UNIVERSITY
OF SKÖVDE

DOCTORAL DISSERTATION

WHAT DID YOU EXPECT?

A human-centered approach to investigating and reducing the social robot expectation gap

JULIA ROSÉN
Informatics

WHAT DID YOU EXPECT?

A human-centered approach to investigating and reducing the social robot expectation gap

DOCTORAL DISSERTATION

WHAT DID YOU EXPECT?

A human-centered approach to investigating and reducing the social robot expectation gap

JULIA ROSÉN
Informatics

UNIVERSITY
OF SKÖVDE

Julia Rosén, 2024

Doctoral Dissertation

Title: What did you expect?

A human-centered approach to investigating and reducing the social robot expectation gap

University of Skövde 2024, Sweden
www.his.se

Printer: Stema Specialtryck AB, Borås

ISBN 978-91-987906-9-6

Dissertation Series No. 55 (2024)

ABSTRACT

We live in a complex world where we proactively plan and execute various behaviors by forming expectations in real time. Expectations are beliefs regarding the future state of affairs and they play an integral part of our perception, attention, and behavior. Over time, our expectations become more accurate as we interact with the world and others around us. People interact socially with other people by inferring others' purposes, intentions, preferences, beliefs, emotions, thoughts, and goals. Similar inferences may occur when we interact with social robots. With anthropomorphic design, these robots are designed to mimic people physically and behaviorally. As a result, users predominantly infer agency in social robots, often leading to mismatched expectations of the robots' capabilities, which ultimately influences the user experience.

In this thesis, the role and relevance of users' expectations in first-hand social human-robot interaction (sHRI) was investigated. There are two major findings. First, in order to study expectations in sHRI, the social robot expectation gap evaluation framework was developed. This framework supports the systematic study and evaluation of expectations over time, considering the unique context where the interaction is unfolding. Use of the framework can inform sHRI researchers and designers on how to manage users' expectations, not only in the design, but also during evaluation and presentation of social robots. Expectations can be managed by identifying what kinds of expectations users have and aligning these through design and dissemination which ultimately creates more transparent and successful interactions and collaborations. The framework is a tool for achieving this goal. Second, results show that previous experience has a strong impact on users' expectations. People have different expectations of social robots and view social robots as both human-like and as machines. Expectations of social robots can vary according to the source of the expectation, with those who had previous direct experiences of robots having different expectations than those who relied on indirect experiences to generate expectations.

One consequence of these results is that expectations can be a confounding variable in sHRI research. Previous experience with social robots can prime users in future interactions with social robots. These findings highlight the unique experiences users have, even when faced with the same robot. Users' expectations and how they change over time shapes the users' individual needs and preferences and should therefore be considered in the interpretation of sHRI. In doing so, the social robot expectation gap can be reduced.

SAMMANFATTNING

Vi lever i en komplex värld och för att kunna hantera denna komplexitet formar vi förväntningar. Förväntningar är antaganden om framtida tillstånd och är en vital del av vår perception, uppmärksamhet och beteende. Genom att interagera med omvärlden och andra människor blir våra förväntningar mer precisa och korrekta över tid. I en social interaktion behöver vi förstå den andra personens syften, avsikter, preferenser, övertygelser, känslor, tankar och mål. Sociala robotar är utformade för att skapa liknande inferenser när användare interagerar med dem. Detta kan leda till missbedömningar mellan vad vi förväntar oss av sociala robotar och vad dessa artefakter är kapabla till, vilket påverkar användarupplevelsen av sociala robotar.

I den här avhandlingen presenteras den forskning som har utförts för att studera rollen och relevansen av människors förväntningar i social människa-robotinteraktion (sMRI). Resultaten kan delas in i två större fynd. Det första fyndet är ett utvärderingsramverk som ämnar att systematiskt studera användares förväntningar av sociala robotar i en interaktion, med fokus på hur förväntningar ändras över tid i en interaktion, med interaktionens unika kontext i åtanke. Ramverket är menat för designers av sociala robotar och forskare inom sMRI-fältet för att bättre studera, hantera, och förstå förväntningar, både i robotarnas design och i robotarnas agerande. Det andra fyndet består av de empiriska resultat som visar hur tidigare erfarenheter påverkar användares förväntningar. Förväntningarna baseras till stor del på vilka typer av tidigare erfarenheter användare har, där de med direkta erfarenheter av robotar har andra förväntningar än de med indirekta erfarenheter. Vidare visar resultaten att användare ser sociala robotar både som människolika och som maskiner samtidigt.

Förväntningar kan också ses som en bakomliggande variabel inom sMRI-forskning eftersom tidigare erfarenheter kan påverka deltagare i kommande interaktioner med sociala robotar. Resultaten visar även att användarupplevelsen är unik för varje användare, även om roboten är densamma, vilket bör tas i åtanke när resultat tolkas i en sMRI-kontext. Genom att ha förväntningar i åtanke kan vi minska det gap som uppstår mellan människors förväntningar av sociala robotar och robotarnas faktiska förmågor. På så sätt kan vi främja positiva användarupplevelser och förbättra interaktionen mellan människa och robot.

ACKNOWLEDGEMENTS

This thesis owes its existence to the incredible people in both my professional and personal life. It's not often we get the chance to express our gratitude to those who mean the most to us. While words may fall short of capturing the depth of my appreciation, I will try nevertheless.

I would like to begin by thanking my advisors: Erik B, Jessica, Maurice, and Christian. Your tireless commitment to listening to my ideas, offering invaluable insights, and providing unwavering support sustained me throughout this journey. This work would have not been possible without your advisement on the different aspects of expectations. Perhaps the most profound advice I received during these years was that "philosophy is like a healthy snack." I think, with my thesis complete, that I get it now... *maybe*. Thank you all.

My gratitude extends to my friends in academia, both in Skövde and beyond, who have demonstrated remarkable patience and understanding, never tiring of my endless discussions about my work. Special appreciation goes to my office roommates, Erik L and Kajsa, who've shared the trenches with me daily, listening to my concerns and offering their support. The thought of not seeing you at work every day is difficult to fathom. This journey would have been far more challenging without your presence. To my co-workers and friends Vipul and Sara, who embarked on this journey with me, from our early days taking courses together to the later years pushing each other towards the finish line. To Maja, who from our Master's in Lund to now has been an unwavering source of support through it all. I struggle to find the right words to convey how much you mean to me. To Katie, who has provided steadfast support since our first meeting. To all the members of iLab, the PhD Student Council, and the Research Ethics Council. Your roles in my academic life have been invaluable.

I want to acknowledge my friends who may not fully grasp my academic pursuits but have consistently shown me love and appreciation. To Victoria, my lifelong #1 supporter, you've been a constant presence through every phase of my life starting in our childhood. You have followed me around the world, showering me with love along the way. Our relationship is a treasure, and I'm immensely grateful for your enduring presence. To Daniel, the master of top-tier memes and karaoke enthusiast, thank you for being my friend and for bringing out the fun side of me (GGSC+ forever). To Philip, the friend who never fails to make me laugh, thank you for bringing much needed laughter into my life. To my friends from my Bachelor's days in New York. Over a decade later, you remain cherished parts of my life, and I'm immensely proud of our

collective growth. To Joella, with whom I shared a home in Brooklyn and learned to be adults together with. I know you literally never take my advice but your guidance and support have been invaluable for me. Without you I would undoubtedly be worse off. To Josefine, our friendship has been marked not only by the fun moments we've shared but also by the challenging parts we've faced together. Thank you for being in my life. To Dan, despite the physical distance that separates us, you've managed to keep our connection strong. Your consistent prioritization of our friendship, your frequent visits to Stockholm, and the fun we have together all hold a special place in my heart. To TK and Petter, even though our meetings are infrequent, the moments we share are always filled with fun and laughter. To the Varberg gang, who always ask how I am, and with whom I always have the most fun. Special thanks to Louise, Victor, and Elias. I am so grateful for you.

My family and relatives have been a source of strength throughout my life, providing unwavering support during good and bad times. To my parents, Annika and Björn, your constant support, love, and encouragement have been a driving force throughout my life and academic journey. To my brothers, I'm forever grateful for the unique connection we share. To Cim and my big brother Adam, whose warm dinner invitations, thoughtful care of our beloved dogs, and delightful concoctions of fun drinks have added joy to this journey. Thank you for your sound advice and consistent support. To my little brother David, whose constant presence has allowed me to grow as a sibling and a friend. Although I will never forgive you for growing up, I want to thank you for the privilege of being your big sister. To my aunt Stina, who has been a comforting presence throughout my life. To Karin, Lars, and Ellen, who have provided solace during life's most challenging moments. To Lotta and John, who have been like second parents, offering support, love, and care. To my Victorin family: Lotte, Ulf, Johan, Caroline, and Georg. Your warmth and support after welcoming me into your family have been giving me comfort and strength throughout the trials of pursuing this degree.

And to my love, Mattias, who deserves at least half of this degree. You have carried, pushed, and pulled me across the finish line. Thank you for managing everything I had to neglect to complete this work. For providing snacks, taking care of the dogs, allowing me to blast Beyoncé at odd hours, enduring my three separate marathons through The Last of Us (this year alone), and all the other ways you've supported me. You've never asked for anything in return, but I will spend the rest of my life thanking you for this chapter in our lives.

Lastly, to my loyal companions, Hayley and Stella. Thank you for the unconditional love and for always being there, reminding me that sometimes a walk in the park is the best break from academia.

PUBLICATIONS

PUBLICATIONS WITH HIGH RELEVANCE

I Rosén, Julia, Lindblom, Jessica, and Billing, Erik (2021). “Reporting of Ethical Conduct in Human-Robot Interaction Research”. In: *Advances in Human Factors in Robots, Unmanned Systems and Cybersecurity*. Ed. by Matteo Zallio, Carlos Raymundo Ibañez, and Jesus Hechavarria Hernandez. Springer International Publishing, pp. 87–94.

II Rosén, Julia (2021). “Expectations in Human-Robot Interaction”. In: *Advances in Neuroergonomics and Cognitive Engineering*. Ed. by Hasan Ayaz, Umer Asgher, and Lucas Paletta. Springer International Publishing, pp. 98–105.

III Rosén, Julia, Lindblom, Jessica, and Billing, Erik (2022). “The Social Robot Expectation Gap Evaluation Framework”. In: *Human-Computer Interaction. Technological Innovation*. Ed. by Masaaki Kurosu. Springer International Publishing, pp. 590–610.

IV Rosén, Julia, Billing, Erik, and Lindblom, Jessica (2023). “Applying the Social Robot Expectation Gap Evaluation Framework”. In: *Human-Computer Interaction*. Ed. by Masaaki Kurosu and Ayako Hashizume. Springer International Publishing, pp. 169–188.

V Rosén, Julia, Lagerstedt, Erik, and Lamb, Maurice (2023). “Investigating NARS: Inconsistent Practice of Application and Reporting”. In: *The 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2023), Busan, South Korea, 2023*. IEEE, pp. 922–927.

VI Rosén, Julia, Lindblom, Jessica, Lamb, Maurice, and Billing, Erik (Under review). “Previous Experience Matters: An In-Person Investigation of Expectations in Human-Robot Interaction”. In: *Under review for scientific journal*, pp. 1–19.

VII Lindblom, Jessica, Rosén, Julia, Lamb, Maurice, and Billing, Erik (Manuscript). "Disentangling People's Experiences and Expectations when Interacting with the Social Robot Pepper: A Qualitative Analysis". In: *Manuscript for scientific journal*, pp. 1–41.

PUBLICATIONS WITH LOW RELEVANCE

VIII Rosén, Julia, Richardson, Kathleen, Lindblom, Jessica, and Billing, Erik (2018). "The Robot Illusion: Facts and Fiction". In: *Workshop in Explainable Robotics System, in conjunction with 13th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2018), Chicago, USA, March, 5–8, 2018*, pp. 1–2.

IX Billing, Erik, Rosén, Julia, and Lindblom, Jessica (2019). "Expectations of Robot Technology in Welfare". In: *The second workshop on social robots in therapy and care, in conjunction with the 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2019), Daegu, Korea, March 11–14 2019*, pp. 1–4.

X Rosén, Julia, Lindblom, Jessica, Lamb, Maurice, and Billing, Erik (2020). "Digital Human Modeling Technology in Virtual Reality–Studying Aspects of Users' Experiences". In: *DHM2020*. IOS Press, pp. 330–341.

XI Rosén, Julia, Lindblom, Jessica, Billing, Erik, and Lamb, Maurice (2021). "Ethical Challenges in the Human-Robot Interaction Field". In: *TRAITS Workshop, in conjunction with the 16th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2021), Boulder, USA, March 8–12, 2021*, pp. 1–2.

XII Rosén, Julia, Lagerstedt, Erik, and Lamb, Maurice (2022). "Is Human-Like Speech in Robots Deception?" In: *HRI'22 Workshop—Robo-Identity 2, in conjunction with the 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2022), online, March 7–10, 2022*, pp. 1–3.

XIII Rosén, Julia and Lagerstedt, Erik (2023). "Speaking Properly with Robots". In: *HRI'23 Workshop—Inclusive HRI II, Equity and Diversity in Design, Application, Methods, and Community, in conjunction with the 18th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2023), Stockholm, Sweden, March 13–16, 2023*, pp. 1–3.

XIV Billing, Erik, Rosén, Julia, and Lamb, Maurice (2023). "Language Models for Human-Robot Interaction". In: *Companion of the 18th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2023), Stockholm, Sweden, March 13–16, 2023*. Sweden, USA: ACM, pp. 905–906.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

CHAPTER 1
INTRODUCTION

We live in a world that is full of people, events, and things. In order to cope with the constant information flow from the world, we require a wide range of capabilities that aids us in parsing this information. A major factor in understanding the world is the expectations people have of these people, events, and things; i.e., looking ahead to an imagined future state of affairs and acting accordingly. The human mind uses past events to organize and reorganize itself to drive responses to external stimuli. This process is constant, happens on all cognitive levels, and over time we learn to understand the world better and make better predictions of the future (Olson, Roese, and Zanna, 1996; Roese and Sherman, 2007; Kveraga, Ghuman, and Bar, 2007; Hohwy, 2013).

Social interactions are an integral part of human life, and other people's thoughts, feelings, and actions, in relation to oneself is managed in real time by forming expectations of others (Premack and Woodruff, 1978; Dennett, 1989; Ward, 2015; Mascolo and Bidell, 2020). In turn, we adjust our own behavior in unison with these expectations in order to create successful social interactions. The expectations we form of people are quite different than the expectations we form of things. If a person wants a soda and interacts with a soda machine, that interaction is (usually) not characterized as social, and subsequently our expectations are not as complex as in social interactions. If the person inserts a coin and presses a button, they will expect to receive a soda. If a person, instead, wants a soda and asks a friend for one, that social interaction will require many complex expectations. The person may first consider the appropriateness of asking such a question by imagining what their response will be to such a request. The person may also consider how to ask the questions, and what to say after the friend's response depending on if they say yes or no. Then, if the friend agrees and gets a soda, the person may consider how the friend will get the soda and what to say once it is done. If the friend ends up not giving the soda, the person may feel hurt by their friend. In contrast, if the soda machine fails to dispense a soda, there might be disappointment, but there will probably not be any feelings of hurt because the soda machine cannot have premeditated malicious intent.

Indeed, our social expectations are different than other kinds of expectations, but there are artifacts that blur the line between what is considered social or not (Alač, 2016; Clark and Fischer, 2023). Specifically, there are robots that are made to act socially in numerous environments with the aim to meet the social and emotional needs of people (Breazeal, 2003; Fong, Nourbakhsh, and Dautenhahn, 2003; Thrun, 2004; Meister, 2014). These so-called social robots are designed with the intention of mirroring peoples' way of interacting socially. Depending on if people view a social robot as human-like or as a machine, the expectations formed may be from a human-human perspective, or it may be from a human-computer perspective. The maturity of people's expectations is influenced by society, and these expectations, in turn, shape what is perceived as ordinary, extraordinary, normal, or abnormal (Floridi, 2016). In science fiction, social robots are portrayed as "almost humans" with internal states, which may bleed into real human-robot interactions and create high expectations (Sandoval, Mubin, and Obaid, 2014; Alves-Oliveira et al., 2015; Oliveira and Yadollahi, 2023). The gap between a social robot's capabilities and its perceived capabilities can thus be vast. In terms of the soda machine example, it is harder to predict what people will expect from a social robot. A soda machine and a social robot are both machines, however, social robots are designed as being human-like in many regards which may generate expectations akin to that of a person when asked about receiving a soda. For example, the person may ask politely for a soda and have similar expectations as toward another person in relation to what kind of response the person may get.

Due to the duality of social robots (both human-like and a machine) it is harder to determine what kinds of expectations are formed. These expectations will ultimately affect the interaction between humans and social robots. Therefore, this Ph.D. thesis focuses on the role and relevance of users' expectations in social Human-Robot Interaction (sHRI). This topic is of major importance because these expectations are a considerable indicator of behavior, and ultimately determine the success of social interactions (Olson, Roese, and Zanna, 1996). Understanding the expectation of social robots and their many components are still an emerging area within the Human-Robot Interaction (HRI) field, and more specifically in the sub-field of sHRI. This line of research is surprisingly understudied. In fact, it is not uncommon that expectations are generally ignored when conducting sHRI studies. It is well-acknowledged, from the fields of psychology and cognitive science, that expectations are an underlying aspect of social interactions. This makes it especially problematic when sHRI ignores expectations because much of the research that is being done in sHRI today holds the assumption that human-robot interaction is similar to human-human interaction, with borrowed methodology from these human-centered fields. In addition, in the fields of Human-Computer Interaction (HCI) and User Experience (UX), the importance of users' expectations when interacting with interactive digital systems is a core aspect, both before, during and after the interaction (Roto et al., 2011). Although there have been a few attempts to explicitly study expectations in HRI, the role of expectations has not been systematically investigated and analyzed within the sHRI field (Lohse, 2009; Lohse, 2010; Meister, 2014; Kwon, Jung, and Knepper, 2016; Jokinen and Wilcock, 2017; Edwards et al., 2019; Horstmann and Krämer,

2020a; Manzi et al., 2021). Expectations are especially important to understand when the expectations are not met because this can guide and inform the future design of social robots, which ultimately would lead to more successful interactions and increased interaction quality. When expectations are not met an expectation gap is created, where the expected capability and the actual capability of the robot are not met or poorly aligned.

1.1 AIM AND RESEARCH QUESTIONS

The aim of this thesis is to investigate the role of user expectations play when interacting socially with social robots. I approach this aim from an interdisciplinary perspective to map out the concept of expectation and how it affects humans' interactions with social robots. The research questions are the following:

Research Question 1 (RQ1): How can users' expectations be studied in sHRI?

