaviour more positively than male users. No other significant difference of between factor has been found in this analysis study. As a consequence, we can partially confirm the validity of our research hypothesis: Only the social behaviors of robots, i.e., *Perceived Responsiveness*, *Perceived Interactiveness* and *Perceived Naturalness* are influenced by users' gender. Consequently, the designers of social robots should make sure that the interaction style of the robot fits the user's gender and the users' individual attributes.

6.4.3 Analyzing users' role effect

As Scholtz [149] emphasizes the importance of human roles in HRI, in E4, we are interesting in validating the research hypothesis on *users' role effect*.

We first completed the missing data using mean imputation method, and then conducted Mixed-ANOVA to check how role A, role B and role C perceived differently the robot's behavior. We found no *interaction effect* between with-in subject factor (i.e., teams) and between-subject factor (i.e., role), meaning that the impact of between-subject factor does not depend on the level of with-in subject factor. However, we found highly significant difference of *main effect* among the with-in subject factor as well, and this finding has been further discussed (see Subsection 6.4.1).

For the following items related to *Proxemics between human and robot*: *Have you paid attention to the conversation with the robot?* and *Have you understood well the meaning of conversation?*, we found significant differences between roles. Furthermore, we conducted pairwise comparisons to check the effect among role A, role B and role C, we found remarkable significant difference between Role C and Role A in *Have you paid attention to the conversation with the robot?*(p=0,008), significant difference between role C and role A in *Have you understood well the meaning of conver-*

sation?(p=0,039) and significant difference between role C and role B in *Have you understood well the meaning of conversation?*(P=0,016). The finding means users' role make a difference on behaviors highly related to spoken languages or dialogues. As a consequence, we can partially confirm the validity of our research hypothesis: Only the behaviors that highly related to spoken language or dialogues are influence by users' role.

6.5 Discussion and Conclusion

In this chapter, we present a new approach for designing task scenarios in a social context. We exploit the most sociable episode of the SciRoc challenge "Take the elevator (E4)" to implement our approach for task scenario design. The approach proposed consists two steps: (i) analyzing the elements of the task scenario; (ii) sketching out the layout of the task scenario.

For the human-centered view, HRI researchers can start by brainstorming possible interaction scenarios which may happen in regards to a person and the particular robot or interface [190]. HRI researchers can implement the interaction scenarios by adopting our approach for designing the concrete task scenario.

Moreover, we present an HRI confirmatory research study by conducting the experiment studies in E4. We got some interesting findings concerning *gender effect*, *users' role effect* and *robot's behaviors of different participating teams perceived by users.* Designers use form to balance the needs of people, the capabilities of technology and the context of users when they introduce a single product [15]. For instance, the results of users' role analysis confirm that users' role affect the spoken behaviors of robot perceived by users in social contexts, HRI designers should "involve" these findings in design process [75].

While drawing up the scoring system of a robot competition in social contexts, we should establish the rules that encourage researchers for addressing *human behaviors toward robot* and *robot's behaviors perceived by human* in the field of social HRI.

Chapter 7

Toward an HRI mixed-method research approach

7.1 Introduction

As a community, we recognise that the technical fields of engineering, control theory and computer science do not provide necessary tools for the scientific research on the *human* and *interaction* parts of HRI [78]. For this reason, we take inspiration and ground research methodologies in establishing results from the social science, social psychology, and sociology. Hence, this thesis takes users' perspectives and aims at understanding *human factors*, *human behaviors toward robot* and *robot's behaviors perceived by human* in the field of HRI.

"A theory can be proved by experiments, but no path leads from experiment to the birth of a theory". This is a well famous science quote by Albert Einstein. For one thing, exploratory research is about putting one's self deliberately in a place - again and again - where discovery is possible and broad, usually non specialized interests can be pursued, exploratory research requires lengthy periods of fieldwork (of various kinds) and the sort of personal concern and long standing interest in a topical area that sustains such fieldwork [157].

HRI researchers usually approach research problems by conducting *exploratory research study (i.e., qualitative approach)* and *confirmatory research study (i.e., quantitative approach)*.

The qualitative approach is used for exploratory research study in HRI, it is the approach that allows you to examine people's experiences in detail by using a specific set of research methods such as in-depth interviews, focus group discussion, observation, content analysis, visual methods and so on [70]. Hoffman and Zhao [72] believe that qualitative research methods have been much underutilized in HRI research and they encourage researchers to familiarize themselves with these methods by seeking out other sources.

For another thing, quantitative approach is used for confirmatory research study in HRI, it is an approach that provides deeper, valid and relevant findings to confirm people's experiences and expectations in detail by using query techniques and observational techniques. These non-standardised quantitative approaches are often used in the field of HRI by running experimental studies and imposing experimental conditions and expectations.

Similar to discussions within psychology on what the "right approach" is, HRI researchers often involve discussions on quantitative, large scale studies versus qualitative small scale studies based on case studies. In order to advance the field of HRI research, it is important to 'make peace' between proponents of one or the other methodology, acknowledging that there are several paths one can take in order to illuminate the issues under investigation [42].

There are many contexts of social science and psychology where qualitative and quantitative methods can be used in conjunction to build and refine theory [55, 80, 181]. Shah and Corley [150] motivate researchers to engage in this practice – to use both qualitative and quantitative empirical methods to fully understand their phenomenon of interest – or at least to convince quantitative researchers to draw insights from qualitative research in their area, and vice versa. Psychology and cognitive engineering topics have tools and methods to measure HRI, joint attention theory, teams, and user-centered design. For instance, Giusti and Marti [61] carry out a qualitative and quantitative speech and behavior analysis in a social HRI context, Schillaci et al. [148] adopt two techniques for evaluating saliency detection and attention manipulation mechanisms in HRI: user experience as measured by qualitative and quantitative questions in questionnaires and proxemics estimated from recorded videos of the interaction, Syrdal et al. [163] explore qualitative research methods of video prototyping in HRI in order to evaluate user experience of prototype systems by interviewing, the results were used to inform the general outline of quantitative questionnaires for each video. Damholdt et al. [41] confirm that qualitative and quantitative methods can be used in unison in HRI research to achieve more fine-grained analyses of relevant experience.

However, HRI researchers note that there was no concise list of qualitative and quantitative evaluation methods or tools, nor was there a clear mapping of particular techniques to desired outcomes [110].

In this chapter, we improve the existing research approaches and methodologies and find complementary quantitative as well as qualitative approaches towards HRI studies aiming at contributing more reliable research results. In particular, we first discuss *exploratory research study (i.e., qualitative approach)* and *confirmatory research study (i.e., quantitative approach)* in the field of social HRI, in particular, we intend to take these arguments further. We end this chapter by proposing a new HRI qualitative and quantitative (i.e., mixed-method) research approach in this field.

7.2 Qualitative approach for exploratory research study

As mentioned the differences between qualitative approach and quantitative approach in Section 2.2.5, we decide to introduce a qualitative approach for exploratory research study that adopts subjective measurements in order to measure beliefs and desires of HRI experts (see Fig. 7.1).

It begins with:

- Selecting representative samples: *type* and *number* of the participants selected for exploratory research study (i.e., qualitative approach) are highly dependent on the qualitative technique chosen. In our approach, we select *HRI experts* as participants. Since we adopt subjective measurements as qualitative evaluation techniques for qualitative approach, HRI experts' responses can avoid the influencing by possibly technical incompetence or impairment.

- Determining the duration of the experimental trial: *short-term*, *mid-short term*, *mid-long term*, and *long-term* are often used in an experimental study. In the context of social HRI, long-term experimental can encounter some unforeseen impairments due to technological failures or personal inconveniences. We suggest that HRI researchers conduct *short-term* experimental trial as first choice as it guarantees higher efficiency.

- Introducing legal ethical and social considerations: as robotics technology forays into our daily lives, research, industry, and government professionals in the field of HRI in must grapple with significant ethical, legal, and normative questions [140]. The use of robots in social context poses unanticipated risks and ethical and social problems. Two main areas of potential ethical risk are considered here: children and elder [122]. The ethical discussion of ownership of human biological materials has two major aspects. The most prominent argument revolves around the effects such ownership would have on our views of humans, on whether humans would become commodities. The second argument, raised clearly by John Moore's suit, is the fairness of the distribution of benefits: if any profit is made as a result of research with human biological materials, how should it be shared between the source of the materials and the researchers [67]. Hence, HRI researchers should check the following two rules before any exploratory research study: *Consent form*, any research study that involves human participants, whether exploratory or confirmatory, qualitative or quantitative, online or in person, requires participants' informed consent before the research is started [14]. *Ethical, legal and social regulators*, any research study that involves human participants, whether exploratory or confirmatory, qualitative or quantitative, requires that the experiments comply with the requirements of local and relevant overseas regulators in this regard, comply with local social customs and culture as well.

  It is worth to notice that Riek and Howard [138] present a preliminary sketch of HRI ethical principles. It can be used as a preliminary check, whether experimental settings are in compliance with ethics regulations or not. Otherwise, ethics committee of HRI research institutions may provide more detailed feedback regarding your specific research experiment. Note that ethics approval is a requirement for publication in many scientific journals as well [14].

- Selecting a robot and implementing the robot's functionalities: the robot should be selected according to the task scenario. In the context of social HRI, robots are often humanoid robots or service robots (see in Chapter 2.1).

- Conducing the experiment in the realistic environment: as we mentioned in Chapter 2, the experiment environment and the experiment settings should reflect realistically the application domain and the situations that participants would likely to be encountered. For this reason, we create the realistic environment to conduct experiments.

- Subjective measurements: *focus group*, *interview* and *questionnaire* are subjective measurement techniques for conducting experimental studies, where *focus group* and *interview* are often used for qualitative techniques when small size of samples are involved as participants. The selection of measurement technique is highly dependent on *sample size* of representative samples. Since experiments of qualitative approaches require HRI experts (i.e., small size), *interview* and *focus group* techniques are recommended.

- Analysing the qualitative data: as we mentioned in Section 2.2.5, researchers use *content analysis*, *thematic analysis*, *text analysis* and *discourse analysis* to analyze qualitative data.

- Determining the HRI research statement: we declare the HRI research statements according to our relevant findings.

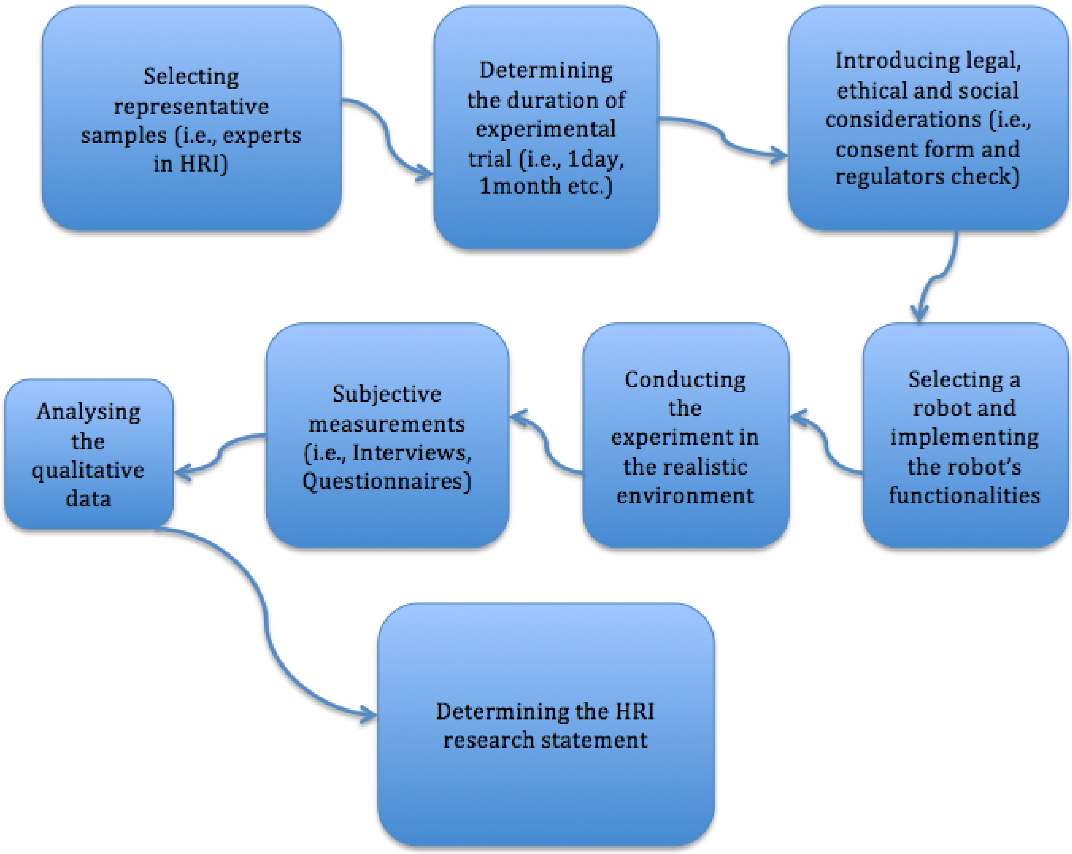


Figure 7.1: HRI qualitative research approach with experts for exploratory research study

7.3 Quantitative approach for confirmatory research study

In [173], the authors define *confirmatory research studies* as *researchers preregister their studies and indicate in advance the analyses they intend to conduct. Only these analyses deserve the label "confirmatory," and only for these analyses are the common statistical tests valid.* Other analyses can be carried out but these should be labeled "exploratory". Exploratory research explore patterns with no a-priori articulated hypotheses, whereas confirmatory research that explicitly tests a-priori formulated hypotheses [119]. Therefore, to complement qualitative approach, we decide to introduce a quantitative approach for confirmatory research study that adopts objective measurements in order to measure behaviors of experimental participants and confirm the HRI exploratory statements (see Fig. 7.2).

It begins with:

- Defining a research question: before any experiment concerning confirmatory research study (i.e., quantitative approach), HRI researchers should formulate a research hypothesis, analyze dependent and independent variables and establish relationships between variables. Since qualitative and quantitative methods can be used in sequential form, the research question of quantitative approaches can be established on the basis of the HRI research statements of qualitative approaches.

- Selecting representative samples: large-scale of participants are required in quantitative approach. In the context of social HRI, we suggest to use real users (i.e., users randomly selected in the experimental field) as participants.

- Determining the experimental design: *between-subjects*, *within-subjects* and *mixed-factorial design* are often used in the field of HRI [22, 115]. In psychology that make comparisons using either within or between designs (or both) that sometimes yield the same results and sometimes do not [36]. A between-subjects experiment would remove learning effect for individual participants, but it would be more difficult to control for variation in learning style between participants. On balance, therefore, a within-subjects design is preferred, with order of presentation controlled [51]. A mixed factorial design involves two or more independent variables, of which at least one is a within-subjects factor and at least one is a between-groups factor.

- Determining the duration of experimental trial: as we mentioned in the previous section, we suggest to conduct *short-term* experiment as first choice.

- Introducing legal, ethical and social considerations: as we mentioned the importance of legal and ethical issues in social HRI, we suggest to impose the same regulations as described in the previous section.

- Selecting a robot and implement the robot's functionalities: the robot selected is highly dependent on the task scenario.

- Objective measurements: objective measurements are behavioral indicators you can measure independently of people's stated opinion. Objective measures include a broader range of possibilities: reaction speed, physiological reactions and so on [72]. *Observation* of HRI in naturally occurring social situations can be used to create new theoretical and practical models of appropriate social robot behavior and design. Along with the promise of novel technical challenges that can lead to developments in and beyond conventional approaches to robotics, social robotics has potential for advancing our knowledge about human society by observing how people explore and interact with social robots in non-laboratory social environments [144]. Hence, we suggest HRI researchers to adopt *observation techniques by using camera or sensors* as evaluation techniques.

- Conducting the experiment in the real environment: to complement qualitative approach, we suggest to conduct our experiment and impose our experimental settings in a real environment in social contexts.

- Collecting the quantitative data: we collect quantitative data from cameras and sensors.

- Data analysis: statistical method and machine learning method can be used for analyzing quantitative data.
- Hypothesis validation and reporting findings: we conclude by validating the proposed hypotheses and reporting the relevant findings.

7.4 HRI qualitative and quantitative (mixed-method) research approach

The specific mixed methods approaches are defined by the ordering of the application of the quantitative and qualitative methods (simultaneously or sequentially), as well as at what point the mixing of the methods occurs. Qualitative and quantitative data collection can occur in parallel form or sequential form [106].

In the context of social HRI, several researchers have proposed mixed-method for conducting the HRI experiments. For instance, Dautenhahn and Werry [45] present a technique for quantitatively and qualitatively describing and analysing HRI in terms of low-level behavioural criteria (so-called micro-behaviours). In this section, we aim at improving the existing research frameworks and proposing a new HRI qualitative and quantitative (i.e., mixed-method) research approach.

The proposed approach, HRI qualitative and quantitative (i.e., mixed-method) research approach, is developed base on three steps, as shown in Fig. 7.3. First step, we rely on HRI task scenario development, which analyzes the elements of task scenarios and makes the HRI task scenario layout available (e.g., see Chapter 6). Second step, we conduct the experiment in the realistic environment by conducting the exploratory research study (i.e., qualitative approach) with HRI experts. To validate the HRI research statements obtained from exploratory research study, we conduct by the end the experiment in the real environment by conducting confirmatory research study (i.e., quantitative approach) with large samples (i.e., real users). The content is described as follows:

1. First step, *development of HRI task scenario*, by analyzing the elements of the task scenario and sketching out the layout of the task scenario.
2. Second step, *exploratory research study (i.e., qualitative approach)*, by conducting the experiment in the realistic environment with HRI experts according to the task scenario.
3. Third step, *confirmatory research study (i.e., quantitative approach)*, by conducting the experiment in the real environment with real users according to the task scenario.

Moreover, we consider our approach for HRI qualitative and quantitative (mixed-method) as a first small step towards HRI mixed research approach in HRI social contexts. The novelties of our approach proposed are:

- *Task scenario analysis*, our approach analyzes task scenarios in advance, layouts of task scenarios are useful for implementing exploratory and confirmatory studies.
- *Legal, ethical and social norms and regulators*, our approach emphasizes the importance of legal and ethical considerations when conducting HRI experiments.

- *Representative samples*, our approach involves both HRI experts and real users as participants when conducting HRI experiments.
- *Realistic and real environment*, our approach explores the possible research statements in the realistic environment and validates the findings in the real environment.
- *Qualitative data and quantitative data*, our approach collects qualitative data and quantitative data in order to analyze human behaviors toward robots and robot's behaviors perceived by human.
- *Subjective measurements and Objective measurements*, our approach complement subjective measurements and objective measurements.
- *Qualitative approach and quantitative approach*, our mixed-method research approach conjuncts qualitative approach and quantitative approach in sequential form, i.e., the HRI research statement of qualitative approach is the research question to be validated in the context of quantitative approach.