Research Question 2 (RQ2): What is the role and relevance of users' expectations in sHRI?

RQ1 was formulated with the motivation that there are no widely adopted methods, approaches, or techniques to study users' expectations of social robots within the sHRI field. Methods in sHRI are typically focused on users' preferences in and of themselves without paying any focus on how these preferences are actually formed. Understanding how these preferences are formed would provide additional important dimensions for analysis towards gaining a deeper understanding of social human-robot interactions. Understanding the role and relevance of user expectations allows for managing those expectations that ultimately may lead to successful interactions and positive experiences, which is one of the major goals within the sHRI field.

For RQ1, additional insights need to be acquired in relation to the methodological practices in the sHRI field. As mentioned, sHRI methods, approaches, or techniques are to a great extent borrowed from and inspired by other fields that focus on human-human interactions, such as psychology and cognitive science, and along with the fast-growing development and the nature of sHRI, there is a risk that certain practices become the norm without careful considerations (Irfan et al., 2020). The intended contribution of answering RQ1 is therefore the development of a framework for studying expectations and an increased methodological awareness within the sHRI field.

RQ2 was formulated with the motivation that users' expectations play a role in social human-robot interactions, serving as a potential confounding variable, and need more attention within the sHRI field. It is important to gain a deeper understanding of expectations because social robots stand out from other kinds of artifacts in the way that people may expect social robots to be and act both as human-like and as machines (Alač, 2016). In addition, people have less personal experience with these robots, relying more on the indirect experiences of social robots, such as those portrayed in media, which may lead to too high expectations

of social robots when interacting first-hand with them (Oliveira and Yadollahi, 2023). By gaining a richer understanding of users' expectations we can also manage these expectations by aligning expected robot capabilities and actual robot capabilities. Ultimately, having well-aligned expectations can contribute to successful interactions and, consequently, result in more positive user experiences. However, it's worth noting that there are instances where users have extremely low or negative expectations, and merely meeting these may not always guarantee a good user experience. The intended contribution of answering RQ2 is therefore to reduce the users' expectation gap of social robots.

1.2 THESIS OUTLINE

In chapter 2, I present background on the HRI and sHRI fields including a historical view of how the fields have evolved since their early years. Then I present an overview of expectations and how it relates to social robots, followed by concepts and theories that relate to expectations. Lastly, I briefly introduce previous and representative research on expectations in sHRI.

In chapter 3, I present the methods I have used in my thesis to investigate my two research questions. The methods include the theoretical development of an evaluation framework, two literature studies, one empirical study, and one survey.

In chapter 4, I summarize each paper that is included in this thesis. I also summarize workshop contributions that relate to ethical considerations of users' expectations of social robots.

In chapter 5, I present the major findings in this thesis. These findings are divided into the two research questions.

In chapter 6, I discuss the findings identified in this thesis and put them into a bigger context of the HRI field and the implications of using social robots in society. I also offer a set of guidelines for understanding and reducing the social robot expectation gap. Then, I discuss the limitations of my work and present future directions this research topic can take. Finally, I offer some concluding remarks.

BACKGROUND

CHAPTER 2
BACKGROUND

In this chapter, I present the research field of HRI, including the field's development and the sub-field of sHRI as well as its relation to the field of UX, which offers context to why and how expectations play such an integral part in human-robot interaction. Then, I define and present the concept of expectations and how it relates to sHRI. Because expectations is such a broad concept, there are other concepts and theories that relate to expectations which are presented in this chapter. Lastly, I present previous research on expectations in sHRI.

2.1 HUMAN-ROBOT INTERACTION

The field of HRI aims to understand the interaction between human and robot – it is the study of the behaviors, feelings, and opinions that individuals have towards social robots. The goal is to create interactions that are acceptable and meet the social and emotional needs of an individual (Dautenhahn, 2013). The HRI field is relatively new, with its conception around the first HRI conference IEEE RO-MAN in 1992 (Dautenhahn, 2007b; Goodrich, Schultz, et al., 2008). The HRI field consists of researchers with varying backgrounds, including computer science, artificial intelligence (AI), engineering, psychology, philosophy, cognitive science, human factors, HCI, UX, anthropology, linguistics, human-animal interaction, and other disciplines (Winkle et al., 2023). The shaping of HRI is therefore influenced by different ideas of what it should look like (Dautenhahn, 2007b; Weiss, 2016; Lindblom and Andreasson, 2016; Lagerstedt and Thill, 2020). Many HRI researchers agree that the field should keep its interdisciplinary nature while still finding common ground when conducting research (Baxter et al., 2016). Thus, due to the nature of being a growing interdisciplinary research field, HRI is facing several challenges, including building a foundation of frameworks, terminologies, theories, models, methods, and tools (Baxter et al., 2016; Lagerstedt and Thill, 2023). It would be beneficial to consider and incorporate research from a wider outlook that may challenge and enhance existing frameworks and embark on new frontiers within the field. It should be noted, however, that the broader

outlook also implies that researchers within the field need to adopt a broader set of literature, theories and and methods (Lindblom and Andreasson, 2016).

Having researchers from varying backgrounds results in different motivations for approaching the field of HRI. Roboticists might aim for building advanced robotic systems for real-world applications (e.g., collaborative robots that are able to assist human operators in manufacturing) (Goodrich, Schultz, et al., 2008). Cognitive science and AI researchers might want to implement complex autonomous systems, using robots as embodied or physical test beds, or proofs of concepts of such artificial cognitive systems (Aly, Griffiths, and Stramandinoli, 2017). HCI and UX researchers might focus on the user by creating successful interactions and striving for good interaction quality with robots (Lindblom, Alenljung, and Billing, 2020). Psychologists and anthropologists might use robots as tools in order to deepen the understanding on fundamental issues of how humans interact socially and communicate with interactive systems (Kahn Jr et al., 2007).

Although there are several characterizations and definitions of HRI, there appear to be three broader views of the field: robot-centered view, robot-cognition centered view, and human-centered view. These views have notably been summarized by Dautenhahn (2007) in her conceptual space of HRI. Meister (2014) later added an AI view, however, I focus on Dautenhahn's (2007) views. All three views are considered in HRI and are not mutually exclusive. A research project might have a researcher go into the technical aspects of voice recognition in a social robot whereas another researcher might focus on how participants interpreted and understood the interaction. Sometimes these views are combined in the same study.

From a robot-centered view, there is an interest to develop various kinds of robots. Robots that had been used until the early 2000s were mostly handled by professionals such as developers and researchers in laboratory settings with little need for the robot to be social because they were mainly operated without direct human interaction (Breazeal, 2003). Therefore, this view has been the most developed and studied within the HRI field because the field was created from this tradition (Dautenhahn, 2007b). Although there has always been an interest in human-like robots, the interest has grown significantly since the start of the HRI field. From this technical point of view, Goodrich and Schultz (2008, p. 210) framed HRI in the following way:

> *Taking a very broad and general view of HRI, one might consider that it includes developing algorithms, programming, testing, refining, fielding, and maintaining the robots. In this case, interaction consists primarily in discovering and diagnosing problems, solving these problems, and the reprogramming (or reiting) the robot.*

Beyond the robot-centered view, the robot-cognition centered view is based on the interest to create intelligent systems for robots (Dautenhahn, 2007b). This view focuses on the social robot and the cognitive and social skill it can possess. Social robots are defined as robots with anthropomorphic and zoomorphic features that are intended to work in social interactions with people (e.g., figure 2.3, 2.1, and

2.2). The objective is using a social robot as an intelligent system, with research challenges regarding the development of cognitive robot architectures, machine learning, and problem solving. As further explained by Dautenhahn (2007, p. 684):

> *Defining socially acceptable behaviour, implemented, for example, as social rules guiding a robot's behaviour in its interactions with people, as well as taking into account the individual nature of humans, could lead to machines that are able to adapt to a user's preferences, likes and dislikes, e.g., an individualized, personalized robot companion.*

Thus, social robots need to be able to perform and learn tasks in a flexible and adaptive manner and should be able to be personable. A social robot that is truly personalized to its user, and can be called a companion, needs to have a specific set of characteristics. The social robot needs to have a constant interaction with the user (including social skills) while serving several functions. Therefore, the robot should be viewed as a unique companion designed to establish a social bond with its user. All social robots have, in one way or another, been developed from a robot-cognition centered view because they are attempting to be social robots equipped with a complex interaction pattern, e.g., speech, voice recognition, response to touch, face recognition and other features that contribute to establish a dynamic interaction between the human and the social robot.

Although researchers working in the robot-centered view focuses on creating complex cognitive and social skills in social robots, typically with the aim to be human-like, it is worth noting that social robots should not be considered *actual* human. There is an ongoing debate regarding the future of AI and whether or not it will be able to possess consciousness, usually referred to as strong versus weak AI (Cole, 2023); however, in this work I focus on the social robots that are possible today, which only is the illusion of life. Even though social robots today may have complex cognitive and social skills, social robots do not possess actual genuine life.

The human-centered view focuses on the individuals using and interacting with the robots. HRI is not exclusively concerned with the *development* of robots, but also the *interaction* that occurs between an individual and a robot, especially social robots. As explained by Dautenhahn (2013):

> A number of people are interested in studying the *interaction* of people and robots, how people perceive different types and behaviours of robots, how they perceive social cues or different robot embodiments, etc. The means to carry out this work is usually via 'user studies' (...) Such research strongly focuses on humans' reactions and attitudes towards robots.

Thus, the objective for an HRI researcher with the human-centered view is to investigate and analyze how humans interact with robots. The focus is therefore not on what is happening inside the robot but rather on how humans understand and interpret the interaction with the robot. In these types of studies, it is common to use social robots (i.e., robots that can perform a task that is comfortable to the human) (Dautenhahn, 2007b). From a human-centered perspective, the

Figure 2.1: The social robots Nao by Aldebaran (2023)

primary concern is not making the robot inherently social, but having the robot be *perceived* as social. My doctoral work has focused on the human-centered view, specifically in relation to what we expect of these social robots.

2.1.1 THE HUMAN-CENTERED VIEW: SOCIAL ROBOTS, sHRI, AND METHODS

Studying HRI from the human-centered view usually involves the use of social robots that are able to interact socially. Studying the social aspects of human-robot interaction can further be categorized as the sub-field sHRI (where the socio-cognitive view oftentimes also falls under). Social interaction plays a key role in sHRI and for social robots. Social robots can be applied to various settings (e.g., as research platforms, toys, educational tools, and therapeutic aids) (Fong, Nourbakhsh, and Dautenhahn, 2003; Jost et al., 2020). Social robots can also take on different social roles, such as partners, companions, peers or assistants. Most social robots are meant to act human-like to fit these roles. There are also pet-like social robots that serves as a companion for humans. Consequently, robots can be both, so called, anthropomorphic and zoomorphic. As previously mentioned, robots that are used in sHRI and have anthropomorphic or zoomorphic traits are defined as social robots, though there are several other terms for such robots (Breazeal, 2003; Fong, Nourbakhsh, and Dautenhahn, 2003; Thrun, 2004; Meister, 2014). For example, according to Brezeal (2003; 2004), sociable humanoid robots (as she has chosen to call them) have changed the way we think of autonomous robots because these actively interact with humans in the same environment as humans, and are not remotely controlled. Instead, these robots

"can communicate in a manner that supports the natural communication modalities of humans. Examples include facial expression, body posture, gesture, gaze direction, and voice" (Breazeal, 2003, p. 120). In addition, these robots are able to engage humans' expressive social cues.

Figure 2.2: The social robot Furhat by Furhat Robotics (2023)

Social robots are thus used with the purpose to behave like humans, and there are a number of ways in which we understand other human beings in an interaction. We can use verbal and body language, gestures, facial expressions, and other features that make up a typical human-human interaction (HHI). Human behavior and interaction have been widely researched and continue to be investigated in research fields such as psychology, sociology, cognitive science, philosophy, HCI, and UX. Because human behavior and interaction have been so rigorously researched for a long time, this research can be used as a foundation for modeling interaction behavior in a social robot. Thus, the field of sHRI is, in part, based on much of the research from HHI (Jost et al., 2020). Hence, there is an underlying assumption that the interaction between people represents an ideal for human robot interaction and that knowledge about HHI can be transferred to HRI. Moreover, the development of social robots is often seen as an attempt to have humans interact with technology in *human-like ways*, and robots are therefore equipped with human-like modalities, which separates social robots from other kinds of technology like computers or smartphones. Because social robots are purposely designed to be human-like, they can be viewed as a metaphor for a human, where social robots are created in the image of humans. Social robots have, therefore, great potential to create smoother interactions between humans and technology.

It is worth noting that this strong connection to HHI is an assumption and not always correct, yet it is an integral part of sHRI research.

Various kinds of user studies are common practice in the human-centered view and thus empirical work is prevalent, with many ways to gather data. Questionnaires are popular in the human-centered domain, and there are numerous kinds depending on what is being assessed (Rosén, Lagerstedt, and Lamb, 2023). For example, The Negative Attitudes towards Robot Scale (NARS) is a Likert scale used to measure negative attitudes. Attitudes are a psychological construct that can be defined as a *"relatively stable and enduring predisposition to behave or react in a certain way toward persons, objects, institutions, or issues; its source is cultural, familial, and personal"* (Nomura et al., 2008, p. 442-443). Attitudes, in general, is a popular research topic in sHRI (Jost et al., 2020). Users' feelings and attitudes towards social robots can also be assessed using qualitative methods (Lindblom, 2015). These include in-depth and long term studies, focusing on the the meaning of the interaction. In addition, several measures, both qualitative and quantitative analysis, can be combined in order to collect relevant data to be analyzed.

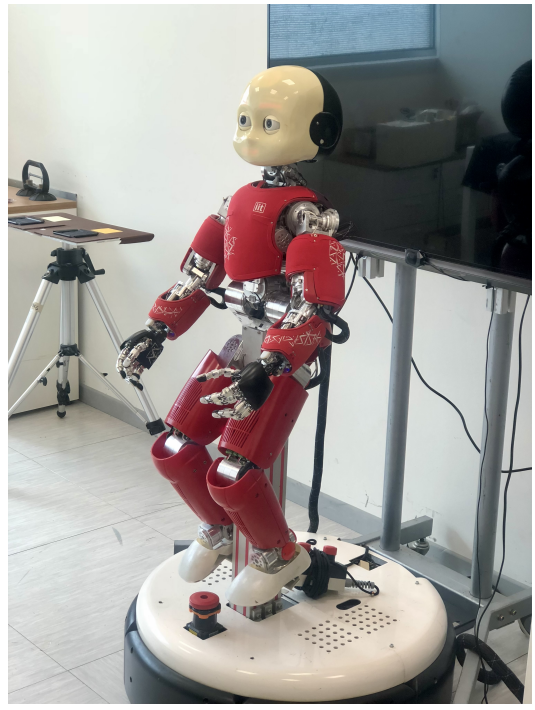

Figure 2.3: The social robot iCub (Metta et al., 2010)

Because the focus is on the human, the way the social robot achieves certain social skills can be manipulated. One way is to remotely control the social robot via the Wizard of Oz (WoZ) set-up where the social robot is controlled by the researcher without the participant's knowledge. The WoZ set-up is usually used in order to act out certain behaviors that are not yet possible with autonomous solutions (Riek,

2012). It is typically used to evaluate an iteration process with a human user before a technical solution is fully implemented. The ethical implications of using this method are an ongoing debate in the HRI community because WoZ is viewed as deception by many researchers because the participants are not informed about the purpose of the study or the capability of the robot (Duffy, 2002; Riek, 2012; Sharkey and Sharkey, 2021).

Another method employed in the human-centered view is the Theatrical Robot (TR) method where humans play social robots (Dautenhahn, 2013). The advantages of using TR include being able to prototype a robot in the early stages of its development. It can also be useful if the objective is to study how humans would react to highly realistic social robots, which are not a reality today. The TR method is useful in therapy for children with autism because research shows that, while they express avoidance to strangers, they respond positively to computer systems and social robots (Robins, Dautenhahn, and Dubowski, 2004). For example, Robins et al. (2004) investigated how children with autism react socially to a mime artist, who either has a socially robotic appearance or a human appearance. Results showed that the participants did indeed react more positively when the mime artist was acting like a social robot, using the TR method.

Both WoZ and TR are established methods in human-centered HRI research, especially when there is a lack of social robots with these social skills implemented. However, the question of whether or not these methods constitute actual human-robot interaction and not human-human interaction is up for debate. The use of these methods also begs the question of what counts as natural human behavior? As stated by Dautenhahn (2013):

> There appears to be little argument to state that a particular behaviour X is natural for a robot Y. Any behaviour of a robot will be natural or artificial, solely depending on how the humans interacting with the robot perceive it. Thus, naturalness of robot behavior is in the eyes of the beholder, e.g., the human interacting with or watching the robot; it is not a property of the robot's behaviour itself.

These methods highlight the complicated nature of studying and creating social robots, because they can be interpreted as humans to the point that they are sometimes even played by a person (unbeknownst to the user). The question of what people expect from these social robots can therefore be raised. Whether people interpret them as human-like or as machines (Alač, 2016). At its core, the sHRI field needs to further understand the experiences of the user interacting with the social robot, from a human-centered view. The field of UX could be useful in this pursuit.

2.1.2 USER EXPERIENCE IN HRI

The UX field is considered crucial in the development of interactive systems, products, and services. This progress has resulted in higher expectations and demands on the interaction quality of digital artifacts, which goes beyond the more traditional aspects of usability and acceptance that are usually addressed within

the HCI field (Hartson and Pyla, 2018). Although a UX perspective within the HRI field does exist, and is getting more traction in the literature, it is often overlooked (Lindblom and Andreasson, 2016; Tonkin et al., 2018; Alenljung et al., 2019; Khan and Germak, 2018; de Wit et al., 2019; Shourmasti et al., 2021). The UX field aims to design for and create a positive user experience. User experience is defined by ISO (9241-210:2010, 2.15) as:

> User's perceptions and responses that result from the use and/or anticipated use of a system, product or service. (...) User's perceptions and responses include the users' emotions, beliefs, preferences, perceptions, comfort, behaviors, and accomplishments that occur before, during and after use.

Accordingly, trying to guarantee a certain user experience, rather than designing for a certain user experience, is not attainable because UX encompasses the subjective and personal inner state of individual human use. Subsequently, designing for high interaction quality with the intended users and the usage context in mind makes it possible to impact the user experience (Lindblom, Alenljung, and Billing, 2020). User experience can therefore be viewed as a result of the quality of the human-technology interaction which embraces a holistic perspective (Hartson and Pyla, 2018). Thus, understanding, researching, and designing for a better interaction quality and improving this experience is the focal point of the UX field.

UX researchers and practitioners aim to analyze, design, evaluate, and implement artifacts with the users' experience in mind (Hartson and Pyla, 2018). Thus, a goal in UX is to evaluate how well a user can carry out a task with an artifact in a certain context, and subsequently design the artifact through an iterative process. Hassenzahl and Tractinsky (2006) stressed three main factors that make up UX; (1) the internal states of the user; (2) the designed systems characteristics, such as its purpose, complexity, and usability; and (3) the context and environment for the interaction. Being aware of these factors allows designing for a positive user experience. Accordingly, these main factors can be applied in HRI research; for example, user's expectations (internal state) of a social robot (designed system characteristic) in an assisted living facility (context).

User experience includes pragmatic and hedonic qualities (Hartson and Pyla, 2018). Pragmatic quality refers to fulfilling the do-goals of the users, meaning that the social robot should support the users to reach task-related goals in effective, efficient, and secure ways. Pragmatic quality is similar to the well-known HCI terms of the usability and usefulness of interactive systems. Hedonic quality refers to fulfilling the be-goals of the users, meaning that the social robot should address the psychological and emotional needs of the users. Similar to other interactive systems developed for human usage, interactions with social robots evoke different kinds of feelings with different kinds of intensities. User experience is thus a subjective phenomenon of the interaction quality (Lindblom, Alenljung, and Billing, 2020).

From a UX perspective, a social robot is seen as a tool to achieve a certain goal in a certain context of use. The goal, therefore, defines what the role of the social robot should be and what kind of tasks it should carry out to achieve that goal, based on

its end-users and the particular usage context (Alač, 2016). On the one hand, if the aim of a robot is to serve as a companion for older adults with the goal for these users to experience being less alone; the robot should, in the best of worlds, fulfill some identified social needs and would be expected to exhibit social behaviors. The robot should be designed so it is perceived as an artificial agent that is able to interact socially, which is a cornerstone for sHRI research (Dautenhahn, 2007a). On the other hand, if the aim of a robot is to vacuum floors in a home with the goal for the user is to experience less stress over cleaning, it would not be expected to behave socially to the same extent. Thus, because the robot's goal is to clean and not interact with the user, the robot does not need to be designed as a social robot perceived as an artificial agent. There still might be certain interactive features of the robot, such as notifying the user that the robot is done with vacuuming, but the expectation of social interaction and human-likeness will be lower than with a social companion robot.

The UX design and development process always starts with the user's preferences, needs, or motives - then it builds and designs the artifact/tool after the UX goal(s) it is meant to fulfill (Hartson and Pyla, 2018). It is worth noting that, although from a UX designer perspective, the robot is a thing that is used to fulfill a goal, the designer can design the robot to be perceived as an artificial agent (Alač, 2016). It is therefore important to understand what users will expect of the social robot in order to design for a certain user experience. As mentioned, social robots are quite different from other technological artifacts because they are purposely designed to look and behave like humans to create expectations of its social capabilities. Although anthropomorphism has been discussed with other artifacts like computers (Reeves and Nass, 1996), social robots takes anthropomorphism to its edge (Duffy, 2002). The agential and social perspectives of social robots have developed into the, so called, CASA paradigm (i.e., Computers As Social Actors)(Nass et al., 1993; Lee and Nass, 2010). Within this paradigm, social companion robots are considered as relational artifacts.

Another important aspect of UX, sometimes overlooked in sHRI research, is the temporal dimension of interaction. It's crucial to consider the temporal perspective, as its absence can impact overall user experience negatively. In the abovementioned ISO definition of UX, user experience has an obvious temporal aspect in which expectations play a central role. A key aspect of user experience occurs during the actual interaction with a system; however, this is not the only relevant perspective to consider (Roto et al., 2011). Users are also affected by experiences before their first encounter with a system. Such experiences are created from preconceived notions or existing exposure from related systems advertisements, presentations, or demonstrations. The exposure may also come from media and movies, or other people's opinions. In the same way, users' experiences may exist after the actual usage situation, such as reflecting on previous usage and previous expectations, or through the impact from other users' assessment of using the system which may alter the actual user experience (Roto et al., 2011).

There are many methods in UX that involve temporal aspects that can be applicable to HRI. For example, the UX design life cycle, or UX wheel, which is a model of the core activities in UX (Hartson and Pyla, 2018). The UX wheel consists of four

iterative steps: Analyze, Design, Prototype, and Evaluate. The main purpose of the UX wheel is to ensure that the digital artifact supports its intended end-users in a certain context and that the UX goals are fulfilled. The UX wheel provides support to systematically study how user expectations have an impact on the experience of social robots – before, during, and after the interaction at various steps in the UX wheel.

As demonstrated in this chapter so far, social robots create certain social expectations that affect human-robot interaction. This is oftentimes overlooked in HRI, although it appears to be an underlying variable in HRI. One of the main goals of UX is to consider the users' experiences, including their expectations. It is therefore beneficial to understand expectations in sHRI from a UX perspective. In the next section, I present expectations and how it relates to social robots and existing research in sHRI.

2.2 EXPECTATIONS

It is hard to overstate how large of a role expectations have on our perception, attention, and behavior. Expectations are beliefs regarding the future state of affairs. Humans are able to regulate behavior by, in part, vividly conjuring images of possible outcomes, even in novel situations. Expectations are thus the behavior in the present, made up of the past and the future (Roese and Sherman, 2007).

It is well acknowledged that expectations play a huge part in life for humans. In fact, a position in cognitive science is that cognition is predictive, emphasizing the forward looking aspects of cognition that is imagining a future state of affairs (Kveraga, Ghuman, and Bar, 2007; Bubic, Von Cramon, and Schubotz, 2010; Hohwy, 2013; Vernon, 2014). *"The predictive brain is"* as Clark (2013, p. 229) puts it, *"a restless, proactive (Bar, 2007) organ, constantly using its own recent and more distant past history to organize and reorganize its internal milieu in ways that set the scene for responses to perturbation by external stimuli."* The predictive mind works by forming predictions and testing them (i.e., hypothesis testing) as a way to understand the world and *"getting the world right"* (Hohwy, 2013, p. 2), ultimately navigating it more effectively (Roese and Sherman, 2007). There are strong similarities between what the predictive mind hypothesis says about how cognition works and the higher-level concept of expectations.

The predictive mind is formulated in terms of two perspectives: *bottom-up* and *top-down*. The bottom-up perspective entails processing finer details from a stimulus, forming increasingly complex and stable patterns (Kveraga, Ghuman, and Bar, 2007; Hohwy, 2013). The top-down perspective goes in the other direction, where bigger concepts are processed, with wider pattern connections, working from complex to simple. Top-down processes derive information from context on a global scale, and with gist information (i.e., capturing the essence of complex information). Top-down processes derive from previous experiences with the world in order to predict future outcomes. These two processes work simultaneously from both the lower levels and the higher levels of cognition towards each other until recognition is achieved, followed by semantic analysis and object name

information (Kveraga, Ghuman, and Bar, 2007; Hohwy, 2013). When a mismatch occurs between the two processes, an iterative process initiates where the higher-level predictions are matched with the lower-level predictions and are refined over a trial-and-error process until the issue is resolved (Kveraga, Ghuman, and Bar, 2007). This process should, over time, decrease as the prediction becomes refined during the trial-and-error process. The process works to reduce error, i.e., reducing the gap between the predicted outcome and the actual outcome. The predictive mind does, therefore, not stabilize as long as error signals are present in the prediction and actual outcome processes.

Kveraga et al. (2007) proposed three main components of the predictive mind, namely, association, analogy, and generation of predictions. Associations refer to the accumulation of experiences that become linked together over re-occurrences. Analogy refers to new input being mapped to existing representations in memory, and inferences can be made to similar situations. The generation of predictions refers to the predictions we can create based on association and analogy. The predictive mind, with these three components in mind, can be useful when explaining social interactions. For example, when forming first impressions of someone, we might make associations between the individual and someone we already know based on similar looks, and then we project certain personality behaviors onto that person based on the person you know via analogies (Kveraga, Ghuman, and Bar, 2007). Thus, we can expect certain outcomes of the interaction based on our generated predictions. A similar process, the expectancy process (Olson, Roese, and Zanna, 1996), has been identified and described in the context of expectations which is a much higher-order cognitive process, and more in line with the top-down perspective of the predictive mind (Kveraga, Ghuman, and Bar, 2007). The expectancy process (Olson, Roese, and Zanna, 1996) is especially useful in understanding the role and relevance of expectations in sHRI.

2.2.1 DEFINITIONS

Much of this work is inspired by the research done on expectancy in social psychology, and I, therefore, ground my definition of expectations in work from this research (e.g., Olson, Roese, and Zanna, 1996; Roese and Sherman, 2007; Borg Jr and Porter, 2010). Olson et al. (1996) as well as Roese and Sherman (2007) defined expectations as beliefs regarding the future state of affairs. Beliefs are defined as taking something to be true or accepted (Schwitzgebel, 2019). Expectations can be derived from beliefs; however, not all beliefs are expectations because not all beliefs are future-oriented. For example, a person's belief that a soda machine cools beverages will generate the expectation that a soda from the machine will be refreshing in the summer heat. Moreover, beliefs are persistent through situations and events, whereas expectations are typically related to a specific instance. Once the state or event that is subject to the expectation occurs, that specific expectation has run its course. Beliefs generate a unique expectation for each specific event. In turn, a unique expectation may have an effect on the overall belief over time. Moreover, some expectations are based on knowledge, but not all knowledge is related to expectations. Knowledge is commonly defined as justified true beliefs,

Table 2.1: Definitions for expectation and associated terms

Belief	Taking something to be true or accepted
Knowledge	True justified beliefs
Expectation	Beliefs regarding the future state of affairs
Expectancy	The act of expecting
Anticipation	Positive or negative emotions regarding the future state of affairs
Prediction	The act of estimating the likelihood of a specific outcome
Certainty	The estimated likelihood of a specific outcome
Assumption	Taking something for granted, or regarding it as true, in face of uncertainty

though not without debate (Steup and Neta, 2020). The overlap between expectations and knowledge occurs when expectations are based on the knowledge that some thing is true. For example, a person *knowing* from direct experience that a soda machine needs coins to dispense a soda, will generate the expectation that a soda will come out when a coin is inserted.

There are also other related terms and concepts such as anticipation, assumption, belief, certainty and prediction, presented in table 2.1. In one way or another, these terms relate to the concept of expectations and have more specific connotations. For example, anticipation is similar to expectations but with emotional connotations, like suffering or enjoyment, of future state of affairs. Expectations alone do not have such emotional implications. There may be a subsequent emotional effect of expectations, but it is not tied to emotional implications until the expectation is confirmed or disconfirmed. Anticipation can also refer to the bottom-up perspective, as the 'predictive mind' theory of cognition (Hohwy, 2013).

2.2.2 THE EXPECTANCY PROCESS

The expectancy (expectation) process (figure 2.4) by Olson et al. (1996, p. 231) presents the mechanisms underlying the process of expectations. In this section, I present the steps an expectation takes from its formation to its consequences.

Sources of expectations

Olson et al. (1996) described three sources of beliefs that are the basis for expectations: direct experience, indirect experience, and inference. Beliefs based on *direct experience* are generated through sensory experience, including perception and emotion. These sources of beliefs are considered more trustworthy and held with a higher certainty (Olson, Roese, and Zanna, 1996). Beliefs based on *indirect experience* are generated through interaction with others. These kinds of beliefs are usually less trustworthy and held to a lower degree of certainty (Olson, Roese,

and Zanna, 1996). From the predictive mind perspective (Kveraga, Ghuman, and Bar, 2007), the top-down perspective of prediction is based on previous experience which would make this process fit under these two generated beliefs of the expectancy model by Olson et al. (1996). Lastly, beliefs based on *inferences* are generated through reasoning regarding other beliefs, such as abduction, induction, or deduction. Inferences, as per the expectancy process (Olson, Roese, and Zanna, 1996), could be likened to associations and analogy from the framework by Kveraga et al. (2007) related to the predictive mind theory.

Properties of expectations

Expectations can vary along four different dimensions: certainty, accessibility, explicitness, and importance (Olson, Roese, and Zanna, 1996). *Certainty* refers to the confidence level that an individual has that an expectation will be true. *Accessibility* refers to the likelihood of an expectation to be activated. *Explicitness* refers to what degree an expectation is consciously generated, ranging from implicitly to explicitly generated. Some expectations are implicitly assumed, usually related to the degree of certainty, whereas other expectations are consciously and explicitly thought about. Lastly, *importance* refers to the expectation's significance, with higher importance having a higher impact on needs and motives.

Consequences of expectations

The rest of the expectancy process relates to the consequences of expectations (Olson, Roese, and Zanna, 1996). Expectations can be confirmed or disconfirmed. Based on confirmed and disconfirmed expectations, different outcomes will be elicited. Confirmed expectations often lead to positive affective responses, they often are handled implicitly (and with ease), and the expectations are upheld with greater certainty. Confirmed expectations may produce secondary affect (positive or negative), given the inferences are made after the confirmation of the expectation. If, for example, a person has been feeling sick for a while and expects to receive bad news from the doctor and then this expectation is confirmed, the instance where it is confirmed is still viewed as a positive affect in this model. However, the actual negative feeling of being sick is the secondary affect, as explained by this process. Arguably, the initial positive affect will be overshadowed and not noticeable by the secondary affect of the news of being sick.

In contrast to confirmed expectations, disconfirmed expectations will often lead to negative affective responses and are handled explicitly because they are surprising and need further processing in order to make sense of what went wrong. Secondary affects may also occur for disconfirmed expectations. In the case of the person feeling sick and expecting bad news from the doctor and then when this expectation is disconfirmed, the instance where it is disconfirmed is still viewed as a negative affect in this model. The actual positive feeling of not being sick is the secondary affect, as explained by this process. Again, arguably, the initial negative affect will be overshadowed and probably not noticeable by the secondary affect of the good news of actually not being sick.

As expectations relate to predictions of the future, it can lead to disappointment and negative affect when an expectation is disconfirmed (Olson, Roese, and Zanna,

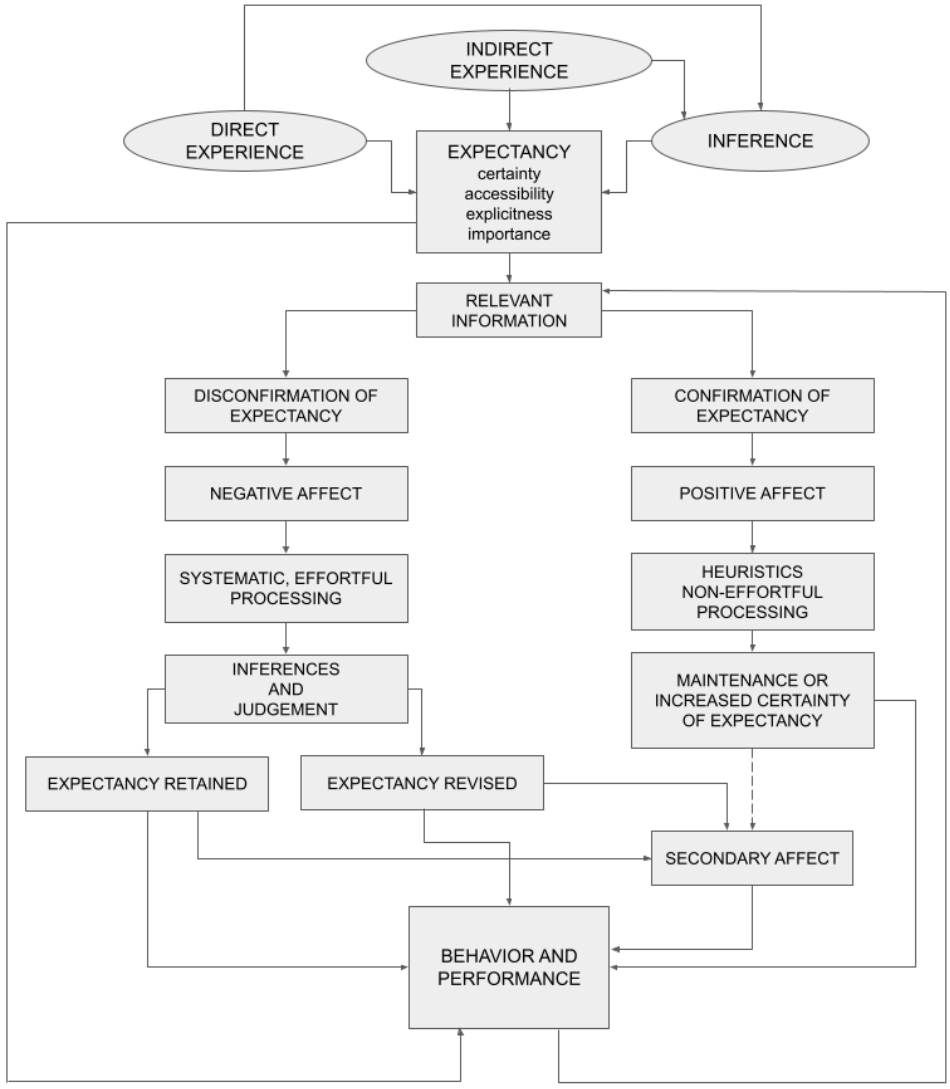

Figure 2.4: A model of the expectancy (expectation) process by Olson et al. (1996, p. 231), modified (with permission) for clarity.

1996; Roese and Sherman, 2007; Borg Jr and Porter, 2010). A constant pattern matching is unfolding between the previous outcome, the expected outcome, and the actual outcome. This is sometimes is called fluency processing (Borg Jr and Porter, 2010). This process of expectations analysis is mostly carried out implicitly and happens swiftly with low cognitive effort. When an expectation is confirmed, it usually goes unnoticed. When an expectation is disconfirmed, however, it usually produces a negative affect and is brought to awareness. Humans generally like a predictable world, and when this is not the case, it is usually perceived as unpleasant (Olson, Roese, and Zanna, 1996). A parallel can be drawn to the predictive mind theory, as a constant matching of predictions and outcomes occurs in order to improve predictions over time, including how ”manual control” kicks in when a prediction does not match the actual outcome (Hohwy, 2013).