The improved mixed-method approach can give us a complete view on how to perform an HRI experiment in social contexts, and we believe the findings obtained by our mixed-method approach are *reliable*.

7.5 Discussion and Conclusion

In this chapter, we first propose a qualitative approach for exploratory research study in the field of social HRI and discuss the main concerns of qualitative approach in exploratory research study as follows: *selecting representative samples*, *ethical, legal and social considerations* and *subjective measurements*. Secondly, we propose a quantitative approach for confirmatory research study in the field of social HRI and discuss the main concerns of quantitative approach in confirmatory research study as follows: *formulating a research hypothesis*, *objective measurements*. We end this chapter by proposing a new qualitative and quantitative (mixed-method) research approach, and listing the improvements comparing to the existing HRI research frameworks.

Conducting HRI experiments is challenging, we believe the improvement of our research approach is a small step toward *human behaviors toward robot* and *robot's behaviors perceived by human* research topics. The new research approach examines structure, concepts, and procedures, with an emphasis on human factors in the field of social HRI. The conditional and dynamic procedure of approach should be discussed in-depth in the future, as well as multiple elements of approach along with quantitative and qualitative (mixed-method) research approach validation should be discussed.

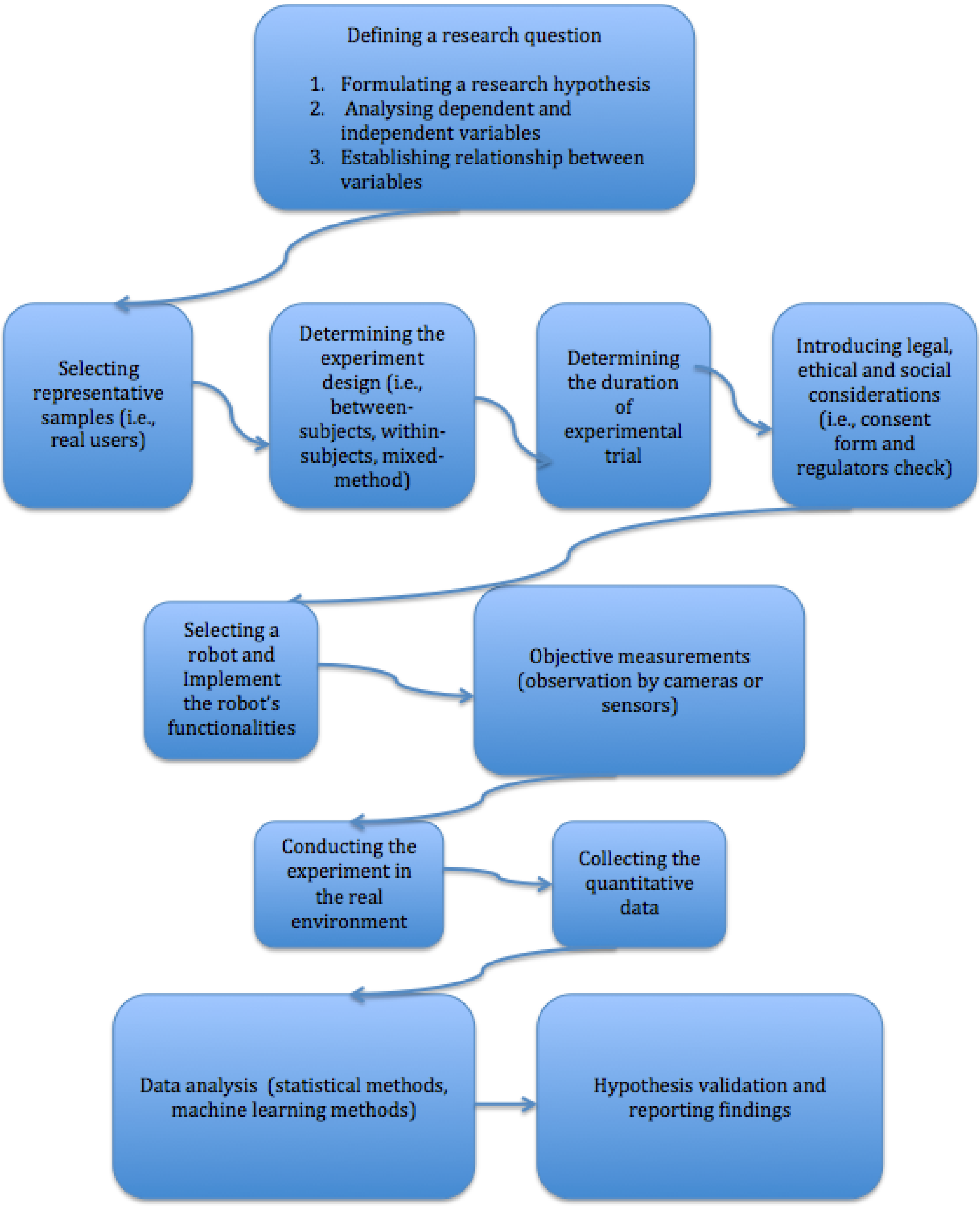


Figure 7.2: HRI quantitative research approach for confirmatory research study

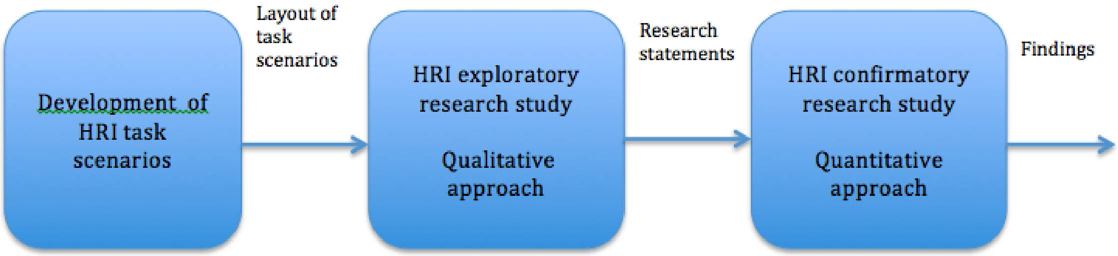


Figure 7.3: HRI qualitative and quantitative (mixed-method) research approach

Part III: Conclusion

Chapter 8

Discussion and Conclusion

8.1 Conclusion

HRI, a cross-discipline, encompasses a broad range of disciplines, as *robotics*, *social science*, *artificial intelligence (AI)*, *robotics*, *psychology*, *human-computer interaction (HCI)*, etc. It is argued that current research efforts and directions are not sufficient in HRI research, and that future research needs to further address interdisciplinary research in order to achieve long-term success of socially interactive robots [4]. To fulfill this gap, we try to bring together the multi-discipline to a unique research area - *Social HRI*.

Social HRI research aims to understand *interaction* and *interactivity* between human and robot in social contexts. Dautenhahn [43] presents three different ways on HRI: robot cognition-centred view, robot-centred view and human-centred view (see Fig. 8.1). For this purpose, our thesis takes the perspective of *human-centred view*, in particular, *human behaviors toward robot* and *robot's behavior perceived by human* and investigates in-depth HRI research methodologies in social contexts.

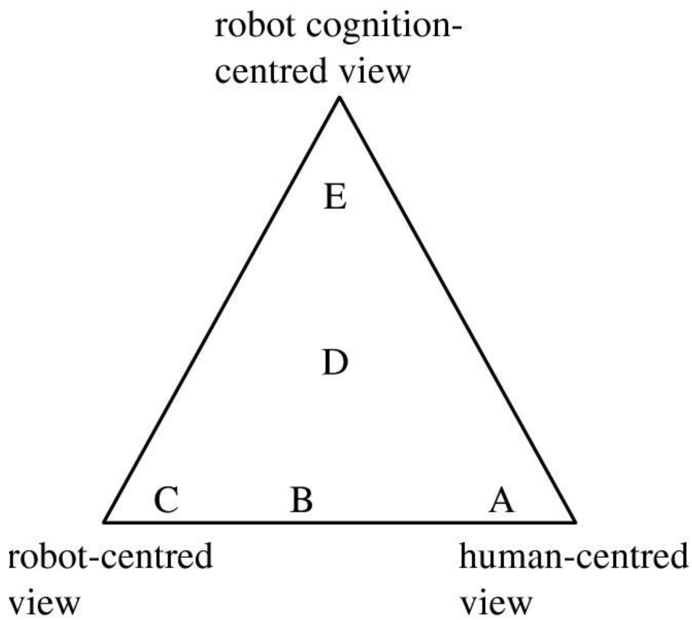


Figure 8.1: Different view on social HRI [43]

Letters presented in Figure 8.1: A, socially evocative; B, socially situated; C, sociable; D, socially intelligent; E, socially interactive

8.1.1 Contributions in HRI research methodologies

We start by reviewing some relevant research methodological approaches in the field of multi-discipline, and try to improve these methodologies by conducting some research studies in HRI. In this thesis, we present a list of research studies :

- HRI confirmatory research study in a social context in order to investigate if user perception may be affected by user's gender. Leveraging on the standardized questionnaire - Robot Social Attribute Scale (RoSAS), we analyze the quantitative data collected by the questionnaire and reported interesting findings.
- HRI confirmatory research study in the realistic context of the SciRoc challenge to investigate gender effect, user's role effect, and robot's behaviors perceived by users among different robots developed by different participating teams. Leveraging on the dedicated questionnaire designed for the tested scenario, we collect quantitative data and reported the findings.
- HRI exploratory research study in the realistic context of the SciRoc challenge in order to explore the correlation relationship between task performances evaluated by scoring and robot's behaviors perceived by participants. Leveraging on the same questionnaire designed for the test scenario and the task performance scoring, we collect qualitative data and reported the findings.

To improve the existing research methodological approaches, we added *participant recruitment consideration*, *ethical consideration*, *environment consideration*, *robot's autonomy consideration* to procedures of our research methodological approaches.

Questionnaire is the most common evaluation technique for collecting data from experiments. However, there are few standardized questionnaires for HRI research. In this thesis, we present an approach to develop a new type of questionnaire as a task-driven evaluation technique. Consequently, we validate the reliability of the questionnaire developed. The approach proposed is easy to replicate and reproduce.

Moreover, we introduce a novel approach for designing of task scenario which consists two steps, then we exploit the approach proposed for designing task scenario "take the elevator" of the SciRoc challenge.

We end this thesis by improving *qualitative research approaches for exploratory research study*, *quantitative research approach for confirmatory research study*, and *qualitative and quantitative (mixed-method) research approach* in the field of social HRI.

8.1.2 Contributions in scientific competitions

SciRoc Challenge is a repeatable and general-purpose test method (benchmark) developed for HRI performance evaluating investigating users' attitudes using HRI research methodologies. The challenge aims to introduce the robots in the European Robotics League (ERL) Smart Cities, in order to show how robots will integrate in the social context of the future as physical agents living in smart cities. On basis of the robots' functionalities developed by participant teams, robots autonomously cooperated with the simulated smart city (i.e., realistic environment) infrastructure and interacted with its citizens (i.e., users selected by SciRoc organization), accomplishing tasks such as assisting customers, providing professional services and supporting during emergency situations. To create the realistic environment, we provided a simulated smart city in the shopping mall of Milton Keynes which includes data acquisition system, data management system and physical communication infrastructure.

The key novelty of the first SciRoc challenge was the introduction of a novel concept for designing, implementing, and evaluating HRI tasks based on scientific competitions. For instance, we concretely enacted HRI experimental studies to investigate users' attitudes as well as develop the HRI task scenario "take the elevator". Involving *real users* in the task scenario had relevant impacts on scientific findings as well as it could promote dissemination of HRI scientific research, and improve the visibility of AI and robotics technologies for audience.

Our colleagues from the university of London mentioned *"Doing research just inside a lab, you forget about how it's going to be used, and how people are going to react, and think about it, and that's so central to whether this technology, robotics, is going to be acceptable and be useful to society."* In the context of the first SciRoc challenge, developers of robotic technology and HRI researchers had a good opportunity for users' engagements which were usually missing.

Experimental studies conducted in the first SciRoc challenge bring a lot of value to scientific community, they fulfilled with the multi-discipline sectors where we were going to develop such as robotic technology, artificial intelligence (AI), communication technology in 4G or 5G, HRI, interaction design, etc. Moreover, we discussed the scoring system in robotic competitions of HRI by establishing the rules that encouraged developers to study in-depth *human factors*, *human behaviors toward robot*, and *robot's behaviors perceived by human* in the field of social HRI.

8.1.3 Lessons learned

In the field of HRI, researchers have attempted to develop the robot's functionalities and the interactions between human and robot, but few have attempted to characterize its research problems as cross-discipline problems, and few have explicitly developed an entire research approach taking the perspective of users in social contexts. Hence, we try to improve the existing research frameworks in the field and attempt to propose a new HRI mixed-method research approach. In this way, We hope we might gain a broad perspective of the multi-discipline in our thesis by explaining the characteristics of research studies. We learn some lessons that we believe could be useful for other HRI researchers:

- *Participants*, Ju and Takayama reveal that the experts are positive about robots performing highly compensated jobs but have concerns about the robots' sensory or physical capabilities [81]. The non-experts, on the other hand, are interested in having robots undertake social tasks like responding to people or answering the phones. This study confirms that the experts view differently from non-experts. For this purpose, our HRI qualitative and quantitative (i.e., mixed-method) research approach should involve both HRI experts and non-experts (i.e., possibly real users) in HRI research studies.
- *Evaluation techniques*, the responses to subjective measurement (i.e., interview or other evaluation technique) may influenced by the mood and state of mind on the day of the study. For this reason, our HRI qualitative and quantitative (mixed-method) research approach consists two types of evaluation techniques, *subjective measurement* and *Objective measurement.* Subjective measurements are self-reported attitudes, thoughts, emotions and moods of participants, collected through participants' verbal responses. objective measurements can measure

independently of person's stated opinion [72].

- *Task scenario*, refers to the tasks performed by a person or a group of person in relation to a social robot or a group of social robots. HRI researchers collect the essential information by analyzing the elements of task scenarios. For this purpose, the experiments, either exploratory research study or confirmatory research study, should be conducted on basis of the analysis of task scenarios.

8.2 Limitations

In this section, we first discuss an ambiguous research problem regarding *evaluation technique.* since comparing to observation techniques, *questionnaires* can be used to reach a wider participant group in a social context, it takes less time to administer, and it can be analyzed more rigorously [51]. *Questionnaire* is also the most adopted evaluation method in the field of social HRI to collect quantitative data. Hence, we can say that *questionnaire* is an evaluation technique with efficiency.

In theory, *questionnaire* is more subjective than objective as evaluation technique, i.e., it is a subjective evaluation technique restricted in the specific task scenario. HRI researchers emphasize the accuracy and consistency of the questionnaire proposed, but they have neglected the *effectiveness* of a questionnaire comparing to an *objective measurement.* Since an objective measurement is more convincing, interesting and generalizable to the real word [72], we need to discover the concise concepts regarding weaknesses and strengths between objective measurements and questionnaires.

Secondly, traditional confirmatory research study (i.e., hypothesis testing) relies on statistical methods such as p-values, bayesian methods, etc. HRI researchers should notice that descriptive statics of quantitative data are useful in some specific circumstances, since p-value or beyesian coefficient is not the unique criteria [72]. Our approach suggests to use *objective measurement* and *subjective measurement* to collect data. We hope that HRI researchers could "treat" data carefully by running data analysis with traditional approach and computational approach in order to avoid possibly missing of underlying findings.

In the end, the latest advances concerning robots in real-world environments have brought a series of challenges to be faced. In fact, taking robots out from the laboratory ecosystem is a complex task as it introduces new problems related to unpredictable scenarios, where it is necessary for the machines to act autonomously and irrespective of changes in their surroundings [8, 144]. Due to widely unforeseen variables in real-world environments, HRI researchers should consider all these variables in a practical way. Each variable presented in social contexts can be a research topic in the field of HRI. It is worth to notice that even fields like psychology do not have one agreed upon methodology in how to carry out all these variables. Therefore, HRI researchers should make "appropriate" assumptions in social contexts before any experimental study. These assumptions can influence the relevant findings, but HRI community have not a concise list regarding this concern.

8.3 Future work

Researchers in the area are increasingly aware that new methods, methodologies, and in general a theoretical and conceptual basis needs to be formed, if HRI is to establish itself as a research field along-side e.g. HCI or psychology. As a research field, HRI is still in its infancy [42]. Our research approach proposed is a small contribution concerning research methodologies and approaches in the field of HRI. However, we need to conduct numerous experiments in the field to validate reliability as well as replicability of our approach proposed.

Following the success of the first SciRoc challenge, held in Milton Keynes, UK in 2019, the second SciRoc event will settle in a new Smart City, Bologna Italy, in September 2021. Due to COVID-19 restrictions, the second SciRoc challenge propose a simulation environment (see Fig. 8.2) to design and implement HRI experiments. The goal of the simulation environment is to allow for the development of the whole architecture, main functionalities as person and object perception are simplified by the simulation set up. Real users will be recruited and they will interact through computer devices. The second SciRoc challenge offers a good opportunity to investigate *human behaviors toward robots through computer devices*, the relevant results should be useful for HRI communities.

In SciRoc 2 (Episode 1), the robot will assist the staff of a coffee shop to take care of their customers. The robot is required to recognise and report the status of all tables inside the shop, to take orders from customers and to deliver objects to and from the customers' tables.

The immediate future work is to adopt our research proposed to design and implement the task scenario both in the simulation context and in the realistic context.

As the field of social HRI, gains more visibility by researchers from multi-discipline. Any kinds of suggestions, contributions or even criticisms from the community can push toward understanding of *human behaviors toward robot* and *robot's behavior perceived by human*. We hope that this thesis can inspire HRI researchers to move forward standardization of metrics and measurements.

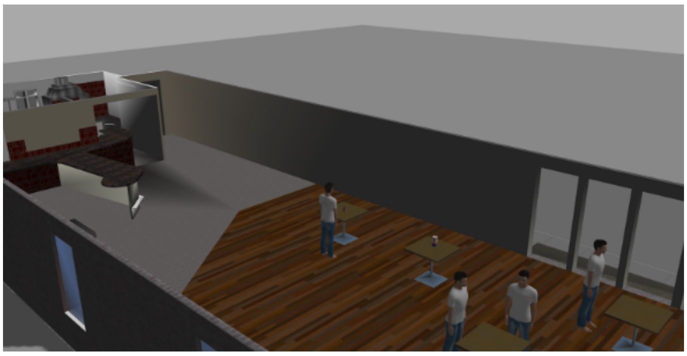


Figure 8.2: Simulated environment in the second SciRoc challenge (Episode 1, coffee shop)

Bibliography

[1] I. Aaltonen, A. Arvola, P. Heikkilä, and H. Lammi. Hello pepper, may i tickle you? children's and adults' responses to an entertainment robot at a shopping mall. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 53–54, 2017.