Once the process of figuring out what went wrong in a disconfirmed expectation begins, the expectations are evaluated, i.e., inference and judgment occurs for future expectations. There are two outcomes of this evaluation, namely, revising the expectations or retaining the expectation. The revised expectation is when an expectation is adjusted, meaning that the expectation is updated so it aligns better with the event and thus should lead to confirming the expectation over time. Retained expectation, however, is when an expectation is kept while ignoring contradictory information. The disconfirmation may be uncomfortable, but it may not lead to revising the expectation. This can happen for numerous reasons. An expectation may be retained due to the robust view the person may have, e.g., based on norms and stereotypes (Olson, Roese, and Zanna, 1996). We can see this phenomenon in recent events, where people refuse to get vaccinated despite the myriad of scientific backing for its benefits and safety, which has become even more prevalent during the COVID-19 pandemic (Nasr, 2021; Ahmed, 2021). The cognitive dissonance theory states that, if faced with contradictory information, unpleasant feelings will be experienced and ultimately the view (or expectation) that is least likely to change will be retained (Festinger, 1957; Roese and Sherman, 2007). This means that, if one has a robust view of something, that expectation will be retained despite being faced with lots of discrepancies.

The last consequences of an expectation are behavior and performance. This can mean virtually any behavior that an individual may exhibit, which means that it has a major role in interactions because expectations shape the way we behave (Olson, Roese, and Zanna, 1996). Typically, people will behave in accordance with the particular content of an expectation. This is demonstrated by the example of not getting vaccinated due to fear of getting sick from it. The expectation that the person will get sick leads to the behavior of choosing not to be vaccinated. Another example is from a classic psychology experiment that studied learned helplessness, which is a behavior exhibited by people who keep failing at a task which leads to the expectation that they will keep failing and therefore gives up trying to achieve a desired outcome (Seligman, 1972). This concept is related to self-fulfilling prophecy, where an individual’s expectation influences their behavior, causing an outcome that aligns with the initial expectation, affirming its accuracy (Olson, Roese, and Zanna, 1996). People behave in ways that are consistent with their expectations. This relates to the types of evidence that are used to generate expectations. For

example, expectations generated from direct experience are usually held with greater certainty, which implies that people without direct experience should have an easier time adjusting their behavior because they have less strongly supported expectations.

Another aspect of the behavioral outcome of expectations is hypothesis testing which is the deliberate behavior a person engages in in order to test an expectation (Olson, Roese, and Zanna, 1996). If a person puts a coin in a soda machine and no soda comes out, the person may alter their behavior in order to figure out what went wrong. The person may think that the soda is stuck in the machine and thus tries to shake the machine with the expectation that the soda will fall out. If it does not come out, that expectation is disconfirmed, and thus the person could move on to other hypotheses and test them until their expectations are confirmed, ultimately getting the soda.

2.2.3 SOCIAL EXPECTATIONS

So far, I have presented expectations on a general level, including the process expectations take from forming expectations to the outcome of those expectations. I will now turn to social expectations, which are, arguably, the highest level of expectations. The expectations we form of future states are context-specific and situated in the world, which is full of several things, events, and people. To cope and interact with others, a wide range of social capabilities are required (Rudling, 2023). Other people's thoughts, feelings, and actions (i.e., psychological processes) need to be managed in relation to oneself in real-time, including before, during and after an interaction (Heider, 2015). In the case of interaction, humans form expectations of how the person may act or what the person may be thinking or feeling and plan to act accordingly. According to Beavis and Ross (1996), expectations are the bare minimum to make sense of a shared experience and actions between two people. If we replace the machine from the soda machine example with another person bringing a soda, the expectations can be quite different. If one person says that they will bring a soda and fails to do so, the other person may feel let down and subsequently act differently towards that person. If a soda machine fails to give a soda, there would arguably be other emotions, such as possibly frustration with technology and decreased trust towards it, and it is not directed to the social aspects of this kind of interaction. As explained in the seminal work by Heider (2015), there are two kinds of perceptions – *thing perception* and *person perception*. The former refers to how we perceive and interact with inanimate objects whereas the latter refers to how we perceive and interact with other people. The expectations from these types of perception vary because we expect *people* to have internal psychological processes whereas we do not expect *things* to possess such processes. We do not expect that an object can act upon us as people usually do.

If we turn to human-robot interaction (e.g., figure 2.5), the matter becomes more complex. Although social robots are things, they can be perceived as people (Alač, 2016), which makes an individual's expectations of social robots harder to gauge. Social robots are purposely designed to be perceived as people, and thus higher

expectations are formed. In addition, many people lack direct experience with robots, and consequently, their expectations are to a greater extent based on the depictions of robots in media and science fiction (Alves-Oliveira et al., 2015; Rosén et al., 2018; Oliveira and Yadollahi, 2023). Therefore, expectations play a central role in an interaction between humans and social robots as people's expectations of, both other people, and human-like social robots from science fiction, creep into our expectations of real life social robots. In the case of the soda machine example, if we switch the soda machine to a social robot, the expectations generated by the person would likely be different for the particular robot. Although a soda machine and a social robot are both machines, the expectations humans generate are different because social robots are designed to be human-like. When a social robot fails to give a soda to the person, the person may feel emotions, like awkwardness. If the robot *would* give a soda, however, the person might be impressed with the robot. A person would probably not be impressed with a soda machine successfully dispensing a soda.

Although the concept expectation is not as widely discussed in sHRI as in other fields (e.g., cognitive science and psychology), there have been some attempts to highlight this topic (e.g., Lohse, 2009; Lohse, 2010; Meister, 2014; Edwards et al., 2019; Horstmann and Krämer, 2019; Kwon, Jung, and Knepper, 2016; Schramm, Dufault, and Young, 2020; Berzuk and Young, 2023). Moreover, the concept expectation can be found in many sHRI articles, however, the term is mostly used colloquially, in passing, or without explaining the concept.

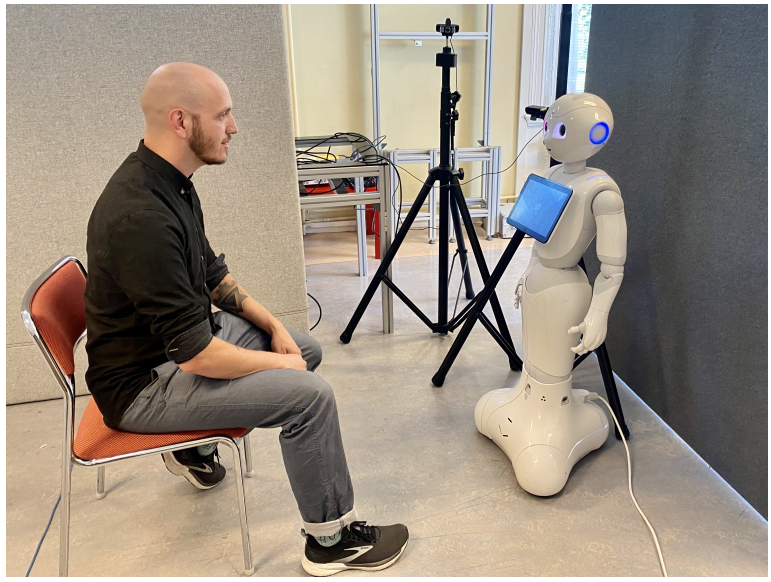

Figure 2.5: A social human-robot interaction, with Pepper by Aldebaran (2023)

Building on the expectancy process by Olson et al. (1996) and putting it into an sHRI context, I have illustrated what could happen in an interaction between an individual and a social robot (figure 2.6) (Rosén, Lindblom, and Billing, 2022).

If the user's expectations are either too high or too low, the expectations are disconfirmed, creating a gap between what the user expects and the display of the robot's actual capabilities. As introduced in Paper III, I refer to this gap as the *Social Robot Expectation Gap*. The diagonal line of the figure 2.6 represents the ideal case, when user expectations and robot capabilities are properly aligned. For example, if a human interacts with a social robot and expects, based on exposure to science fiction movies, that it is capable of expressing emotions and then it fails to do so, the expectation will be disconfirmed in the form of high expectations relative to low capabilities (falling in the blue space of disconfirmed expectations). Alternatively, if an individual does not expect that a robot will be capable of any verbal communications and then it does strike up a conversation, the expectation will be disconfirmed in the form of low expectations relative to high capabilities (falling in the green space of disconfirmed expectations). A conclusion that can be drawn from the expectancy process by Olson et al. (1996), is that we can achieve high interaction quality with robots both with high and low capabilities, given that the expectations are met on the diagonal line. An interaction can go smoothly given that expectations are met, regardless of actual capabilities of the social robot.

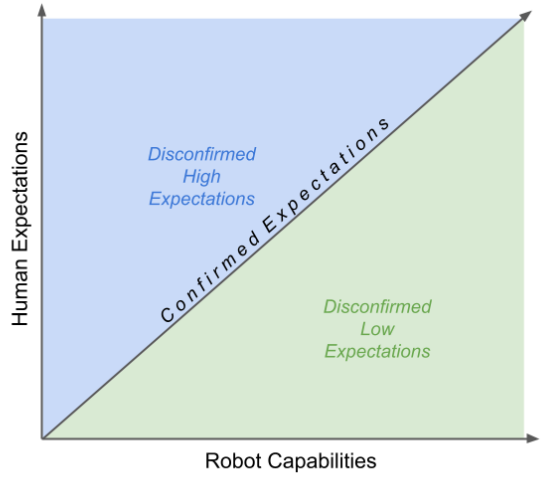

Figure 2.6: The social robot expectation gap, with the two spaces of disconfirmed expectations that occur when a robot's capabilities do not align with the expected capabilities

Disconfirmed high expectations have been proposed previously by Kwon et al. (2016), called the expectation gap. However, the authors did not account for disconfirmed low expectations. Figure 2.6 includes both too high and too low expectations of social robots. Moreover, another similar gap has been proposed by Moore (2017), called the habitability gap. This gap presents the mismatch between the capabilities and expectations of advanced interactive agents, with a focus on voice-based agents (Jokinen and Wilcock, 2017). The authors claimed that, as

flexibility of the system increases so does usability. This happens up to the point where users do not know the capabilities of the system which leads to an inability to properly use the system. Although this gap is certainly related to the social robot expectation gap, it focuses on different aspects (e.g., flexibility and usability, voice-based systems only) of interacting with new technology. Besides the obvious central theme of expectations, the social robot expectation gap, the expectation gap (Kwon, Jung, and Knepper, 2016), and the habitability gap (Moore, 2017) all demonstrate how the interaction is affected by human expectations of artifacts, such as social robots.

2.3 RELATED CONCEPTS AND THEORIES IN sHRI THAT MAKE UP EXPECTATIONS

Although expectations are not always explicitly mentioned, they are prevalent in sHRI research. There are concepts and theories from other disciplines outside sHRI that, to various extents, have been applied for sHRI research and being closely related to expectations. These concepts and theories can be used to explain how people are able to perceive robots as agents with agency, and not just inanimate objects (Heider, 2015; Alač, 2016).

2.3.1 EXPECTING AGENCY IN SOCIAL ROBOTS

As I have repeatedly highlighted throughout this thesis, people generally tend to expect agency from robots even though they are inanimate things. The tendency to do so can be explained in many ways. One of the most popular explanations is the theory of mind. The theory of mind is a psychological theory which describes the capacity to understand others' purpose, intention, liking, belief, emotions, thinking, and goals through their actions (Premack and Woodruff, 1978). A person does not act in the physical world alone - there are other agents that a person will interact with. We cannot look into another person's mind; therefore, we infer their intentions and we expect certain behaviors of them. We do this by observing their actions, creating an internal representation of these actions, and subsequently using it to understand and predict future actions (Mascolo and Bidell, 2020). Thus, we are able to expect certain actions and behavior from another person (Dennett, 1989; Ward, 2015). We also are able to do so with social robots, expecting, representing or mirroring a similar kind of mind to that of our own (Alač, 2016; Clark and Fischer, 2023). These expectations of action behavior can stem from numerous cues exhibited by the social robot, including verbal and non-verbal ones (Gazziniga, Ivry, and Mangun, 2014). An action by a person can occur based on the assumption that a social robot's action is goal directed. There have been several attempts to create a theory of mind in social robots (Scassellati, 2002), aligned with the cognition-centered view from Dautenhahn (2007).

Heider and Simmel (1944) showed that humans are capable of inferring intent in non-living objects. An experiment was performed where participants were asked to observe a short clip depicting circles and triangles moving around inside and

outside a larger square. Despite the fact that the objects were simple drawn figures, the movement combination resulted in participants viewing this as a scene with animated beings (even as persons) equipped with intent. Participants described these figures as being, among other things, scared, terrified, helpless, happy, or glad. The figures appeared to have motivations for their actions, and participants expected certain behaviors and intentions from these objects. This experiment highlights peoples' tendency to infer intent even in the simplest of objects. It is therefore no surprise that similar inference of agency can occur with social robots, especially considering that these robots are designed as human-like.

The intentional stance from philosophy is another theory which describes the tendency to infer agency. It describes how people infer states (i.e., adopting the intentional stance) in other people or non-humans. Dennett (1989) explains that, because humans are overly social beings, we tend to assume intent in others, human or otherwise. This is hypothesized to happen because humans are social in nature which causes certain expectations. This means that we will expect higher intelligence in animals than what is necessarily accurate (Andrews, 2020). There are many similarities between theory of mind and intentional stance. However, theory of mind assumes an intentional stance but the intentional stance does not assume a theory of mind. Theory of mind focuses on inferences of states and how they are reciprocated in other agents, whereas intentional stance focuses solely on the inferences of states whether they are true or not.

The intentional stance has been studied in HRI (e.g., Thellman, Silvervarg, and Ziemke, 2017; Bennett, 2021; Marchesi et al., 2021). Thellman et al. (2017) investigated the role of intentional stance in sHRI. Participants were asked to observe a robot and a person perform the same household task and rate their behaviors. Results show that participants rated the robot and the person similarly in term of mental states, however, the participants were less confident in their rating for the robot. These results may be due to the higher uncertainty related to the expectations of social robots.

2.3.2 ANTHROPOMORPHISM: DESIGNING AND DISPLAYING SOCIAL ROBOTS

Having the assumption that users expect agency in social robots has lead sHRI researchers to aim to achieve acceptable and positive social interactions between people and robots. Creating social interactions where the users expect human-likeness in the robot is one of the key aspects of the human-centered view of HRI (Dautenhahn, 2007a). Achieving the expectation of agency in social robots can be done in several ways, from the way we design the robots to the ways we display them. Fundamentally, all social robots are to some degree designed and displayed anthropomorphically. Anthropomorphism occurs when certain characteristics of an entity or thing leads to the attribution that that entity or thing is like a human, including cognitive and emotional states, as a way to rationalize and understand their behavior (Duffy, 2003; Fink, 2012). It is well-acknowledged that anthropomorphism has a major impact on humans' expectations when interacting with artificial artifacts like social robots (Duffy, 2003). For example, autonomous

motion seems to be enough for children to attribute life towards artifacts (Suchman, 2007).

Designers can use anthropomorphism to design social robots to invite social interaction. Because social robots are intended to work in social environments, designing them with human-like features will help the users to better understand its purpose and use (Fischer, 2011). Anthropomorphic features can be designed in numerous ways. A social robot may have eyes to *see* with and a mouth to *talk* with. The appearance of a robot might thus offer users hints about its capabilities (Fischer, 2011). A robot may also be anthropomorphically described and displayed, including the previous sentence – the robot as seeing with eyes, or talking with its mouth. We describe robots with human features in such a way to make sense of the actions and behavior of the robot. This creates an expectation of not only means for social interaction, but also of agency.

Even though designing for interaction is an effective way to create smoother human-robot interaction, anthropomorphization of social robots comes at a cost, including ethical concerns when deceiving users into believing that the robot is more capable and intelligent than what is actually possible today (Coeckelbergh, 2011; Sharkey and Sharkey, 2021; Winkle et al., 2021). Although social robots can have strong physical resemblances to humans, similarities beyond that tend to be superficial (Duffy, 2003).

Another way we use anthropomorphism is the way we present and talk about social robots. In general, we use metaphors in our everyday language, which has consequences for the expectations we form. As explained by Lakoff and Johnson (2008, p. 5), *"the essence of metaphors is understanding and experiencing one kind of thing in terms of another."* Metaphors help us understand the world around us and is a fundamental characteristic of the human mind. We can say that someone is at a crossroads when having to make a decision even though the person is not literally choosing what road to walk; or someone's smile is magnetic in the sense that people may smile back or have their mood lifted, not that it is literally magnetic. Shakespeare wrote in *As You Like It: "All the world's a stage, and all the men and women merely players. They have their exits and their entrances."* The world and human experience is not literally a play, but it is *like* a play.

A byproduct of using metaphorical language is that it changes the ways humans experience things, and ultimately hides certain concepts that are inconsistent with the metaphor. An argument, as explained by Lakoff and Johnson (2008), is often likened with war – you can win an argument and you can shoot down someone's argument. However, arguments can also be cooperative and a valuable exchange between people, which is usually overlooked because the focus is on the war aspects.

A common metaphor used in Western culture is likening our minds to machines. We can say statements like "my mind isn't operating today" or "I'm a little rusty" (Lakoff and Johnson, 2008). This has also been found to happen the other way around – we use human metaphors to explain machines (McDaniel and Gong, 1982). When we liken machines with humans, we also build up an expectation of the two working in similar ways. This way of arguing lays the ground for

misunderstandings because the human mind and the internal states of a robot are in fact very different in many ways. It may also hide aspects of machines that are mechanical and not comparable to humans. We know, more or less, to what extent the mind as machine metaphor works because we have plenty of experience with our own mind. Most people know less about machines and how they work and therefore it may not be as clear what is a metaphor and what is not. When we say a machine is ”feeling overworked” we mean that the system is overheated, not that it is literally feeling the stress of working hard. When we say that the machine is ”thinking” we mean that the machine is running through lines of code, not that it is literally thinking as a person does. Words are context-sensitive and will mean different things to different people. When we use human metaphors to describe machines, it may be obvious to the computer scientist that this is a mere metaphor. However, someone with little experience with machines and computers may not have the correct context to correctly understand the extent to which the metaphor is merely a comparison. The same kind of language is used for social robots, especially because these robots are designed as human-like. According to McDaniel and Gong (1982, p. 179), many authors in professional journals use metaphorical language when writing about robots:

> When a robot is endowed with these kinds of physical human attributes, it is predictable that people will also credit it with a ”brain” to control them. Rarely does one call the central processing unit, chip, or integrated circuits in an Apple computer a ”brain”

Although this quote is from 1982, it is evident that such issues have not been resolved over time.

Metaphors may also hide many aspects of the machine because we talk about it as like the human mind. In social robots, these metaphors become even more dangerous because social robots are deceptive in their anthropomorphic features (Duffy, 2003; Sharkey and Sharkey, 2021) – both by design (e.g., human-like eyes) and by describing them (e.g., ”the robot sees”). Most people lack first-hand experience with social robots as well, thus relying on indirect experiences which are considered more unreliable (Olson, Roese, and Zanna, 1996). Thus, by using human metaphors when presenting social robots, we build expectations of what they *are* and what they are capable of.

Another way to explain how using anthropomorphic design and presentation creates human-like expectations of things is the term believability, which is a common term in fiction to describe the illusion of life. For example, the magic carpet in Aladdin is an inanimate object that manages to captivate the audience and make them *believe* that it is alive (Simmons et al., 2011; Bates, 1994). In sHRI, believable robots refer to peoples’ tendency to believe the anthropomorphic features as being life-like and thus creating more successful and natural social interaction (Simmons et al., 2011; Moetesum and Siddiqi, 2018). From a believability point of view, these anthropomorphic features can be: the look of the robot, e.g., human-like face; the displayed verbal and non-verbal emotions and mood, e.g., happy and smiling; its made-up backstory, e.g., family background; and its ability to express social cues

and respond to its sociocultural context, e.g., religious expressions (Simmons et al., 2011).

Believability is characterized as the *"strong subjective sense of realism"* and not actual genuine life (Bates, 1994, p. 6). This distinction is important to note, and it inevitably opens up a Pandora's box of discussions on strong versus weak AI (Cole, 2023); however, as mentioned previously, in this work I focus on the social robots that are possible today which only demonstrates the illusion of life. Because believability relies on what the user will believe, despite the reality, expectations play a big part here. In HRI, the illusion of life is built on what kinds of expectations the user has of the social robot. If the user's expectations of the robot is met,i.e., the robot is acting the way people expect, the user will believe the robot as life-like and ultimately accomplish a successful interaction.

2.3.3 THE IMPACT OF SOCIAL ROBOTS IN DESIGN AND PRESENTATION

There are several implications of expecting agency from social robots. An outcome of users' expectations of social robots are the norms we create from these expectations. Norms originate from simplifications of complex concepts and act as guidelines in the real world (Dewey, 1933; Malle et al., 2015; Malle and Scheutz, 2019; Sparrow, 2020). They can thus be viewed as expectations that are held with greater certainty and being stabilized into patterns (Olson, Roese, and Zanna, 1996). Norms, implicit and explicit, determine what is considered typical and desirable (Olson, Roese, and Zanna, 1996). They permeate society at large and affect all kinds of interactions (Pereira, Baranauskas, and Liu, 2015). Norms are historically studied in psychology, anthropology, and gender studies, but have implications for sHRI research (Lohse, 2009; Carlucci et al., 2015). In sHRI, humans have norms about social robots, and these norms guide the design of social robots (Carlucci et al., 2015). Norms can also be held by the designer that is designing the robot. If, for example, a designer aims to develop a social robot for a certain group of people, the designer's norms will affect the user in various ways. Such norms are sometimes explicitly stated, but they are often implicit ones which dramatically increases the risk of people being unconsciously affected in unwanted ways. For example, gendering social robots is a common practice in sHRI (Winkle et al., 2022). This is done both by design and by the users. Male presenting robots are able to reject commands of the user more than female-presenting robots, related to politeness norms found in HHI (Winkle et al., 2022). Female digital assistants are often programmed to be docile and to be tolerant of both sexual- and verbal abuse. The cultural norms of women being subservient are therefore at a risk of carrying over from HHI of sHRI (Winkle et al., 2022). Reflective design is one approach for designers to combat and mitigate negative norms and to encourage positive norms in their designs (Dewey, 1933; Sengers et al., 2005; Bulman and Schutz, 2013; Nyhlén and Gidlund, 2019; Fronemann, Pollmann, and Loh, 2021). In terms of gender norms, Winkle et al. (2022) found that gender bias could be reduced by showing a female presenting social robot protest when it was being abused by users.

Relying on norms can also be useful in HRI. For example, creating robots that match normative expectations is sometimes used to develop trust between the person and the robot (Malle et al., 2020). Trust refers to the *"attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability"* (Lee and See, 2004, p. 51). Trust is considered an important aspect for creating a successful human-robot interaction and is a popular topic in sHRI research (e.g., Lee and See, 2004; Billings et al., 2012; Schaefer, 2016; Ullman and Malle, 2018; Lyons and Guznov, 2019; Natarajan and Gombolay, 2020). There are different ways humans can trust robots. For example, trusting that a robot will be able to carry out its assigned task without error or trusting that a robot will be honest and open in its communication. Trust is complex, but at its core it is the outcome of the expectations people have of robots' capabilities (Zhang and Wei, 2020; Henschel, Hortensius, and Cross, 2020). Social robots are designed to be able to adapt to complex environmental inputs as well as be able to act socially towards people. Social robots are envisioned to handle these tasks in various environments, including the workplace and the home. Therefore, it has been emphasized that a social robot needs to be both tools and teammates for humans, as well transition between the two fluently (Lewis, Sycara, and Walker, 2018). However, as social robots are making the transition from being merely a tool to becoming a teammate, humans need to accept and trust robots more (Billings et al., 2012). Studies show that anthropomorphic features in social robots result in higher trust in the robot (Lewis, Sycara, and Walker, 2018; Natarajan and Gombolay, 2020).

2.4 PREVIOUS RESEARCH ON EXPECTATIONS IN sHRI

To my knowledge, Lohse (2009) was the first researcher to explicitly introduce expectations to the HRI field. The author highlights the importance of understanding the concept of expectation because it aids the design of robots. To illustrate how to investigate expectations, a case study was carried out. In this study, participants interacted with a social robot in a home tour scenario. The participants were instructed in how to use the robot and tasked with guiding the robot through an apartment as well as showing certain parts of the apartment. The scenario included a breakdown in communication due to the robot not being able to perceive the participant properly. Thus there is an interruption in the interaction which the participant needed to figure out and mend. The authors explained this as an unexpected event which would lead the participants to update their expectations of the robot. Afterwards, participants were interviewed and asked to fill in a questionnaire regarding the robot's usability, how much they liked the robot, and attributions made towards the robot. Results show that certain behaviors from the robot created specific expectations of what it was capable of. Moreover, results show how participants updated their expectations over time to adjust to new information, thus showing how expectations are influenced by the robot's behavior. Lohse (2009) concluded that expectations need to be considered when designing robots in order to more efficiently solve tasks in human-robot

interaction.

Fourteen years have passed since Lohse's (2009) publication and much has happened during that time. First, the technical advancement of robots has progressed and, thus, the capabilities of state-of-the-art social robots today should have an affect on users' expectations. Second, the public views of robots have progressed, especially considering the growing concern with AI and its perceived threat to society (Cugurullo and Acheampong, 2023). Lastly, the HRI field as a whole has matured and more focus has been put on the human perspective. There is therefore an identified need to continue the work on expectations in HRI. Below, I present some more recent work on how expectations have been studied in HRI.

Building on the theoretical work on expectations in sHRI, Meister (2014) introduced four steps of expectations, based on action theory from sociology. Meister's (2014) model of expectations is briefly summarized in the following way: the first step includes the person's perception and expectation of the situation; the second step includes consolidating and confirming or disconfirming the expectations from the first step; the third step is creating a social order of generalized expectations, based on the two previous steps; and the fourth step includes these consolidated expectations and forming the perceptions that are used in the first step and the subsequent course of actions. These steps can be viewed as an optimization of how to successfully deal with social situations. Thus, this model specializes in the interaction, rather than just any expectation one may have.

As mentioned in section 2.2.3, Kwon et al. (2016) lifted the paradoxical nature of the idea, that underlies most HRI research, that human-robot interactions are improved the more advanced socio-cognitive capabilities a robot possesses even though this usually increases the users' expectations. The authors explained that an expectation gap may occur due to users' ability to generalize social robots from mental models based on human capabilities and skills. In order to demonstrate this expectation gap, an experiment was conducted with three different agents: a social robot, an industrial robot and a human. Participants were given photos of each agent, followed by a description of a human-agent scenario where the agent interacted with a person in the home or in a factory. Results show that the social robot was viewed as having more capability in the home scenario than the industrial robot. Thus, the expectations of the robots varied based on the robot's anthropomorphic design.

In a second study, Kwon et al. (2016) continued to investigate how robot behavior could alter expectations. The authors studied this by having participants view videos of either a social robot or a person performing a block-building task. The results show greater variance in the expectations of the robot being able to complete the task in comparison to the person. It seems that participants, thus, modify their expectations of the robot based on its behavior and performance. From this paper, it seems like the robot's physical appearance and its actual behavior influenced how people form their mental models. This means that it is likely that humans unintentionally can be manipulated by the robot's appearance to form incorrect mental models of its overall capabilities solely by displaying a sub-set of capabilities that resemble the one's that humans intrinsically have, e.g., speech

and turn-taking abilities. Kwon et al. (2016) concluded that if the expectation gap is not modified, the outcome could paradoxically result in less effective human-robot collaboration when robots' capabilities will be further developed.

Jokinen and Wilcock (2017) stressed the importance of narrowing the high expectations and actual experience in order to foster positive human-robot interaction, including long-lasting relationships and solid trust towards the robot. In order to investigate the expectations that are created in human-robot interaction, the authors conducted a study using the Expectations and Experience (EE) method. The EE method initially was developed for evaluating multi-modal dialogue systems and was intended to assess the mismatch between the users' original expectations before using a digital application and their experience after using it. The authors examined if the tendency reported in some earlier work, which indicated that humans have higher expectations about interacting with a robot compared with their concrete experience of interacting with the robot, also occured in spoken dialogue when a social robot was equipped with human-like natural communication via a dialogue system. They applied the EE method, which is based on the SASSI questionnaire (Hone and Graham, 2000), and applied it before and after the interaction sessions with the Nao robot (figure 2.1). After a short introduction, the participants filled in the questionnaire about their expectations of the upcoming interaction with the robot. Then the experimental session followed in which the participants interacted spontaneously with the robot for 15 minutes and asked questions verbally that they considered interesting, and then the participants filled in the questionnaire again. Results show that the participants' expectations were higher than the actual experience with the robot. Overall, participants had positive experiences although they reported some negative tendency towards not being understood by the robot. The authors also pointed out that the participants with the most experience reported being the most critical towards the robot.

Edwards et al. (2019) investigated how expectations can be changed and confirmed after the initial impressions that participants have. The change was manipulated by first-hand social interaction experience with a social robot or with a person. Results show that participants interacting with a robot experienced positive feelings of affinity and connectedness. Participants that interacted with a person experienced the opposite feelings. The results demonstrate how expectations vary depending on the agent even when the interaction and script were the same. Moreover, it is possible that the limited conversation to accommodate the robot's capability might have affected the interaction quality. It is clear that participants alter their expectations which may result in the tendency to magnify the robot's limited responses and ability to offer behavioral feedback contrasted to human behavior. The implication for the robot confirmation tendency is that participants may anthropomorphize the robot and subsequently create disconfirmed expectations.

Horstmann and Krämer (2019) investigated expectations and the sources of expectations, drawing inspiration from uncertainty reduction theory (Berger and Calabrese, 1974). The uncertainty reduction theory is a communication theory that centers on the initial interaction between individuals before the actual communication process. The theory states that, in order to reduce uncertainty, the person

interacting with another person needs to gather information about that person. The more information that is gathered, the more one person can predict another's behaviors and actions (Berger and Calabrese, 1974). The authors paid attention to the mechanisms found in the uncertainty reduction theory and their impact on humans' expectations of social robots, specifically what people prefer versus what they expect of robots. In their study, data was collected on expectations of previous first-hand experiences of social robots compared to fictional robots via semi-structured interviews and a quantitative online survey, applying a mixed-methods approach. The purpose of the analysis of the collected data was to disconnect expectations from preferences, and the online survey complemented the interviews with a questionnaire. The results show that the expectations of social robots' ability to be integrated into society at large and their personal life are affected by the expectations participants have of fictional robots. In addition, participants with negative feelings towards fictional robots also scored higher on robots' perceived threat to society. In contrast, participants who had more knowledge regarding robots' actual capabilities also showed reduced anxiety towards social robots. These results indicate that people form their expectations from different sources which subsequently affect what kind of expectations people may have.

In another publication, Horstmann and Krämer (2020) studied expectancy violations theory (EVT) in human-robot interaction (Burgoon and Jones, 1976). EVT is another theory of communication that focuses on how people respond when another person violates their expectations. The outcome of expectancy violation is altered behavior and feelings, either positive or negative, which is based on the personal relationship between the interaction partners and how well the violation is received. In order to study this in an HRI context, Horstmann and Krämer (2020) performed a study where a social robot violated expectations, and then measured the participants' desire to interact with the robot by evaluating the overall quality of the interaction. The obtained results indicate that in the situations when the social robot negatively violated the participants' expectations, they evaluated the robot in a more negative manner regarding its competence, sociability, and interaction skills.

Lastly, Manzi et al. (2021) investigated whether or not expectations and interaction quality differed towards social robots depending on physical features as well as the behavior of the social robot. The study was driven by the well-acknowledged assumption in sHRI that human-like design features of social robots are effective in improving the interaction quality. In their study, participants were asked to observe human-robot interactions performed by an experimenter. The experimenter played a card game with the robot, including the robot explaining the rules of the game, as a way to display the robot's speech recognition, response, and movements. The two social robots Nao and Pepper were used. The data was collected to analyze the variability of attributions of mental states, expectations of robotic development, and negative attitudes toward the robot. Results show that both the type of robot and the interaction had an effect of the attribution of mental states, with higher attribution to Pepper. Surprisingly, the type of robot had no effect on the expectations and showed only an effect on the interaction. These

results demonstrate that not all social robots are interpreted equally. Specifically, the design of social robots affects the mental attributions of the robot, and the observation of an interaction affects peoples' expectations of social robots.

The studies presented in this section so far demonstrate how expectations do indeed affect human-robot interaction and interaction quality. There are many dimensions to peoples' expectations of social robots, ranging from the physical design of the robot, and the context the robot is in, to the behavior of the robot. In addition, people's previous experience with social robots also had an affect on expectations, whether it is from science fiction (indirect experience) or from first-hand interaction (direct experience). With this in mind, the results presented in this section seem to fit in relation to the expectancy process by Olson et al. (1996) as expectations are shown to vary depending on participants' previous experiences with robots and that expectations can be altered after exposure in-person to social robots.

The above authors' work stresses the importance of decreasing the disconfirmed expectations to better match with the actual capabilities of current social robots. In addition, the authors lift the issue of studying expectations due to these many different aspects of expectations. Many studies in sHRI lack first-hand interaction with robots, relying instead on video recordings and imagined interactions as a way to measure expectations. In addition, because expectations change over time, there is a lack of studies where the temporal aspect is specifically examined. For the remainder of this section, I briefly present two frameworks created for studying expectations of social robots which aimed to address some of the issues of studying expectations in sHRI.

Schramm et al. (2020) identified the need for a framework to address disconfirmed expectations, and expectancy violations, of social robots from a robot design perspective. The authors argued that people with limited to no previous experience of interacting with social robots create expectations from the robot's immediate physical appearance and behavior, which are formed by indirect experiences of robots depicted in social media and movies. Consequently, the human's constructed expectations of the robot do not necessarily relate to the actual robot's capability, resulting in what they explained as an expectation discrepancy. When people experience this discrepancy, people may experience frustration, disillusionment, and trust reduction. Thus, a better understanding of this expectation discrepancy will aid in grasping the challenges that designers face when developing social robots, especially in the initial encounter when the robot's capabilities are still being revealed. Schramm et al. (2020) therefore created a framework that could address expectations in order to inform designers of social robots.

The framework was developed from workshops and analysis of the HRI research which resulted in two components of developing expectations of social robots. These components are based on robot's appearance and behavior which imply certain capabilities by emitting so-called capability signals. The first component is thus the *emitted* signals from the robot to the user regarding its capability. The second component is the *construct* that humans create based on this emitted signal, which results in a mental model and consequently the user's expectations

of the robot. The authors stressed the importance of understanding these two components separately in order to gain a richer understanding of what design choices affect the expectations.

The emitted signals that the robot can signify of its potential capability are divided into three categories. The first is *life-like* signals, which refers to the design choices that are made to make the robot look like humans, animals, or insects. This includes the physical embodiment of the corporeal form and specific body parts such as the robot's face and hands. The physical embodiment implies specific robot actions like moving, gesturing, showing facial expressions, or manipulating objects. The second is *consequential* signals, which is when people may assume that a robot equipped with specific elements related to specific functions encompasses the related capability. For example, a robot with a camera may be assumed to be able to ”see.” The third is *exposition* signals, which relate to how a robot is introduced to its users, how it introduces itself, and what tasks it is used for. For example, using the terms ”tutor” or ”companion” may impact they expectations of the social robot.

The above signals determine the expectations. Expectations are further categorized after the amount of information that is emitted in order to create certain expectations. The types of expectations are divided into two categories, one focused on *mechanical capabilities* and the other on *life-like capabilities*.

The mechanical category consists of physical ability and computational ability. The physical capability refers to the robot's ability to move in the physical world, including movements, sensing and responding, and performing more advanced tasks like manipulating objects. The computational capabilities refer to the human tendency to apply their understanding of computerized systems to robots, like the processes of saving and retrieving data, performing calculations, or using the Internet. Hence, humans may consequently assume that robots can record data via their channels (”eyes” and ”ears”), memorize facts, and identify humans.

The life-like capabilities are categorized into the following aspects: non-social cognition, social cognition, the emotional system, the social interaction abilities, and pseudo-consciousness. Non-social cognition refers to a robot that, in its autonomous actions, gives the impression of possessing life-like cognitive capabilities, including the ability to learn and interact with its surroundings. Social cognition refers to situations in which humans may believe that the robot is endowed with a social understanding of others' emotions and feelings, the common practices of interacting socially in various social encounters, and accurately using this social information to interact properly with others. The emotional system refers to as the assumptions humans may make about a robot that is equipped with ”emotions” and its ability to both display and experience these emotions, such as a robot that shows a smile being presumed to have a corresponding internal state of happiness. The social interaction abilities refer to the robot's capability to appropriately seem to use language such as speaking, gesturing, and following gaze in socially appropriate ways. Finally, pseudo-consciousness refers to the avoidance of the philosophical discussion that considers the nature of artificial consciousness, albeit they emphasized that humans' mental models of robot capabilities may be

associated with more advanced socio-cognitive abilities like intentionality, self-awareness, and consciousness.

In summary, Schramm et al. (2020) envisioned that their framework can be a useful tool to support ongoing HRI research to investigate and analyze expectation discrepancy and to inform designers based on users' expectations.

Berzuk and Young (2023) further developed and modified Schramm's et al.'s (2020) framework by presenting a theory of how users form their expectations of social robots, focusing on the sources of expectations. The authors based their theoretical modification to the framework on theories from social psychology (e.g., expectancy violations theory and simulation theory) and applied it to a sHRI. In order to synthesize this information, the authors proposed a survey on existing HRI research on expectations. Ultimately, they wanted to create a descriptive framework that can be applied to user studies in order to systematically evaluate design features and how they affect the users' expectations of the social robots.