[2] R. Agrigoroaie, S.-D. Ciocirlan, and A. Tapus. In the wild hri scenario: influence of regulatory focus theory. *Frontiers in Robotics and AI*, 7, 2020.

[3] B.-K. Ahn, H. S. Ahn, C. Sutherland, J. Lim, and B. MacDonald. Development and evaluation for human-care scenario using social robots. In *Proceedings of HRI 2018 Workshop on Social Human-Robot Interaction of Human-Care Service Robots*, 2018.

[4] B. Alenljung, J. Lindblom, R. Andreasson, and T. Ziemke. User experience in social human-robot interaction. In *Rapid automation: Concepts, methodologies, tools, and applications*, pages 1468–1490. IGI Global, 2019.

[5] P. Alves-Oliveira, T. Ribeiro, S. Petisca, E. Di Tullio, F. S. Melo, and A. Paiva. An empathic robotic tutor for school classrooms: Considering expectation and satisfaction of children as end-users. In *International Conference on Social Robotics*, pages 21–30. Springer, 2015.

[6] F. Amigoni, E. Bastianelli, J. Berghofer, A. Bonarini, G. Fontana, N. Hochgeschwender, L. Iocchi, G. Kraetzschmar, P. Lima, M. Matteucci, et al. Competitions for benchmarking: Task and functionality scoring complete performance assessment. *IEEE Robotics & Automation Magazine*, 22(3):53–61, 2015.

[7] E. André, A. Paiva, J. Shah, and S. Šabanovic. Social agents for teamwork and group interactions (dagstuhl seminar 19411). In *Dagstuhl Reports*, volume 9. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2020.

[8] A. Andriella, C. Torras, and G. Alenya. Short-term human–robot interaction adaptability in real-world environments. *International Journal of Social Robotics*, pages 1–19, 2019.

[9] R. C. Arkin, R. C. Arkin, et al. *Behavior-based robotics*. MIT press, 1998.

[10] I. Asimov. Three laws of robotics. *Asimov, I. Runaround*, 1941.

[11] N. Aspragathos, V. Moulianitis, and P. Koustoumpardis. Special issue on human–robot interaction (hri). *Robotica*, 38(10):1715–1716, 2020.

[12] M. Bajones, A. Weiss, and M. Vincze. Help, anyone? a user study for modeling robotic behavior to mitigate malfunctions with the help of the user. *arXiv preprint arXiv:1606.02547*, 2016.

[13] A. Bannat, J. Gast, T. Rehrl, W. Rösel, G. Rigoll, and F. Wallhoff. A multimodal human-robot-interaction scenario: Working together with an industrial robot. In *International conference on human-computer interaction*, pages 303–311. Springer, 2009.

[14] C. Bartneck, T. Belpaeme, F. Eyssel, T. Kanda, M. Keijsers, and S. Šabanović. *Human-robot interaction: An introduction*. Cambridge University Press, 2020.

[15] C. Bartneck and J. Forlizzi. A design-centred framework for social human-robot interaction. In *RO-MAN 2004. 13th IEEE international workshop on robot and human interactive communication (IEEE Catalog No. 04TH8759)*, pages 591–594. IEEE, 2004.

[16] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, 1(1):71–81, 2009.

[17] C. Bartneck, T. Nomura, T. Kanda, T. Suzuki, and K. Kato. Cultural differences in attitudes towards robots. AISB, 2005.

[18] C. Bartneck, T. Suzuki, T. Kanda, and T. Nomura. The influence of people's culture and prior experiences with aibo on their attitude towards robots. *Ai & Society*, 21(1-2):217–230, 2007.

[19] P. Baxter, J. Kennedy, E. Senft, S. Lemaignan, and T. Belpaeme. From characterising three years of hri to methodology and reporting recommendations. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 391–398. IEEE, 2016.

[20] J. M. Beer, A. D. Fisk, and W. A. Rogers. Toward a framework for levels of robot autonomy in human-robot interaction. *Journal of human-robot interaction*, 3(2):74, 2014.

[21] K. Belhassein, G. Buisan, A. Clodic, and R. Alami. Towards methodological principles for user studies in human-robot interaction. In *Test Methods and Metrics for Effective HRI in Collaborative Human-Robot Teams Workshop, ACM/IEEE International Conference on Human-Robot Interaction*, 2019.

[22] M. Bennett, T. Williams, D. Thames, and M. Scheutz. Differences in interaction patterns and perception for teleoperated and autonomous humanoid robots. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 6589–6594. IEEE, 2017.

[23] D. Benyon and C. Macaulay. Scenarios and the hci-se design problem. *Interacting with computers*, 14(4):397–405, 2002.

[24] C. L. Bethel and R. R. Murphy. Use of large sample sizes and multiple evaluation methods in human-robot interaction experimentation. In *AAAI Spring Symposium: Experimental Design for Real-World Systems*, pages 9–16, 2009.

[25] C. L. Bethel and R. R. Murphy. Review of human studies methods in hri and recommendations. *International Journal of Social Robotics*, 2(4):347–359, 2010.

[26] C. L. Bethel, K. Salomon, R. R. Murphy, and J. L. Burke. Survey of psychophysiology measurements applied to human-robot interaction. In *RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication*, pages 732–737. IEEE, 2007.

[27] L. Bickman and D. J. Rog. *The SAGE handbook of applied social research methods*. Sage publications, 2008.

[28] O. A. Bolarinwa. Principles and methods of validity and reliability testing of questionnaires used in social and health science researches. *Nigerian Postgraduate Medical Journal*, 22(4):195, 2015.

[29] C. Breazeal. Social interactions in hri: the robot view. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 34(2):181–186, 2004.

[30] E. Broadbent, C. Jayawardena, N. Kerse, R. Stafford, and B. A. MacDonald. Human-robot interaction research to improve quality of life in elder care-an approach and issues. 2011.

[31] R. Brooks. A robust layered control system for a mobile robot. *IEEE journal on robotics and automation*, 2(1):14–23, 1986.

[32] K. A. Brownlee. *Statistical theory and methodology in science and engineering*, volume 150. Wiley New York, 1965.

[33] C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner. The robotic social attributes scale (rosas) development and validation. In *Proceedings of the 2017 ACM/IEEE International Conference on human-robot interaction*, pages 254–262, 2017.

[34] J. M. Carroll. Becoming social: expanding scenario-based approaches in hci. *Behaviour & Information Technology*, 15(4):266–275, 1996.

[35] M. Cashmore, M. Fox, D. Long, D. Magazzeni, B. Ridder, A. Carrera, N. Palomeras, N. Hurtos, and M. Carreras. Rosplan: Planning in the robot operating system. In *Proceedings of the International Conference on Automated Planning and Scheduling*, volume 25, 2015.

[36] G. Charness, U. Gneezy, and M. A. Kuhn. Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, 81(1):1–8, 2012.

[37] M. Chita-Tegmark, M. Lohani, and M. Scheutz. Gender effects in perceptions of robots and humans with varying emotional intelligence. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 230–238. IEEE, 2019.

[38] C. Clabaugh and M. Matarić. Escaping oz: Autonomy in socially assistive robotics. *Annual Review of Control, Robotics, and Autonomous Systems*, 2:33–61, 2019.

[39] A. Coles, A. Coles, M. Fox, and D. Long. Forward-chaining partial-order planning. In *Proceedings of the International Conference on Automated Planning and Scheduling*, volume 20, 2010.

[40] D. Cramer. *Advanced quantitative data analysis*. McGraw-Hill Education (UK), 2003.

[41] M. F. Damholdt, C. Vestergaard, and J. Seibt. Testing for 'anthropomorphization': A case for mixed methods in human-robot interaction. In *Human-Robot Interaction*, pages 203–227. Springer, 2020.

[42] K. Dautenhahn. Methodology & themes of human-robot interaction: A growing research field. *International Journal of Advanced Robotic Systems*, 4(1):15, 2007.

[43] K. Dautenhahn. Socially intelligent robots: dimensions of human–robot interaction. *Philosophical transactions of the royal society B: Biological sciences*, 362(1480):679–704, 2007.

[44] K. Dautenhahn, M. Walters, S. Woods, K. L. Koay, C. L. Nehaniv, A. Sisbot, R. Alami, and T. Siméon. How may i serve you? a robot companion approaching a seated person in a helping context. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 172–179, 2006.

[45] K. Dautenhahn and I. Werry. A quantitative technique for analysing robot-human interactions. In *IEEE/RSJ international conference on intelligent robots and systems*, volume 2, pages 1132–1138. IEEE, 2002.

[46] J. Dawes. Do data characteristics change according to the number of scale points used? an experiment using 5-point, 7-point and 10-point scales. *International journal of market research*, 50(1):61–104, 2008.

[47] M. M. de Graaf. An ethical evaluation of human–robot relationships. *International journal of social robotics*, 8(4):589–598, 2016.

[48] M. M. De Graaf and S. B. Allouch. Exploring influencing variables for the acceptance of social robots. *Robotics and autonomous systems*, 61(12):1476–1486, 2013.

[49] M. M. de Graaf and S. B. Allouch. The relation between people's attitude and anxiety towards robots in human-robot interaction. In *2013 IEEE RO-MAN*, pages 632–637. IEEE, 2013.

[50] M. M. de Graaf, S. B. Allouch, and J. Van Dijk. What makes robots social?: A user's perspective on characteristics for social human-robot interaction. In *International Conference on Social Robotics*, pages 184–193. Springer, 2015.

[51] A. Dix, A. J. Dix, J. Finlay, G. D. Abowd, and R. Beale. *Human-computer interaction*. Pearson Education, 2003.

[52] F. Eyssel, D. Kuchenbrandt, F. Hegel, and L. de Ruiter. Activating elicited agent knowledge: How robot and user features shape the perception of social robots. In *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*, pages 851–857. IEEE, 2012.

[53] D. Feil-Seifer and M. Matarić. Using proxemics to evaluate human-robot interaction. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 143–144. IEEE, 2010.

[54] Y. Fernaeus, S. Ljungblad, M. Jacobsson, and A. Taylor. Where third wave hci meets hri: report from a workshop on user-centred design of robots. In *2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 293–294. IEEE, 2009.

[55] G. A. Fine and K. D. Elsbach. Ethnography and experiment in social psychological theory building: Tactics for integrating qualitative field data with quantitative lab data. *Journal of Experimental Social Psychology*, 36(1):51–76, 2000.

[56] S. M. Fiore, T. J. Wiltshire, E. J. Lobato, F. G. Jentsch, W. H. Huang, and B. Axelrod. Toward understanding social cues and signals in human–robot interaction: effects of robot gaze and proxemic behavior. *Frontiers in psychology*, 4:859, 2013.

[57] K. Fischer. Interpersonal variation in understanding robots as social actors. In *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 53–60. IEEE, 2011.

[58] S. T. Fiske, A. J. Cuddy, and P. Glick. Universal dimensions of social cognition: Warmth and competence. *Trends in cognitive sciences*, 11(2):77–83, 2007.

[59] M. E. Foster. Natural language generation for social robotics: opportunities and challenges. *Philosophical Transactions of the Royal Society B*, 374(1771):20180027, 2019.

[60] S. George and L. Mallery. Alfa de cronbach y consistencia interna de los ítems de un instrumento de medida. *Revista de estudios Interdisciplinarios en Ciencias Sociales*, 3(16):3–9, 2003.

[61] L. Giusti and P. Marti. Interpretative dynamics in human robot interaction. In *ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication*, pages 111–116. IEEE, 2006.

[62] D. F. Glas, S. Satake, F. Ferreri, T. Kanda, N. Hagita, and H. Ishiguro. The network robot system: enabling social human-robot interaction in public spaces. *Journal of Human-Robot Interaction*, 1(2):5–32, 2012.

[63] J. A. Gliem and R. R. Gliem. Calculating, interpreting, and reporting cronbach's alpha reliability coefficient for likert-type scales. Midwest Research-to-Practice Conference in Adult, Continuing, and Community . . . , 2003.

[64] K. Gold, I. Fasel, N. G. Freier, and C. Torrey. Young researchers' views on the current and future state of hri. In *2007 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 357–364. IEEE, 2007.

[65] M. A. Goodrich and A. C. Schultz. *Human-robot interaction: a survey*. Now Publishers Inc, 2008.

[66] M. Grandgeorge. Evaluating human-robot interaction with ethology. In *Human-Robot Interaction*, pages 257–268. Springer, 2020.

[67] H. T. Greely. Legal, ethical, and social issues in human genome research. *Annual Review of Anthropology*, 27(1):473–502, 1998.

[68] B. Han, H. W. Kim, and J.-H. Yoo. Deep emotion change detection for human-robot interaction.

[69] M. Heerink. Exploring the influence of age, gender, education and computer experience on robot acceptance by older adults. In *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 147–148. IEEE, 2011.

[70] M. Hennink, I. Hutter, and A. Bailey. *Qualitative research methods*. SAGE Publications Limited, 2020.

[71] C. Heyer. Human-robot interaction and future industrial robotics applications. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4749–4754. IEEE, 2010.

[72] G. Hoffman and X. Zhao. A primer for conducting experiments in human–robot interaction. *ACM Transactions on Human-Robot Interaction (THRI)*, 10(1):1–31, 2020.

[73] A. Hong, N. Lunscher, T. Hu, Y. Tsuboi, X. Zhang, S. F. dos Reis Alves, G. Nejat, and B. Benhabib. A multimodal emotional human-robot interaction architecture for social robots engaged in bidirectional communication\vspace* 7pt. *IEEE Transactions on Cybernetics*, 2020.

[74] Y. Hu, M. Benallegue, G. Venture, and E. Yoshida. Interact with me: An exploratory study on interaction factors for active physical human-robot interaction. *IEEE Robotics and Automation Letters*, 5(4):6764–6771, 2020.

[75] H. Hüttenrauch. *From HCI to HRI: designing interaction for a service robot*. PhD thesis, KTH, 2006.

[76] H. Hüttenrauch and K. Severinson-Eklundh. To help or not to help a service robot: Bystander intervention as a resource in human–robot collaboration. *Interaction Studies*, 7(3):455–477, 2006.

[77] L. Iocchi, D. Holz, J. Ruiz-del Solar, K. Sugiura, and T. Van Der Zant. Robocup@ home: Analysis and results of evolving competitions for domestic and service robots. *Artificial Intelligence*, 229:258–281, 2015.

[78] B. Irfan, J. Kennedy, S. Lemaignan, F. Papadopoulos, E. Senft, and T. Belpaeme. Social psychology and human-robot interaction: An uneasy marriage. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 13–20, 2018.

[79] B. Jensen, N. Tomatis, L. Mayor, A. Drygajlo, and R. Siegwart. Robots meet humans-interaction in public spaces. *IEEE Transactions on Industrial Electronics*, 52(6):1530–1546, 2005.

[80] T. D. Jick. Mixing qualitative and quantitative methods: Triangulation in action. *Administrative science quarterly*, 24(4):602–611, 1979.

[81] W. Ju and L. Takayama. Should robots or people do these jobs?

[82] M. F. Jung. Affective grounding in human-robot interaction. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*, pages 263–273. IEEE, 2017.

[83] T. Kanda, H. Ishiguro, and T. Ishida. Psychological analysis on human-robot interaction. In *Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No. 01CH37164)*, volume 4, pages 4166–4173. IEEE, 2001.

[84] K. Kawamura, R. T. Pack, M. Bishay, and M. Iskarous. Design philosophy for service robots. *Robotics and Autonomous Systems*, 18(1-2):109–116, 1996.

[85] T. Kelley and K. Dickerson. A review of artificial intelligence (ai) algorithms for sound classification: Implications for human-robot interaction (hri). Technical report, CCDC Army Research Laboratory Adelphi United States, 2020.

[86] D. Kember and D. Y. Leung. Establishing the validity and reliability of course evaluation questionnaires. *Assessment & Evaluation in Higher Education*, 33(4):341–353, 2008.

[87] H. Khayrallah, S. Trott, and J. Feldman. Natural language for human robot interaction. In *International Conference on Human-Robot Interaction (HRI)*, 2015.

[88] C. D. Kidd and C. Breazeal. Robots at home: Understanding long-term human-robot interaction. In *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3230–3235. IEEE, 2008.

[89] S. Kiesler and P. Hinds. Introduction to this special issue on human-robot interaction. *Human–Computer Interaction*, 19(1-2):1–8, 2004.

[90] K. L. Koay, D. S. Syrdal, M. L. Walters, and K. Dautenhahn. Five weeks in the robot house–exploratory human-robot interaction trials in a domestic setting. In *2009 Second International Conferences on Advances in Computer-Human Interactions*, pages 219–226. IEEE, 2009.

[91] T. Koulouri, S. Lauria, R. D. Macredie, and S. Chen. Are we there yet? the role of gender on the effectiveness and efficiency of user-robot communication in navigational tasks. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 19(1):1–29, 2012.

[92] C. U. Krägeloh, J. Bharatharaj, S. K. Sasthan Kutty, P. R. Nirmala, and L. Huang. Questionnaires to measure acceptability of social robots: a critical review. *Robotics*, 8(4):88, 2019.

[93] N. C. Krämer, A. von der Pütten, and S. Eimler. Human-agent and human-robot interaction theory: similarities to and differences from human-human interaction. In *Human-computer interaction: The agency perspective*, pages 215–240. Springer, 2012.

[94] B. Kuipers, E. A. Feigenbaum, P. E. Hart, and N. J. Nilsson. Shakey: from conception to history. *Ai Magazine*, 38(1):88–103, 2017.

[95] I. H. Kuo, J. M. Rabindran, E. Broadbent, Y. I. Lee, N. Kerse, R. Stafford, and B. A. MacDonald. Age and gender factors in user acceptance of healthcare robots. In *RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication*, pages 214–219. IEEE, 2009.

[96] A. Lambert, N. Norouzi, G. Bruder, and G. Welch. A systematic review of ten years of research on human interaction with social robots. *International Journal of Human–Computer Interaction*, pages 1–14, 2020.

[97] P. A. Lasota, T. Fong, J. A. Shah, et al. *A survey of methods for safe human-robot interaction*. Now Publishers, 2017.

[98] P. A. Lasota, G. F. Rossano, and J. A. Shah. Toward safe close-proximity human-robot interaction with standard industrial robots. In *2014 IEEE International Conference on Automation Science and Engineering (CASE)*, pages 339–344. IEEE, 2014.

[99] H. R. Lee and S. Šabanović. Culturally variable preferences for robot design and use in south korea, turkey, and the united states. In *2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 17–24. IEEE, 2014.

[100] D. Li, P. P. Rau, and Y. Li. A cross-cultural study: Effect of robot appearance and task. *International Journal of Social Robotics*, 2(2):175–186, 2010.