The framework can be a useful tool for understanding expectations and preventing disconfirmed expectations and also focuses on the design process, with the initial encounter as the main area of interest. The framework, however, does not focus on how to empirically study users' expectations of social robots, specifically the impact of expectations on interactions, before, during, and after the interaction.

METHOD

CHAPTER 3
METHOD

This chapter summarizes the research designs used for the work accomplished in this thesis, which includes the development of the Social Robot Expectation Gap Evaluation Framework (Paper II, Paper III, and Paper IV), two literature reviews (Paper I and Paper V), one experiment (Paper VI), one UX evaluation (Paper IV), and one qualitative analysis (Paper VII). The data collection for the experiment, UX evaluation, and qualitative analysis was gathered from the same empirical study, but the analyses of the collected data differed. The findings from these papers are presented in chapter 5.

3.1 DEVELOPING THE SOCIAL ROBOT EXPECTATION GAP EVALUATION FRAMEWORK

The Social Robot Expectation Gap Evaluation Framework was created with the aim to gain a richer understanding of users' expectations in HRI. This development drew mainly from two fields, namely, social psychology and UX. The development of this framework is presented in Paper II, Paper III, and Paper IV.

A theoretical framework is an aggregation of existing theories and ideas that together create a deeper understanding of a phenomenon (Crick and Koch, 2003; Lindblom, 2015). In addition, putting a theoretical framework into a UX context allows for an evaluation framework that systematically can capture users' expectations in sHRI. As I have stressed in this thesis, HRI deals with inanimate objects that can be interpreted as human-like entities. An underlying assumption in HRI is how HHI insights are transferable to HRI. With this in mind, the method for developing the framework was to synthesize the already existing knowledge on how expectations are described and studied in psychology. The main basis of the framework is, therefore, the expectancy process by Olson et al. (1996), illustrated in figure 2.4. In addition, because HRI deals with users and technical artifacts, another part of developing the framework was to synthesize aspects and methods from UX, specifically UX goals as presented by Hartson and Pyla (2018), and

combine them with research on expectations from psychology. UX deals, to a large extent, with users' expectations with a focus on how time affects the experience and is thus highly relevant to the study of expectations in HRI. After synthesizing the knowledge from psychology and UX, UX goals and metrics were developed that resulted in the framework. The development of the Social Robot Expectation Gap Evaluation Framework was the foundation for the empirical work conducted in this thesis.

3.2 LITERATURE REVIEWS

Two literature studies were conducted towards answering RQ1 as it involves methodology practices in HRI research. Literature review A investigated ethical conduct and reporting in the HRI field, presented in Paper I. Literature review B investigated how NARS (Nomura et al., 2004) is used and reported on in the HRI field, presented in Paper V.

One reason to conduct literature reviews is to examine the scope of the results found in a body of literature for a specific topic; another reason to conduct literature reviews is to gauge *how* these results were achieved, highlighting the methodological practices within the field (Hart, 2018). In this thesis, I have focused on the latter reason, which has guided me in the two literature reviews, my work toward developing the Social Robot Expectation Gap Evaluation Framework, and in the bigger pursuit of answering RQ1.

3.2.1 LITERATURE REVIEW A

This literature review was conducted to investigate how ethical conduct is reported in the HRI field, presented in Paper I. This work is an extension of RQ1 as it involves methodology practices in HRI. The ethical conduct and reporting were based on five principles that are included in several ethical guidelines, including from American Psychological Association (2017), World Medical Association (2018), and the European Commission (2018). Specifically, we sought to answer how often the five ethical principles were explicitly mentioned:

1. Ethical board approval
2. Informed consent
3. Data and privacy
4. Deception (if applicable)
5. Debriefing

Any explicit mention of these principles was noted, however, we did not differentiate further on how these ethical principles were mentioned but rather focused on any type of mention. Because these ethical principles are based on the handling

of participants, we narrowed down the literature review to any type of full-length HRI publication that presented an empirical study with human participants.

To reach an overview of how often the five ethical principles were mentioned in the HRI field in recent years, three major publication outlets were chosen – the 2018 ACM/IEEE International Conference on Human-Robot interaction (HRI), the 2019 IEEE International Conference on Robot & Human Interactive Communication (RO-MAN), and 2018 and 2019 publications from the ACM Transactions on Human-Robot (THRI), which we considered being part of the top tier of contemporary HRI research. For the HRI conference and the THRI journal, all full-length publications were reviewed, which were 49 publications and 31 publications, respectively. For the RO-MAN conference, a random selection of 40 publications was chosen, in order to keep the number of publications reviewed similar across the three outlets. In total, 120 papers were reviewed which became 73 publications after the inclusion criteria were applied, which was to include studies that had an experimental study.

3.2.2 LITERATURE REVIEW B

This literature review was conducted in order to investigate how NARS (Nomura et al., 2004) is used and reported on in HRI research. Much like literature review A, this work is an extension of RQ1 as it involves methodology practices in HRI. Specifically, we sought to answer the following questions:

1. Where are papers using NARS published?
2. How are methods involving NARS reported?
3. How is NARS data analyzed?
4. How are results related to NARS reported?

In order to answer these questions we used the databases IEEE Xplore and ACM Digital Library to find papers in scientific conferences, journals, and book chapters. These two outlets were chosen because they encompass a large majority of the HRI literature, including the main conferences within the field. We used the search term “NARS” or (inclusive) “Negative Attitudes toward Robots” with the starting year 2004 (as the scale was introduced that year), up until 2021. A total of 380 papers were identified, and after an initial round of removing duplicates, 352 papers were left. A second round was conducted in order to remove irrelevant papers (i.e., not experimental studies), resulting in a total of 160 papers. For the analysis phase, two rounds were done in order to answer the above questions. Before the rounds, a random selection of 10 papers were reviewed in order to reach a consensus regarding interpretation of the classifications. The first authors did half each, swapping the order in the second round, and the papers were therefore reviewed by both researchers.

3.3 EMPIRICAL STUDY

The empirical study done for this thesis serves as the foundation for an experiment (Paper VI), a UX evaluation (Paper IV), and a qualitative analysis (Paper VII). The data collection for these papers was done on the same occasion in order to investigate the role and relevance of expectation in a human-robot interaction, related to RQ2.

In order to gain deeper knowledge and richer insights into users' expectations of social robots, several methodological approaches were performed. Although experiments, UX evaluations, and qualitative analyses all are empirical work that focuses on various aspects of behaviors, they differ in the goals, data collection and analyses of the empirical research conducted (Dumas and Redish, 1999). The goal of experiments is to detect if a certain phenomenon is present, where independent and dependent variables are studied. Typically, the data is gathered from a larger sample size and analyzed on a group level. Moreover, usually, several groups are compared to observe whether or not the stimuli affected the participants. Questionnaires and statistical testing are common in experiments. The goal of UX evaluations is to evaluate the task or interaction and to uncover problems. Typically, the data is gathered from a smaller group with a specific user group in mind. The data collected is then analyzed and evaluated in relation to the UX goals. The use of an artifact is evaluated in order to determine if it creates a positive user experience and to identify factors that hinder a positive user experience. The goal of qualitative analyses is to uncover in-depth insights from a certain task, interaction or phenomenon. The data is usually collected on a smaller group by analyzing non-numerical data from field notes, text and/or audio/video recording in order to gain an in-depth understanding of experiences, concepts, or beliefs, by finding patterns or themes that provide findings in addition to those found in the above approaches. I have chosen to explore all three approaches because a comprehensive understanding of expectations is essential to further the sHRI field. This includes delving into the phenomena of expectations, understanding their role and relevance in sHRI, and examining how expectations influence the usability and user experience of social robots.

Below, I present the empirical study, followed by the data collection and analysis related to each methodological approach.

3.3.1 PARTICIPANTS

Participants were recruited through fliers put around the university campus and emails reaching out to faculty and students at the university. A total of 31 (N=31) were recruited, with an age range of 20–54 (M=29); 45% males and 55% females (no one self-described or chose non-binary). A movie ticket was rewarded for their participation in the study. A total of 48% had no previous experience with robots while 52% had some previous experience.

As the interaction was in English but conducted in Sweden, we gathered data on participants first language; 7% of the participants were native English speakers, 55% were Swedish native speakers, 16% were Spanish native speakers, 10% were

native Arabic speakers; the remaining 12% were native German, Portuguese, Turkish, and one participant was native bilingual speaker of Spanish/Arabic.

3.3.2 PROCEDURE

In the study, participants were asked to interact with the robot for 2.5 minutes twice. The participants were told that the purpose of the study was to investigate how robots could work in the home and thus they could ask the robot anything, with a focus to explore what was possible. Thus, the interactions could vary vastly depending on what participants chose to ask. The participants answered subjective measures before, between, and after the interactions. After the interactions and the subjective measures, participant's were debriefed, which included providing information regarding the study's aim and how the robot and the speech function in the robot worked.

3.3.3 THE ROBOT AND TECHNOLOGICAL SETUP

The robot used in the study was the social robot Pepper, created by Aldebaran (2023). The robot could move arms, simulate breathing, and move its head towards the robot. Because participants were asked to explore the interaction freely, we used a customized dialogue system for Pepper which consisted of the OpenAI GPT-3 language model for producing responses to participants' verbal input (OpenAI, 2023). The dialogue system was implemented as text completion, using the *text-davinci-002* language model. The language model was initialized before the participants entered the lab with a prompt: *You are talking to the robot Pepper. We are currently at the Interaction Lab in a town called Skövde. We are in the country Sweden.* Once the participants entered the lab, the interaction was always initiated by the participants. The participants speech was transformed via Google's speech-to-text service to text that GPT could respond to using the NaoQi ALAnimatedSpeech service. Within the interaction, previous speech by the participant was stored and the robot could therefore recall previous conversation. This was reset for each participant.

The study was conducted in a 60m^2 lab room, with the robot in the middle of the room, with screens around it to confine the room. The participants sat approximately one meter from Pepper, with one camera in front of them and one camera to the side (figure 3.1). The interaction was recorded in order to gather data on participant's facial expressions and body movements. The test leader was behind a screen to avoid participants interacting with them. The test leader was by a desk, overseeing the technical aspects of the robot.

3.3.4 ETHICAL CONSIDERATIONS

This project was submitted for ethical review to the The Swedish Ethical Review Authority (#2022-02582-01, Linköping) and was found to not require ethical review under Swedish legislation (2003:615). The experiment is in accordance with the Declaration of Helsinki. No physical or mental health risks were posed

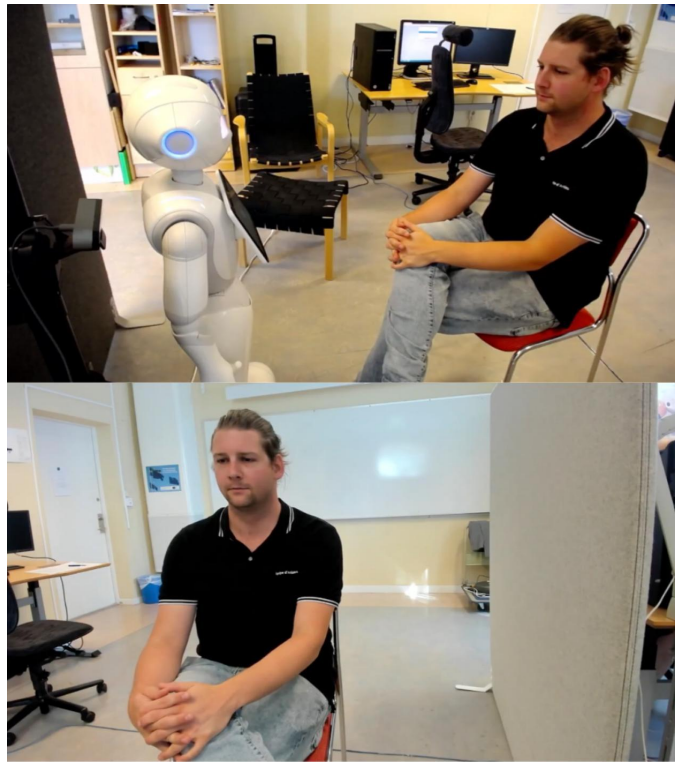

Figure 3.1: Set-up for the interaction, taken from the two cameras that were used to record the interaction.

to the participants of this study. Participants were informed of their tasks prior to receiving an informed consent form and were debriefed after the interactions. All data has been de-identified during collection. No sensitive personal information was collected. Video recordings are stored locally on a computer that is password protected.

3.3.5 DATA COLLECTION AND ANALYSIS RELATED TO THE EXPERIMENT

The analysis for the experiment included a within-subject design and was presented in Paper VI. The dependent variables in this experiment were expectations, measured by several subjective measures. The independent variable was time, i.e., experience from the interactions with the robot.

The measures were intended to gather aspects of expectations and were measured via negative attitudes, anxiety, closeness, and perceived capability. The data collection occurred before the first interaction, after the first interaction, and after the second interaction.

Three hypotheses were formulated. First, hypothesis 1 was that the variability

between participants' expectations towards the robot decreases over time. Second, hypothesis 2 was that previous experience affects expectations of robots. Lastly, hypothesis 3 was that expectations will change based on experience with the robot. To test hypothesis 1, separate two-sided F-tests were used to test the difference in variability between the data collected before the first interaction and after the last interaction. To test hypothesis 2 and 3, a repeated measures ANOVA was performed on each subscale in relation to time and previous experience with robots.

Demographics

Before the interactions with the robot, data on participants gender and age was collected. In addition, participants previous experience and interest in robots were collected. The participants were asked to rate the latter two on a scale of 1–5.

Negative attitudes towards robots

The negative attitudes towards robots scale (NARS) consists of 14 items divided into three subscales (Nomura et al., 2004). Subscale 1 covers *negative attitude toward situations of interaction with robots*, subscale 2 covers *negative attitude toward social influence of robots*, subscale 3 covers *negative attitude toward emotions in interaction with robots*. Each question is rated on a 1–5 Likert scale, 1 being *I strongly disagree*, and 5 being *I strongly agree*.

The robot anxiety scale

The robot anxiety scale (RAS) consists of 11 items divided into three subscales (Nomura et al., 2004). Subscale 1 covers *anxiety toward communication capability of robots*, subscale 2 covers *anxiety toward behavioral characteristics of robots*, subscale 3 covers *anxiety toward discourse with robots*. Each question is rated on a 1–6 Likert scale, 1 being *I do not feel anxiety at all*, and 6 being *I feel anxiety very strongly*.

Closeness

The closeness questions consist of three questions, based on the Other in the Self Scale (IOS) (Aron, Aron, and Smollan, 1992) which is one single item meant to measure how close one may feel towards a person. Because this scale was intended for human-human interaction, we decided to include three questions that have been used to validate this scale. The first question measures to what extent the participant would use the term "we" in relation to the robot. The second question measures how the participant would characterize the relationship with the robot in relation to other relationships. The third question measures how the participants would characterize the relationship with the robot in relation to other people's close relationships. Each question is rated on a 1–7 Likert scale, 1 being *Not at all* for question 1 and *Not close at all* for questions 2 and 3, and 7 being *Very much so* for question 1, and *Very close* for question 2 and 3.

Perceived capabilities

The perceived capabilities question was one item created for this study, rated on a 1–9 Likert scale, 1 being *Not capable at all* and 9 being *Extremely capable*. The

question was created in order to measure how capable the participants thought the robot in this experiment was.

3.3.6 DATA COLLECTION AND ANALYSIS RELATED TO THE UX EVALUATION

For the UX evaluation, we purposely selected a sample of users (N=10) from the criteria of having no previous experience with robots and their first language being either Swedish or English, which is presented in Paper IV. We chose this group in order to get users with similar backgrounds and experiences with social robots. Hence, their user profile was homogeneous. We used the Social Robot Expectation Gap Evaluation Framework as a foundation for the UX goals related to the factors of expectations and set baseline and target levels for each selected metric.

The UX goal related to affect was that the user should expect to have neutral to positive emotions toward the robot. The metrics included RAS (presented in section 3.3.5), facial expressions, and post-test interview. The post-test interview was conducted after the interactions in order to catch aspects of the interaction that could not be collected by the qualitative measures. The following six questions were asked:

1. How did you feel the interactions went?
2. Did you experience any difference in the first and second interaction?
3. Did you have any expectations of how the interaction would go?
4. Was anything surprising about the interaction or the robot?
5. Did you have any specific emotion during the interaction?
6. Is there anything you would like to add?

The UX goal related to cognitive processing was that the user should experience effortless cognitive processing during the interaction. The metrics included observations and post-test interviews.

The UX goals related to behavior and performance were that the user should expect a pleasant and smooth conversation, and that the user should expect to have ease of conversation. The metrics included RAS, field notes, observations, post-test interviews, and the amount of interruptions during the interactions. Both the observations and post-test interviews were video-recorded.

The data was analyzed and evaluated in terms of whether the UX goals were met or not via data triangulation (Patton, 2014). Triangulation means that multiple data sources are used to compare and contrast the data in order to gain a deeper and more reliable understanding of the obtained findings. Several findings that point in the same direction mean that there are identified UX problems that need to be considered. Once the UX goal was identified as met or not met, it was followed by assessing the nature of these UX problems, categorized into

scope (global and local) and severity. Findings that are pointing in the same direction imply a UX problem. Global scope consists of problems that relate to the robot or interaction as a whole, whereas local scope consists of problems that relate to a certain moment of the interaction. High severity problems are those that should be prioritized and are caused by several mismatches between users' expectations and the actual interaction. Low severity problems are those that are of smaller importance that users can easily work around. Lastly, recommendations to decrease the disconfirmed expectations are provided based on the revealed findings.

3.3.7 DATA COLLECTION AND ANALYSIS RELATED TO THE QUALITATIVE ANALYSIS

The data collected for the qualitative analysis was based on the observations from the field notes, the video recordings, and especially the post-test interviews from all 31 participants, which is presented in Paper VII. We decided to include the video recordings and the post-test interviews because it is acknowledged that participants' perspective from the interaction can vary between what they say and what they do (Patton, 2014; Lindblom, Alenljung, and Billing, 2020).