[101] C. H. Lin, E. Z. F. Liu, and Y. Y. Huang. Exploring parents' perceptions towards educational robots: Gender and socio-economic differences. *British Journal of Educational Technology*, 43(1):E31–E34, 2012.

[102] M. Makatchev, R. Simmons, M. Sakr, and M. Ziadee. Expressing ethnicity through behaviors of a robot character. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 357–364. IEEE, 2013.

[103] N. A. Malik, H. Yussof, and F. A. Hanapiah. Interactive scenario development of robot-assisted therapy for cerebral palsy: a face validation survey. *Procedia Computer Science*, 105:322–327, 2017.

[104] N. A. Malik, H. Yussof, F. A. Hanapiah, R. A. A. Rahman, and H. H. Basri. Human-robot interaction for children with cerebral palsy: Reflection and suggestion for interactive scenario design. *Procedia Computer Science*, 76:388–393, 2015.

[105] R. Mead and M. J. Matarić. Proxemics and performance: Subjective human evaluations of autonomous sociable robot distance and social signal understanding. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5984–5991. IEEE, 2015.

[106] D. M. Mertens. *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods*. Sage publications, 2014.

[107] Y. Mizuchi and T. Inamura. Optimization of criterion for objective evaluation of hri performance that approximates subjective evaluation: a case study in robot competition. *Advanced Robotics*, 34(3-4):142–156, 2020.

[108] A. B. Moniz and B.-J. Krings. Robots working with humans or humans working with robots? searching for social dimensions in new human-robot interaction in industry. *Societies*, 6(3):23, 2016.

[109] J. Mumm and B. Mutlu. Human-robot proxemics: physical and psychological distancing in human-robot interaction. In *Proceedings of the 6th international conference on Human-robot interaction*, pages 331–338, 2011.

[110] R. R. Murphy, T. Nomura, A. Billard, and J. L. Burke. Human–robot interaction. *IEEE robotics & automation magazine*, 17(2):85–89, 2010.

[111] R. R. Murphy and D. Schreckenghost. Survey of metrics for human-robot interaction. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 197–198. IEEE, 2013.

[112] B. Mutlu, S. Osman, J. Forlizzi, J. Hodgins, and S. Kiesler. Perceptions of asimo: an exploration on co-operation and competition with humans and humanoid robots. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 351–352, 2006.

[113] B. Mutlu, S. Osman, J. Forlizzi, J. Hodgins, and S. Kiesler. Task structure and user attributes as elements of human-robot interaction design. In *ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication*, pages 74–79. IEEE, 2006.

[114] D. C. Mutz. *Population-based survey experiments*. Princeton University Press, 2011.

[115] E. Mwangi, E. I. Barakova, M. Díaz-Boladeras, A. C. Mallofré, and M. Rauterberg. Directing attention through gaze hints improves task solving in human–humanoid interaction. *International Journal of Social Robotics*, 10(3):343–355, 2018.

[116] C. Nass, Y. Moon, B. J. Fogg, B. Reeves, and C. Dryer. Can computer personalities be human personalities? In *Conference companion on Human factors in computing systems*, pages 228–229, 1995.

[117] C. I. Nass and S. Brave. *Wired for speech: How voice activates and advances the human-computer relationship*. MIT press Cambridge, MA, 2005.

[118] M. Niemelä, P. Heikkilä, and H. Lammi. A social service robot in a shopping mall: expectations of the management, retailers and consumers. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 227–228, 2017.

[119] E. B. Nilsen, D. E. Bowler, and J. D. Linnell. Exploratory and confirmatory research in the open science era. *Journal of Applied Ecology*, 57(4):842–847, 2020.

[120] T. Ninomiya, A. Fujita, D. Suzuki, and H. Umemuro. Development of the multi-dimensional robot attitude scale: constructs of people's attitudes towards domestic robots. In *International Conference on Social Robotics*, pages 482–491. Springer, 2015.

[121] V. Nitsch and T. Glassen. Investigating the effects of robot behavior and attitude towards technology on social human-robot interactions. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 535–540. IEEE, 2015.

[122] S. Noel. Computer science: The ethical frontiers of robotics. *Science*, 322(5909):1800–1801, 2008.

[123] T. Nomura, T. Kanda, and T. Suzuki. Experimental investigation into influence of negative attitudes toward robots on human–robot interaction. *Ai & Society*, 20(2):138–150, 2006.

[124] T. Nomura, T. Kanda, T. Suzuki, and K. Kato. Prediction of human behavior in human–robot interaction using psychological scales for anxiety and negative attitudes toward robots. *IEEE transactions on robotics*, 24(2):442–451, 2008.

[125] T. Nomura and A. Nakao. Comparison on identification of affective body motions by robots between elder people and university students: A case study in japan. *International Journal of Social Robotics*, 2(2):147–157, 2010.

[126] T. Nomura, K. Sugimoto, D. S. Syrdal, and K. Dautenhahn. Social acceptance of humanoid robots in japan: A survey for development of the frankenstein syndorome questionnaire. In *2012 12th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2012)*, pages 242–247. IEEE, 2012.

[127] T. Nomura, T. Suzuki, T. Kanda, and K. Kato. Altered attitudes of people toward robots: Investigation through the negative attitudes toward robots scale. In *Proc. AAAI-06 workshop on human implications of human-robot interaction*, volume 2006, pages 29–35, 2006.

[128] A. Norton, W. Ober, L. Baraniecki, E. McCann, J. Scholtz, D. Shane, A. Skinner, R. Watson, and H. Yanco. Analysis of human–robot interaction at the darpa robotics challenge finals. *The International Journal of Robotics Research*, 36(5-7):483–513, 2017.

[129] S. F. Ong. Constructing a survey questionnaire to collect data on service quality of business academics. 2012.

[130] L. Onnasch and E. Roesler. A taxonomy to structure and analyze human–robot interaction. *International Journal of Social Robotics*, pages 1–17, 2020.

[131] J. Pages, L. Marchionni, and F. Ferro. Tiago: the modular robot that adapts to different research needs. In *International workshop on robot modularity, IROS*, 2016.

[132] A. K. Pandey and R. Gelin. A mass-produced sociable humanoid robot: Pepper: The first machine of its kind. *IEEE Robotics & Automation Magazine*, 25(3):40–48, 2018.

[133] P. Papadakis, P. Rives, and A. Spalanzani. Adaptive spacing in human-robot interactions. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2627–2632. IEEE, 2014.

[134] P. Parashar, L. M. Sanneman, J. A. Shah, and H. I. Christensen. A taxonomy for characterizing modes of interactions in goal-driven, human-robot teams. In *IROS*, pages 2213–2220, 2019.

[135] A. R.-V. D. Pütten and N. Bock. Development and validation of the self-efficacy in human-robot-interaction scale (se-hri). *ACM Transactions on Human-Robot Interaction (THRI)*, 7(3):1–30, 2018.

[136] C. R. Rao, C. R. Rao, M. Statistiker, C. R. Rao, and C. R. Rao. *Linear statistical inference and its applications*, volume 2. Wiley New York, 1973.

[137] B. Reeves and C. I. Nass. *The media equation: How people treat computers, television, and new media like real people and places.* Cambridge university press, 1996.

[138] L. Riek and D. Howard. A code of ethics for the human-robot interaction profession. *Proceedings of We Robot*, 2014.

[139] L. D. Riek. Wizard of oz studies in hri: a systematic review and new reporting guidelines. *Journal of Human-Robot Interaction*, 1(1):119–136, 2012.

[140] L. D. Riek, W. Hartzog, D. A. Howard, A. Moon, and R. Calo. The emerging policy and ethics of human robot interaction. *HRI (Extended Abstracts)*, 10(2701973.2714393), 2015.

[141] B. Robins, E. Ferrari, and K. Dautenhahn. Developing scenarios for robot assisted play. In *RO-MAN 2008-The 17th IEEE International Symposium on Robot and Human Interactive Communication*, pages 180–186. IEEE, 2008.

[142] S. Rossi, F. Ferland, and A. Tapus. User profiling and behavioral adaptation for hri: A survey. *Pattern Recognition Letters*, 99:3–12, 2017.

[143] M. B. Rosson and J. M. Carroll. *Usability engineering: scenario-based development of human-computer interaction.* Morgan Kaufmann, 2002.

[144] S. Sabanovic, M. P. Michalowski, and R. Simmons. Robots in the wild: Observing human-robot social interaction outside the lab. In *9th IEEE International Workshop on Advanced Motion Control, 2006.*, pages 596–601. IEEE, 2006.

[145] M. Salem, G. Lakatos, F. Amirabdollahian, and K. Dautenhahn. Would you trust a (faulty) robot? effects of error, task type and personality on human-robot cooperation and trust. In *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 1–8. IEEE, 2015.

[146] S. Saunderson and G. Nejat. Investigating strategies for robot persuasion in social human-robot interaction. *IEEE Transactions on Cybernetics*, 2020.

[147] P. Schermerhorn, M. Scheutz, and C. R. Crowell. Robot social presence and gender: Do females view robots differently than males? In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pages 263–270, 2008.

[148] G. Schillaci, S. Bodiroža, and V. V. Hafner. Evaluating the effect of saliency detection and attention manipulation in human-robot interaction. *International Journal of Social Robotics*, 5(1):139–152, 2013.

[149] J. Scholtz. Theory and evaluation of human robot interactions. In *36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the*, pages 10–pp. IEEE, 2003.

[150] S. K. Shah and K. G. Corley. Building better theory by bridging the quantitative–qualitative divide. *Journal of management studies*, 43(8):1821–1835, 2006.

[151] S. Shahid, E. Krahmer, M. Swerts, and O. Mubin. Child-robot interaction during collaborative game play: Effects of age and gender on emotion and experience. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*, pages 332–335, 2010.

[152] T. B. Sheridan. Human–robot interaction: status and challenges. *Human factors*, 58(4):525–532, 2016.

[153] B. Siciliano and O. Khatib. *Springer handbook of robotics*. Springer, 2016.

[154] R. M. Siino and P. J. Hinds. Robots, gender & amp; sensemaking: Sex segregation's impact on workers making sense of a mobile autonomous robot. In *Proceedings of the 2005 IEEE international conference on robotics and automation*, pages 2773–2778. IEEE, 2005.

[155] D. Y. Y. Sim and C. K. Loo. Extensive assessment and evaluation methodologies on assistive social robots for modelling human–robot interaction – a review. *Information Sciences*, 301:305 – 344, 2015.

[156] G. Skantze. Predicting and regulating participation equality in human-robot conversations: Effects of age and gender. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*, pages 196–204. IEEE, 2017.

[157] R. A. Stebbins. *Exploratory research in the social sciences*, volume 48. Sage, 2001.

[158] A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, and M. Goodrich. Common metrics for human-robot interaction. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 33–40, 2006.

[159] M. Strait, P. Briggs, and M. Scheutz. Gender, more so than age, modulates positive perceptions of language-based human-robot interactions. In *4th international symposium on new frontiers in human robot interaction*, pages 21–22, 2015.

[160] M. Strait, C. Canning, and M. Scheutz. Let me tell you! investigating the effects of robot communication strategies in advice-giving situations based on robot appearance, interaction modality and distance. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 479–486, 2014.

[161] M. Studley and A. Winfield. Elsa in industrial robotics. *Current Robotics Reports*, pages 1–8, 2020.

[162] D. S. Syrdal, K. Dautenhahn, K. L. Koay, and M. L. Walters. The negative attitudes towards robots scale and reactions to robot behaviour in a live human-robot interaction study. *Adaptive and emergent behaviour and complex systems*, 2009.

[163] D. S. Syrdal, N. Otero, and K. Dautenhahn. Video prototyping in human-robot interaction: Results from a qualitative study. In *Proceedings of the 15th European conference on Cognitive ergonomics: the ergonomics of cool interaction*, pages 1–8, 2008.

[164] L. Takayama and C. Pantofaru. Influences on proxemic behaviors in human-robot interaction. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5495–5502. IEEE, 2009.

[165] L. Tiberio, A. Cesta, and M. Olivetti Belardinelli. Psychophysiological methods to evaluate user's response in human robot interaction: a review and feasibility study. *Robotics*, 2(2):92–121, 2013.

[166] I. Tiddi, E. Bastianelli, E. Daga, M. d'Aquin, and E. Motta. Robot–city interaction: Mapping the research landscape—a survey of the interactions between robots and modern cities. *International Journal of Social Robotics*, 12(2):299–324, 2020.

[167] K. Tsiakas, M. Kyrarini, V. Karkaletsis, F. Makedon, and O. Korn. A taxonomy in robot-assisted training: current trends, needs and challenges. *Technologies*, 6(4):119, 2018.

[168] S. Tuncer. A cross-cultural perspective on human-robot interaction and artificial intelligence. 2020.

[169] F.-W. Tung. Influence of gender and age on the attitudes of children towards humanoid robots. In *International Conference on Human-Computer Interaction*, pages 637–646. Springer, 2011.

[170] J. R. Turner and J. Thayer. *Introduction to analysis of variance: design, analyis & interpretation.* Sage Publications, 2001.

[171] A. Vanzo, F. Riccio, M. Sharf, V. Mirabella, T. Catarci, and D. Nardi. Who is willing to help robots? a user study on collaboration attitude. *International Journal of Social Robotics*, 12(2):589–598, 2020.

[172] V. Villani, F. Pini, F. Leali, and C. Secchi. Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications. *Mechatronics*, 55:248–266, 2018.

[173] E.-J. Wagenmakers, R. Wetzels, D. Borsboom, H. L. van der Maas, and R. A. Kievit. An agenda for purely confirmatory research. *Perspectives on Psychological Science*, 7(6):632–638, 2012.

[174] M. Walters, S. Marcos, D. S. Syrdal, and K. Dautenhahn. An interactive game with a robot: People's perceptions of robot faces and a gesture-based user interface. In *Proc. 6th Int. Conf. Adv. Computer–Human Interactions*, pages 123–128, 2013.

[175] S. Walther and T. Guhl. Classification of physical human-robot interaction scenarios to identify relevant requirements. In *ISR/Robotik 2014; 41st International Symposium on Robotics*, pages 1–8. VDE, 2014.

[176] L. Wang, L. Iocchi, A. Marrella, and D. Nardi. Developing a questionnaire to evaluate customers' perception in the smart city robotic challenge. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1–6. IEEE, 2019.

[177] L. Wang, L. Iocchi, A. Marrella, and D. Nardi. Designing and evaluating hri teaming tasks based on the sciroc challenge. In *ACM Transaction on Human-Robot Interaction*, page Under review, 2020.

[178] L. Wang, L. Iocchi, A. Marrella, and D. Nardi. Hri users' studies in the context of the sciroc challenge: Some insights on gender-based differences. In *Proceedings of the 8th International Conference on Human-Agent Interaction*, pages 287–289, 2020.

[179] L. Wang, A. Marrella, and D. Nardi. Investigating user perceptions of hri in social contexts. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 544–545. IEEE, 2019.

[180] L. Wang, P.-L. P. Rau, V. Evers, B. Robinson, and P. Hinds. Responsiveness to robots: Effects of ingroup orientation &. ommunication style on hri in china. In *2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 247–248. IEEE, 2009.

[181] K. E. Weick. The social psychology of organizing addison-wesley. *Reading, MA*, 1979.

[182] A. Weiss. *Validation of an evaluation framework for human-robot interaction: the impact of usability, social acceptance, user experience, and societal impact on collaboration with humanoid robots.* na, 2010.

[183] A. Weiss, R. Bernhaupt, M. Tscheligi, and E. Yoshida. Addressing user experience and societal impact in a user study with a humanoid robot. In *AISB2009: Proceedings of the Symposium on New Frontiers in Human-Robot Interaction (Edinburgh, 8-9 April 2009), SSAISB*, pages 150–157. Citeseer, 2009.

[184] A. Weiss, J. Igelsböck, M. Tscheligi, A. Bauer, K. Kühnlenz, D. Wollherr, and M. Buss. Robots asking for directions—the willingness of passers-by to support robots. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 23–30. IEEE, 2010.

[185] A. Weiss, T. Scherndl, M. Tscheligi, and A. Billard. Evaluating the icra 2008 hri challenge. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pages 261–262, 2009.

[186] T. Wisspeintner, T. Van Der Zant, L. Iocchi, and S. Schiffer. Robocup@ home: Scientific competition and benchmarking for domestic service robots. *Interaction Studies*, 10(3):392–426, 2009.

[187] P. Wright. What's in a scenario? *ACM SIGCHI Bulletin*, 24(4):11–12, 1992.

[188] Q. Xu, J. Ng, O. Tan, Z. Huang, B. Tay, and T. Park. Methodological issues in scenario-based evaluation of human–robot interaction. *International Journal of Social Robotics*, 7(2):279–291, 2015.

[189] H. A. Yanco, A. Norton, W. Ober, D. Shane, A. Skinner, and J. Vice. Analysis of human-robot interaction at the darpa robotics challenge trials. *Journal of Field Robotics*, 32(3):420–444, 2015.

[190] J. E. Young, J. Sung, A. Voida, E. Sharlin, T. Igarashi, H. I. Christensen, and R. E. Grinter. Evaluating human-robot interaction. *International Journal of Social Robotics*, 3(1):53–67, 2011.

[191] J. Złotowski, A. Weiss, and M. Tscheligi. Interaction scenarios for hri in public space. In *International Conference on Social Robotics*, pages 1–10. Springer, 2011.

Acknowledgements

First of all, I would like to express my deep sense of gratitude to my advisor Prof. *Daniele Nardi* for encouraging me to start my Ph.D Program and giving me the continuous support throughout the whole period of my Ph.D Program. His dedication, patience and immense knowledge helped me to accomplish the research tasks in Human-Robot interaction (HRI). He also inspired me by his hard working and positive attitudes. I acknowledged his guidance and supervision for the writing of this thesis as well.

My sincere gratitude also goes to my Co-advisor Prof. *Andrea Marrella*, who gave me valuable guidance and expressed keen interest in my research topic. His kindness, enthusiasm, and dynamism helped me to complete several significant publications in the field. I really appreciate his vast knowledge and skills in my research field.

My deep gratitude also goes to my co-author of several publications, Prof. *Luca Iocchi*, who taught me first lessons in HRI, gave me the opportunity to acquire knowledge in this research field and provided me meaningful suggestions in different research topics. Moreover, many special thanks to all actual members and ex members of Ro.Co.Co.Lab, in particular, *Mayte*, *Paola*, *Ali Youssef*, *Francesco*, *Roberto*, *Andrea*, *Vincenzo*, *Giorgio*, *Bartolo*, *Dominik*, etc.

Then, I would like to thank a special person, my son *Giovanni Hongyu*. His birth give me incalculable happiness and strengths. Thank a lot for choosing me to be your Mom.

In the end, I would like to express my extremely gratitude to my family for their constantly supporting, caring and guiding. Without the inspiration that they gave me, I might not the person I am today.

Rome, March 2021