The data from the post-test interviews (presented in section 3.3.6) were transcribed and, together with the video recordings, were then analyzed by a reflexive thematic analysis (Braun and Clarke, 2006; Braun and Clarke, 2021a; Braun and Clarke, 2021b; Byrne, 2022). The analysis had six phases:

1. Familiarized ourselves with the data by transcription and review.
2. Generated initial codes by extracting data that seemed to be relevant and created short descriptive and interpretive labels.
3. Generated broader themes by actively identifying and constructing the patterns from the previous step.
4. Reviewed these themes to see if they represented the overall data and if they needed to be modified.
5. Defined and named themes in relation to the data sets and the aim of the research.
6. Wrote and produced the report

The analysis focused on understanding the interaction between the participant and the robot by assessing the interaction quality and whether or not it changed (negatively, stayed neutral, or positively) over the two interactions. Interaction quality was characterized as when participants had success or issues with (1) initiating or continuing an interaction, (2) turn-taking between themselves and the robot, including flow of conversation and potential interruptions, and (3) the dialogue, including well-aligned or misaligned conversations. Moreover, by investigating interaction quality and how it unfolded in the two interactions, we

also obtained insights into participants' interaction strategies, explicit and implicit expectations, as well as positive and negative user experiences.

SUMMARY OF PAPERS

CHAPTER 4
SUMMARY OF PAPERS

4.1 PAPER I

This paper contributes to RQ1 by investigating methodological practices in sHRI. Specifically, this paper investigates how ethical conduct is reported on in the HRI field. There is an ethical dimension to the mismatching expectations because designing social robots as human-like might deceive participants into thinking they are more capable than what they actually are. This is overtly used in HRI when using a WoZ set-up because the participants are meant to believe the robot is acting autonomously, when it is remote controlled by a human (the exception is certain kinds of UX research in which the participants are co-designers and informed about the purpose of the study) . It can be argued that any social robot is designed deceptively, by being designed in ways inviting to social interactions. One way to handle this in HRI is to report on how these ethical dimensions were handled when reporting on empirical studies, which is a standard in fields such as psychology. Therefore, this paper investigates the reporting of ethical conduct, via a literature review where I looked at how often ethical board approval, informed consent, data protection and privacy, deception, and debriefing were explicitly mentioned when reporting results from traditional empirical studies.

Results show that, overall, ethical conduct is under-reported in HRI, with four of the five ethical principles omitted in about one third of the papers analyzed. Informed consent was the most reported ethical conduct, mentioned in 49% of the articles. In 44% of the papers' explicitly using deception in their study, authors also mentioned debriefing. All studies that included deception should disclose to the participants when deception occurs. Uninformed participants may have unrealistic expectations of what robots are capable of based on the study they participated in, which ought to be considered from an ethical perspective.

This paper demonstrates and contributes to identification of potential shortcomings in ethical research practice in HRI and provides explicit discussion of best practices for ethical participant interactions in HRI. Specifically, the paper highlights the overlooked impact of undisclosed deception in the most common HRI

method today.

4.2 PAPER II

This paper contributes to RQ1 by lifting the role and relevance of expectations in human-robot interactions and how we can identify and study expectations with UX methodology. Moreover, I stress how expectations can be a confounding variable in sHRI research which highlights the need to understand and consider expectations further within the field. This paper was the first step towards incorporating UX methodology into my work on expectations, which ultimately set the stage for the development of the Social Robot Expectation Gap Evaluation Framework (Paper III).

A key aspect of UX is understanding that user experiences belong to a context, specific users, and how they can change over time. Although these concepts can be present in sHRI today, they are often overlooked or not framed in a systematic way. Because expectations (in general, and in sHRI) are dependent on context, the users, and can change over time, I propose the novel use of UX in an sHRI context.

In order to prevent negative user experience, I suggest that the UX wheel can be used to assess expectations in the early stages of designing robots. By doing so, expectations can be understood further, creating confirmed expectations, and ultimately design for a positive user experience. The UX wheel consists of understanding users' needs, creating design concepts, realizing design alternatives, as well as verifying and refining designs via evaluations. These steps are meant to be iterative, where each step is worked and reworked until desirable results are achieved. These steps can be beneficial for already developed robots because the iterative UX wheel can be used to design aspects of the robot, such as its behaviors, for specific contexts. The implications of using UX methodology, including the UX wheel, is that designers can understand and design for expectations to create positive user experience. An outcome of negative user experience is that the users will stop using the technology, which can be detrimental in contexts such as healthcare. By managing expectations, specifically by having social robots confirming users' expectations, positive user experience can be created. This means that the social robot can be used for its intended use, which is one of the main goals in sHRI.

4.3 PAPER III

This paper contributes to RQ1 by proposing a framework for which expectations can be studied in an sHRI context. The framework draws inspiration from work on expectations in social psychology and combining them with methodology and aspects from UX. In this paper, I formulated three factors of expectations. Then, I developed four UX goals, with proposed metrics, to study these factors. These metrics are meant to be modular, where certain metrics could be removed or replaced. By doing this, the framework can be tailored after unique scenarios and contexts while still keeping the core of the factors of expectations. Lastly, I

proposed a procedure with four phases, relying heavily on the UX perspective.

Factors of expectations and metrics

The foundation of the framework is based on the expectancy process by Olson et al. (1996). The first step of this paper was to identify what aspects of the expectancy process are relevant to study in sHRI from a UX perspective. Of importance here was to identify how expectations can be studied once an expectation is confirmed or disconfirmed. From this, I identified these factors as being *affect, cognitive processing* and *behavior & performance*, derived from the consequences of expectations. Affect refers to the emotional reaction an individual may have in the interaction with the robot. These can range from positive to negative affect. The metrics NARS and RAS (Nomura et al., 2004) are two questionnaires aiming to measure participants' negative attitudes and anxiety towards robots with a broader scope, considering robots collectively. However, both NARS and RAS are commonly used as subjective measures after interacting with specific robots. Following this practice, the primary objective of using these measures in this framework is to observe how these general expectations evolve throughout interactions with specific robots. Cognitive processing refers to the cognitive strain the interaction may take on an individual. Typically, disconfirmed expectations are surprising and attention is drawn to the occurrence, and cognitive effort is put to making sense of what happened. The metrics proposed for this factor are memory recall (i.e. asking participants to recite the interaction) and reaction time (i.e., measuring the time it takes for participants to react to the robots output, both verbal and non-verbal). Lastly, behavior & performance refers to an individual's deliberate actions. The metrics proposed for this factor are choice of spoken dialogue (i.e., what topics the participant chooses to bring up in the interaction), repeating words (i.e., how often the participant needs to repeat themselves), interruptions of interactions (i.e., the number of times the participant is interrupted by the robot in any way), and the duration of the interaction (i.e., how long the interaction lasted, including how long the participant speaks).

UX goals

In order to study the factors of expectations, four UX goals were formulated. First, for the affect factor: *the user should expect to have neutral to positive emotions towards the robot*. Second, for the cognitive processing factor: *the user should experience effortless cognitive processing during the interaction*. Third, for the behavior & performance factor: *the user should expect a pleasant and smooth interaction*. Fourth, also for the behavior & performance factor: *the user should expect to have ease of conversation*.

Procedure

The framework also entails the procedure in which studying these aspects can be carried out systematically. The procedure is divided into four phases. First, before carrying out any study, the scenario needs to be identified. This can be done by creating two identical scenarios and studying how the factors of expectations change between the two interactions. Another suggestion is to change the second interaction in order to trigger expectancy violation. This way, disconfirmed

expectations are triggered and can be studied. The next step is to collect the data. I stress the importance of studying expectations over time, and thus, collecting data several times is important to catch the change in expectations. I suggest a timeline of the data collection phase in figure 4.1. Third, once the data collection is completed, the data needs to be analyzed. This step combines the findings from the different metrics, quantitative and qualitative, in order to create a bigger picture of how expectations change. During this phase, UX problems need to be identified, relating to the scope and the severity of them. These problems can then be used for further improving the interaction, or even the design of the robot if the study is intended to evaluate a robot under development or re-design. The last step is reporting on findings and to offer recommendations. The recommendations should be related to how disconfirmed expectations can be avoided and how to design for having overall positive user experience for the users interacting with a robot.

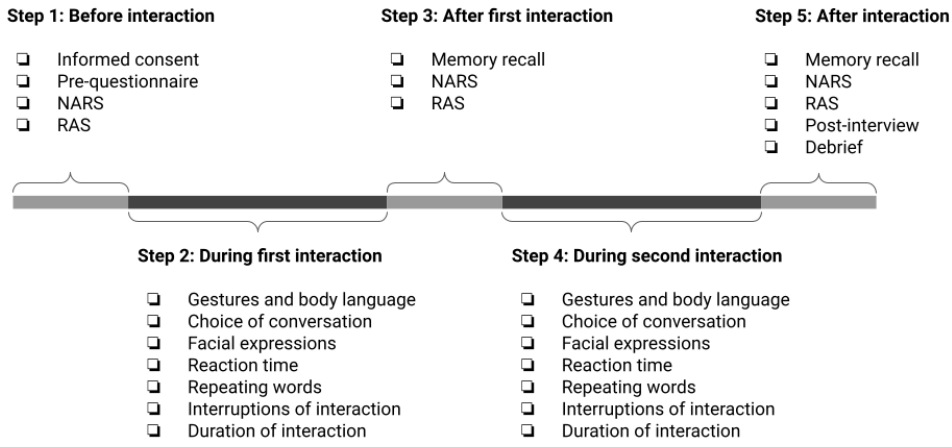

Figure 4.1: Data collection timeline, from Paper III

4.4 PAPER IV

This paper contributes to RQ1 by applying the Social Robot Expectation Gap Evaluation Framework to a user experience evaluation study. This paper focuses on the results that can be applied by the framework. Because the framework is intended to be modular, I kept some of the metrics presented in Paper III and added new ones. The details of the findings are presented in table 4.1. For the affect factor and UX goal 1, I chose to include RAS, facial expressions, and results from the post-test interview. For the cognitive processing factor and UX goal 2, I chose to include observations and data from the post-test interview. For the behavior & performance factor and goal 3, I chose to include observations and data from the post-test interview; and for goal 4, I chose to include interruptions of interaction, observations, and data from the post-test interview. All data was analyzed via triangulation for each and every UX goal.

Four major findings were identified in relation to the framework. First, during the analysis process, I saw, generally, that UX was improved in the second interaction. These findings support the claim that expectations should be studied over time because they are not static but dynamic in the interactions.

Second, I saw the importance of studying expectations in-person because the physical robot had an impact on the users when stepping into the test room. Several users expressed a difference in their expectations once they met and interacted with the robot, with some experiencing negative emotions before the interactions which might not have happened if the study was conducted online.

Third, there was an overlap between the findings of the different expectation factors. For example, subscale 2 for RAS (relating to the affect factor) relates to the behavioral characteristics of the robot, which makes it appropriate for the cognitive processing factor because behavior from the robot can also put a strain on cognitive processing (i.e., the users tries to make sense of the behavior). Moreover, subscale 3 for RAS relates to discourse with robots, which makes it appropriate for behavior & processing because unexpected discourse by the robot can lead to less ease of conversation, because users may not know what to respond. This overlap between factors underpins the importance of triangulation when analyzing the data to further understand how expectations work.

Fourth, the findings from this study alluded to other aspects of the expectancy process by Olson et al. (1996). This implies that the framework can be further improved by looking at other aspects of the expectancy process, specifically the dimensions of expectations: certainty, accessibility, explicitness, and importance (section 2.2.2). For example, accessibility seems to be partly involved in the initial user experiences when the users are meeting the social robot in-person, and explicitness seems to be partly involved for users when they experience the mixed message of being aware that the robot is a machine, but still compare the robot's verbal responses and actions to how human-human interactions commonly unfold.

Moreover, this paper contributes to RQ2 by empirically studying how the experiences of interacting with a social robot affect users' expectations over time from a UX perspective. UX goal 1 was not fulfilled. Facial expressions during the interaction demonstrate that several users were anxious and avoided eye contact with the robot. Several users also displayed puzzled looks in the first interaction, indicating that they did not know what to expect from the robot. The main identified UX problem was that these users lacked first-hand experience.

UX goal 2 was partially fulfilled. The users appear to, overall, start off with high cognitive processing and moving towards it becoming easier over time. Several users expressed during the post-test interviews that they had a hard time figuring out how to interact with the robot, but that it became easier at the second interaction. The identified UX problem was that users did not know what to expect from the robot, and whether or not it should be regarded as a machine or human-like, which caused a cognitive strain on the users during the interaction.

UX goal 3 was not fulfilled. Overall, several users did not have smooth interactions because the robot did oftentimes not recognize the users' speech, and in some cases

Expectation Factors	UX Goal	Metric	Details	Qualities	Baseline before interacion	Baseline after first interaction	Baseline after second interaction	Target level before interacion	Target level after first interaction	Target level after second interaction	Observed results before interacion	Observed results after first interaction	Observed results after second interaction	Meet target before interacion	Meet target after first interaction	Meet target second interaction
Affect	The user should expect to have neutral to positive emotions towards the robot	The Robot Anxiety Scale (RAS)	A questionnaire measuring user's anxiety towards robots	Hedonic	S1: 9 S2: 12 S3: 12	S1: 8 S2: 11 S3: 11	S1: 7 S2: 10 S3: 10	S1: 7 S2: 10 S3: 10	S1: 6 S2: 9 S3: 9	S1: 5 S2: 8 S3: 8	S1: 6 S2: 10 S3: 11	S1: 5 S2: 8 S3: 8	S1: 5 S2: 7 S3: 8	S1: yes S2: yes S3: no	S1: yes S2: yes S3: yes	S1: yes S2: yes S3: yes
		Facial expressions	Observing the kind of facial expressions made by the user			Negative	Neutral		Neutral	Positive		Negative	Negative		No	No
		Post-test interview	Asking the user if the felt any emotions during the interaction													
Cognitive Processing	The user should experience effortless cognitive processing during the interaction	Observations and post-test interview	Observing the interaction and asking the user what was surprising about the interaction	Hedonic		Negative	Neutral		Neutral	Positive		Negative	Positive		No	Yes
Behavior and Performance	The user should expect a pleasant and smooth conversation	Observations and post-test interview	Observing the interaction and asking users about their behavior during the interaction	Hedonic		Negative	Neutral		Neutral	Positive		Negative	Neutral		No	No
	The user should expect to have ease of conversation	Interruptions of Interaction	Measuring the amount of times, and what kind of, interruptions occur for the user during the interaction	Pragmatic		2	1		1	0		0.3	0.3		Yes	No
		Observations and post-test interview	Observing the interaction and asking users about their interaction	Hedonic		Negative	Neutral		Neutral	Positive		Negative	Neutral		No	No

Table 4.1: The applied Social Robot Expectation Gap Evaluation Framework, with set levels and results, from Paper IV.

misheard the commands. In addition, many users reported that the interaction felt one-sided and that it did not feel like a natural interaction because the robot did not ask questions back. There were, however, some users who had a smooth interaction, mainly due to them being able to be ”understood” and no miscommunications occurred. The identified UX problem was that the robot did not respond to all the voices of the users equally.

UX goal 4 was partly fulfilled. Several users were interrupted by the robot. In the post-test interviews, users reported a feeling of awkwardness when the robot did not respond to speech. There were, however, users who stated that the conversation was easy because the robot could understand all the topics. Several users also reported that they thought the conversations were easier at the second interaction. The UX problem was that users could not experience ease of conversations because they did not know what to expect from the robot.

Two of the UX problems were deemed as severe because they ultimately created bad UX – users had problems being understood by the robot and users did not know what to expect from the robot, in terms of if the robot should be handled as a machine or as a human. Despite this, there was an improvement between the first and second interactions. Moreover, many users expressed disappointment when their expectations of the robots did not align with reality, which is in line with the expectancy process by Olson et al. (1996).

4.5 PAPER V

This paper contributes to RQ1 by investigating methodological practices in sHRI. Specifically, this paper investigates how NARS is used and reported on in the HRI field. Results show that the reported use of NARS is increasing. This can be seen as evidence that it is a tool that is relevant, although the actual reported numbers of use are still low. Results also show that NARS was inconsistently applied and not always reported in sufficient detail to confidently interpret the results. Looking specifically at reporting on how the response options were analyzed and how many options the participants could choose from, a great variation in practice was found. For instance, only a third of the papers reported both these aspects, whereas the rest only reported one or neither. Those that explicitly reported their practices were often not using the validated methods, and it was not always possible to infer the used methods from papers with more sparse reporting. When using the validated scale, NARS should result in one number for each of the three subscales. Each of those numbers are calculated by summarizing the responses for the related Likert items. In 56% of the papers, there was no mention on how they evaluated the items to receive the three NARS numbers, neither explicitly nor implicitly. In 31% of the papers, the item evaluation was not in line with the validated method. The number of response options for the respective items also varied a lot. There should be five options, and it was confirmed that that was the case in only 35% of the papers, but in almost half of the papers, it was not possible to tell. In 16% of the papers, other numbers were used, and seven was the most common alternative.

The results found in this paper are important to highlight because NARS has an

important role within the sHRI field, both to capture a common confounder and to study an important phenomenon in itself, specifically how underlying attitudes of robots in general might impact the specific context. The value of any validated scale is twofold, first, it allows for comparison, and second it has a documented correlation with the studied phenomenon. Thus, in order to get meaningful and comparable results, NARS needs to be consistently used. If any modification to the scale occurs, then that new scale would need validation, or at least have the adjustment being clearly reported. There are several reasons why modifications could be appropriate, such as adjusting to specific contexts or users, but the new scale cannot typically be considered NARS.

This work contributes to sHRI methodology by reviewing historical and current practices which in turn informs future improvements of our tools. Considering the results showed the inconsistent use of the scale while still being one of the most popular scales in the field, used across many settings, it makes it worthwhile to discuss how we should move forward as a field when using this, or any other, standardized scale to measure aspects of sHRI. With this paper, I want to encourage consistent use of the scale in order to perform meta-analysis to make it possible to examine more general phenomena related to attitudes and robots, including the expectations people have on robots.

4.6 PAPER VI

This paper contributes to RQ2 by empirically studying how the experiences of interacting with a social robot affects users' expectations over time. These results are from the empirical study presented in section 3.3. Participants were instructed to interact with the robot Pepper two times, for 2.5 minutes each. Expectations were measured via affect (NARS, RAS, and Closeness) and perceived capability. Three hypotheses were formulated – H1: The variability between participants' expectations towards the robot decreases over time, H2: previous experience affects expectations of robots, and H3: expectations will change based on experience with the robot.

H1 was tested by conducting separate two-sided F-tests to test the difference in variability between the data collected before the first interaction and after the last interaction. Results show no decrease in variability for any of the measures and the hypothesis (H1) was thus rejected. I hypothesized that the variability would decrease because the participants would come into the experiment with different expectations and, through direct experience with the robot in the experiment, start to revise their expectations to a more uniform picture of the robot. This hypothesis was formulated with the expectancy process by Olson et al. (1996) in mind. A possible explanation for why a decrease in variability was not found is that the participants' expectations were so robust that the two interactions were not enough to revise the expectations in the interaction. Another possible explanation is that the participants had enough of a different experience in the interaction to induce a decrease in variability. This explanation is supported by the found tendency for increased variability over time in the perceived capability measure. Furthermore,

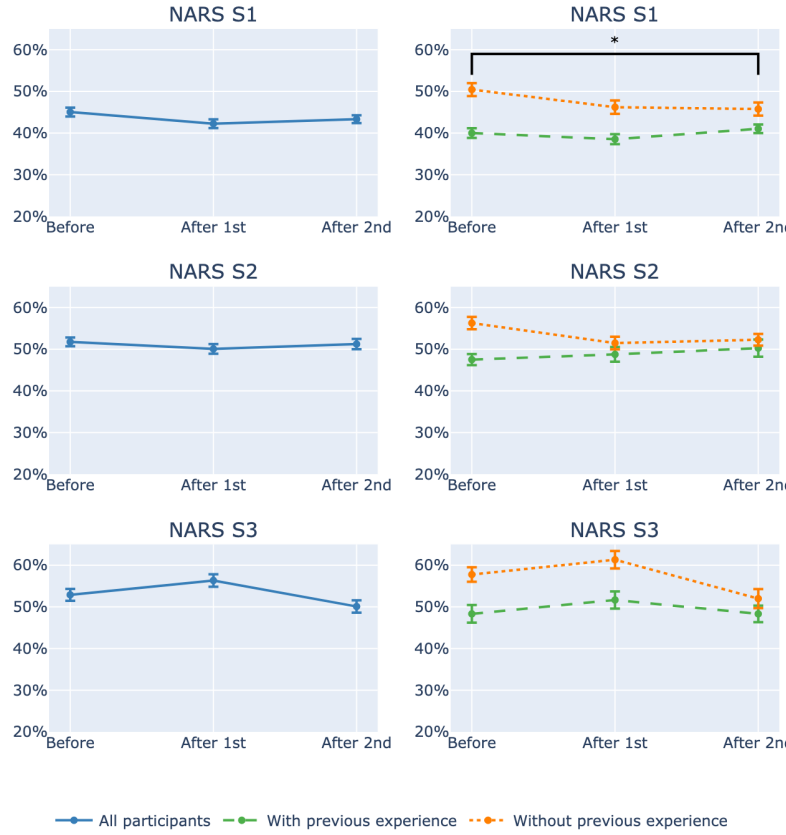

Figure 4.2: The mean scores for the three NARS subscales from Paper VI, both for all participants (left) and when separated based on previous robot experience (right), are presented as a percentage of the maximum score achievable for each component. Error bars are included to represent the standard error of the mean.

a change in expectation was found for the participants but in different directions which also points towards the participants having different experiences.

H2 and H3 were tested by conducting a repeated measures ANOVA, which was performed for each measure in relation to time and previous experiences with robots. The results from NARS, RAS, Closeness, and Perceived Capability are presented in figures 4.2, 4.3, 4.4, and 4.5. H2 was supported whereas H3 was partially rejected. Participants with previous experiences with robots reported more positive responses on the scales. One possible explanation for these results is that the open dialogue system that was used made it possible for more complex interactions with the participants which the experienced participants responded well to. Moreover, these results are in line with the Expectancy model by Olson et al. (1996) in regard to the role the sources of expectations play in expectations. Specifically, those with direct experience with robots had different expectations

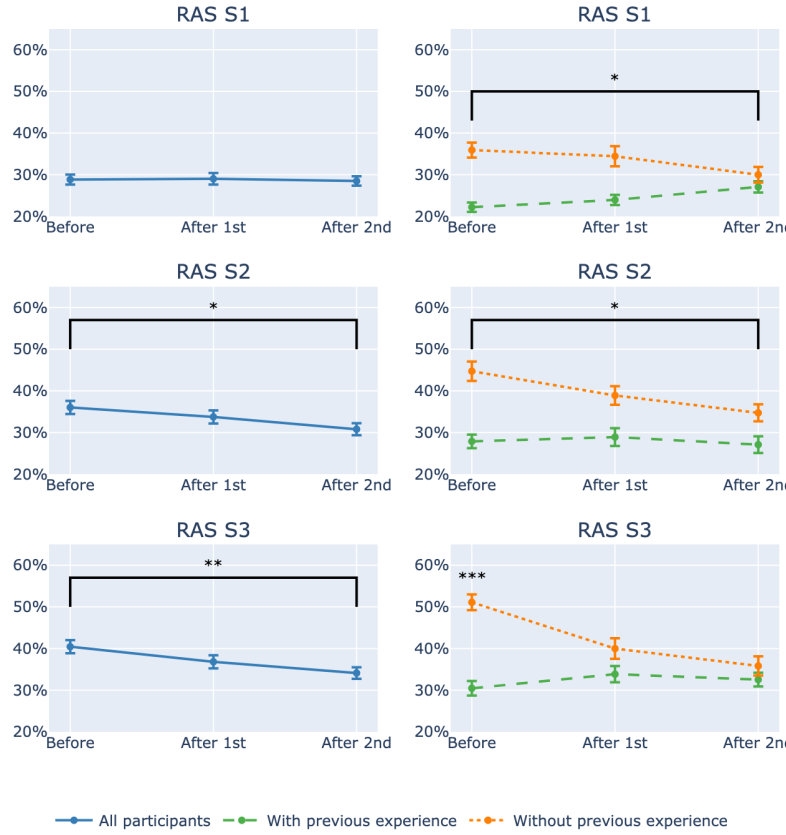

Figure 4.3: The mean scores for the three RAS subscales from Paper VI, both for all participants (left) and when separated based on previous robot experience (right), are presented as a percentage of the maximum score achievable for each component. Error bars are included to represent the standard error of the mean.

of robots than participants that had no direct experience with robots. These expectations were both different before the interaction and after the interaction with the robot. This explanation is also supported by the findings from H1 because the source of the expectations are robust and participants retain them in the interaction.

For the measures with a change, I found that participants became more positive towards the robot over time in the interactions. One possible explanation for this result is that different aspects of expectations stabilize at different times, which has been found previously to happen in first impressions of social robots (Paetzel, Perugia, and Castellano, 2020).

Our main take-way from this study is that previous experiences have a strong impact on the current experiences with robots; participants' expectations of robots are robust, with people with previous experiences scoring significantly differently

on affect, which will not change over time from the experience of interacting a shorter period with a robot. These results contribute to RQ2 by underpinning the importance of previous experiences with robots, and the difference between direct experience and no direct experience, which affects forthcoming interactions. I can further demonstrate how the underlying assumption that HHI works similarly as sHRI is true for expectations, by applying the expectancy process by Olson et al. (1996).

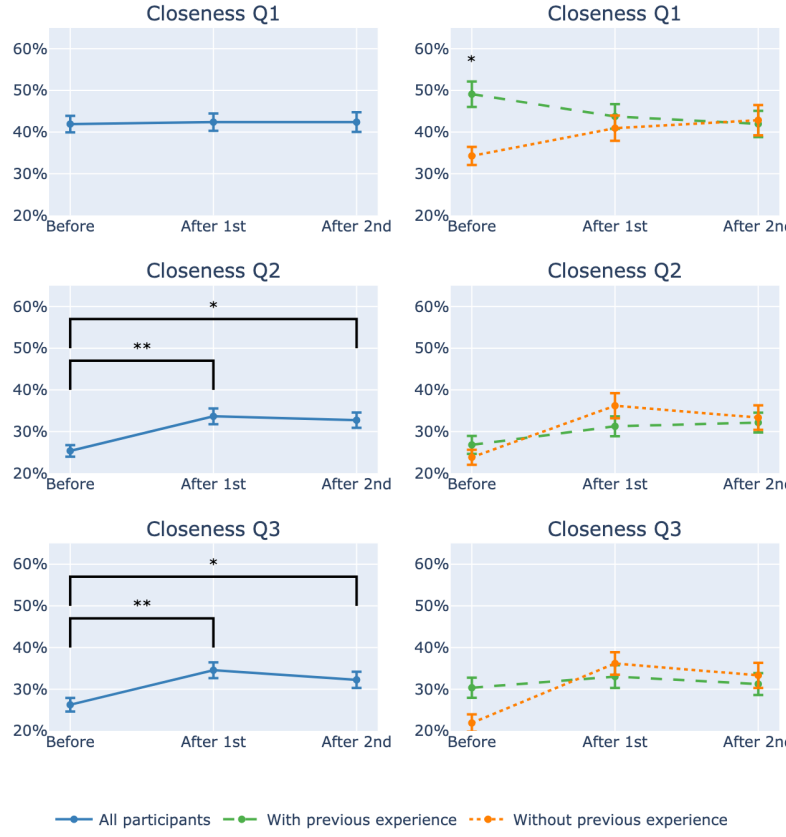

Figure 4.4: The mean scores for the three Closeness questions from Paper VI, both for all participants (left) and when separated based on previous robot experience (right), are presented as a percentage of the maximum score achievable for each component. Error bars are included to represent the standard error of the mean.

4.7 PAPER VII

This paper contributes to RQ2 through an extended qualitative study, delving deeper into the specifics of expectations and user experiences. This paper aimed to gain a deeper understanding of how expectations influenced the social human-

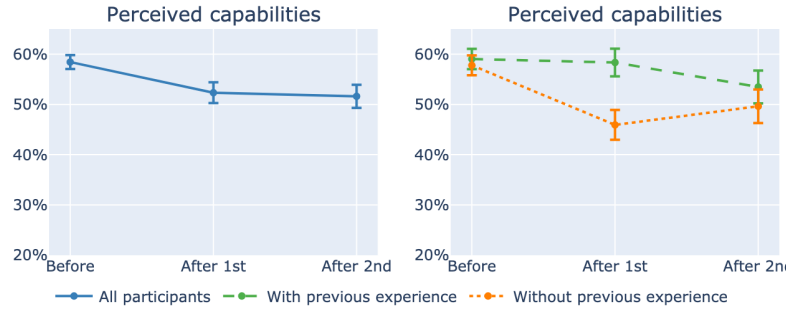

Figure 4.5: The mean scores for the Perceived Capability question from Paper VI, both for all participants (left) and when separated based on previous robot experience (right), are presented as a percentage of the maximum score achievable for each component. Error bars are included to represent the standard error of the mean.

robot interaction as well as how humans experienced it. The results are from the empirical study presented in section 3.3. A reflexive thematic analysis of data extracted from the post-test interviews and video-recorded interactions was conducted. Five objectives were formulated. The first objective was to understand the interaction quality. The second objective was to identify interaction strategies. The third objective was to further examine users' confirmed and disconfirmed expectations. The fourth objective was to identify core aspects that influenced the users' expectations and experiences. Lastly, the fifth objective was to examine users' experiences. The insights gained from this analysis go beyond the acknowledgment of expectations, offering a more nuanced exploration of their nature and thematic categorization.

The first objective, concerning interaction quality was the first step and built the foundation for studying the other objectives. The interaction quality was divided into three levels: low, moderate, and high. This categorization was determined through aspects such as turn-taking, spoken dialogue continuity, interruptions, and alignment of spoken dialogue topics (e.g., the robot responding to the participant's question or responding to something unrelated). The 62 sessions of human-robot interaction were individually rated on their respective interaction quality as well as the trend (positive, neutral, and negative), illustrated in table 4.2. The result shows an overall positive trend of increased interaction quality from the first to the second session. Participants' self-reported assessments of the trend from the first to the second session correspond with the ratings. Participants demonstrated a heightened understanding of how to engage with the robot in the second session compared to the first session. This suggests that their expectations were more accurately aligned in the second session, likely influenced by their firsthand experiences with the initial encounter.

The second objective focused on the users' interaction strategies. From the analysis, five distinct approaches were identified: getting to know the robot, explor-

ing the robot's capabilities, testing the limits of the robot's functionality, asking questions related to the differences between humans and robots, and asking no questions (typically due to now knowing what to do after issues). These interaction strategies did not only share similarities in content but also in how frequently they were employed. All participants posed questions related to the first and second interaction strategies, while the remaining three were utilized differently, largely influenced by time constraints and individual interests. Participants experiencing a lower interaction quality, often due to various reasons, seldom utilized the latter strategies as they had limited time to complete the initial socializing phase before the sessions concluded.

The third objective focused on various aspects of explicit expectations, specifically exploring confirmed and disconfirmed expectations.The analysis included exploring how users land at various points within the social expectation gap (figure 2.6), based on their expectations and their subsequent experiences. The findings not only highlighted the diversity of expectations but also emphasized the participants' adaptability in aligning those expectations based on their previous experiences. Overall, the findings unveil a spectrum within the social robot expectation gap, showcasing that expectations vary within a broad range.

The fourth objective specifically focused on the core elements impacting participants during the study, derived from aspects that surprised them and analyzed as implicit expectations. These elements varied across a range of dimensions from negative to positive. Participants showed varied reactions regarding different aspects of the robot: the dialogue triggered various responses, with some impressed by the robot's answers and others disappointed by its capabilities and response timing. Surprises were found in the robot's arm movements and minimal attention was given to the robot's appearance. Additionally, the portrayal of a human-likeness by the robot surprised participants, provoking thoughts on the robot's potential self-awareness and understanding of emotions. Overall, participants demonstrate a tendency aligning their expectations with their past experiences and encounters.

The fifth objective focused on exploring users' experiences. User experience was divided into positive user experiences and negative user experiences. Experiences were based on the spoken dialogue and the head and body movements of the robot. For many participants, the initial in-person encounter left a positive impression, indicating that the robot surpassed their initial (implicit and explicit) expectations in both its actions and appearance. However, there were also negative user experiences, which were experienced when there was a lack of proper verbal dialogue and turn-taking. These participants expressed that the interaction and the situation did not feel natural and comfortable. Moreover, the robot's head and body movements caused confusion or fright for these participants rather than supporting the interaction between them. Some participants mentioned that they had seen pictures and movie clips of social robots, but interacting with them first-hand was described as qualitatively different and more negative than imagined beforehand.

The findings underscore the intricate nature of user expectations. Recognizing

Participant	Gender	First session	Second Session	Trend	Participant's view of the trend
1	Female	Moderate	Moderate	Positive	Positive
2	Male	Moderate	Moderate	Positive	Positive
3	Male	Low	Low	None	Positive
4	Female	Low	Moderate	Positive	Positive
5	Female	High	High	Positive	Positive
6	Male	High	High	None	Negative
7	Female	Low	Moderate	Positive	Positive
8	Male	Low	Moderate	Positive	Positive
9	Male	Low	Moderate	Positive	Positive
10	Male	Moderate	Moderate	Negative	Negative
11	Female	Low	Moderate	Positive	Positive
12	Male	Low	Low	None	None
13	Female	High	Moderate	Negative	None
14	Female	High	High	Positive	Positive
15	Male	High	High	None	None
16	Female	High	High	None	None
17	Male	Low	High	Positive	Positive
18	Male	Moderate	Moderate	Positive	None
19	Female	Low	High	Positive	Positive
20	Female	High	High	Positive	Positive
21	Female	Moderate	Moderate	Positive	Positive
22	Female	Low	Moderate	Positive	Positive
23	Male	High	Moderate	Negative	Negative
24	Female	Low	Moderate	Positive	None
25	Female	High	Moderate	Negative	Negative
26	Male	High	High	None	Positive
27	Female	Low	Moderate	Positive	Positive
28	Female	Moderate	Low	Negative	Negative
29	Male	Low	Low	Negative	None
30	Male	Low	Moderate	Positive	Positive
31	Female	High	High	None	None

Table 4.2: Interaction quality, from Paper VII

the importance of implicit expectations is critical in understanding user responses, particularly as participants might encounter challenges in expressing explicit expectations. Moreover, this paper's results highlight participants' adaptive strategies, emphasizing adaptation as a crucial aspect of user interaction. Their adaptability significantly shapes their experiences, revealing that positive user experience isn't solely determined by interaction quality.

4.8 OTHER PUBLICATIONS – WORKSHOPS

In addition to the empirical work, there have been several workshop papers that partly have contributed to answering RQ2. Although these are smaller publications, they highlight an important aspect of expectations in sHRI which is an ethical dimension. In Paper VIII, I highlighted how the presentation of social robots impacts the public's expectations of robots. This is illustrated by bringing up three examples of robots in media: existing social robots presented at talks and conventions acting as more capable than what is currently possible, advertisements where existing robots are sold as being able to do more than what is possible today, and science fiction robots portraying highly intelligent "almost humans" capabilities. These examples are discussed in relation to sHRI research that has studied the media's effect on the expectations people may have of robots. Lastly, I raise questions regarding the ethical issues with presenting robots as more capable than what is possible today (e.g., via WoZ), which creates a certain type of deception. The ethical challenges in the sHRI field and how they can create unrealistic expectations of social robots were further presented in Paper XI, where five examples of ethical challenges were presented: emulated emotions in the robot, the presentation form of robots, the terminology used when describing robots, concealing cameras and microphones in the robot design, and lastly using the WoZ technique. In Paper XII, I took a closer look at how using a human-like voice in robots can be viewed as deception because it creates expectations that robots are like humans (i.e., being more capable than what is possible). Lastly, in Paper XIII, I took another approach to the ethical dimensions of our work, namely the language bias that exists in the voice recognition of robots. In this paper, I presented some cases from our empirical study (section 3.3) where there were instances of breakdowns in communication where the robot system was unable to pick up the speech from the participants, tied to participants who deviated from American English pronunciation. These instances were outliers in some way, but too unique to do any form of data analysis on them. We, therefore, used them as examples for a more theoretical discussion on who has access to new technology.

Together, these papers highlight some of the ethical dimensions of expectations including how presenting (fictional and real) robots as more capable than what is possible today can be viewed as deception, which creates a challenge for researchers within the sHRI field. These results also strengthen the findings that the sources of expectations play a larger role in human-robot interaction and have an effect on the results in sHRI research.

FINDINGS

CHAPTER 5
FINDINGS

In my doctoral work, I aimed to investigate the role expectations play when interacting socially with robots, including the subsequent ethical implications of such expectations. I approached this aim, first, by asking how user expectations can be studied in sHRI (RQ1) since there do not seem to be any widely adopted methods to study expectations within the field. Then, I continued to approach the aim by asking what is the role and relevance of user expectations in sHRI (RQ2).

5.1 RQ1: HOW CAN USERS' EXPECTATIONS BE STUDIED IN sHRI?

Knowledge and insights into this research question is based on Paper I, Paper II, Paper III, Paper IV, and Paper V. I have answered this RQ in two ways. The first way is by developing a framework for studying expectations in sHRI, and the second way is by looking at existing methods in sHRI.

5.1.1 A FRAMEWORK FOR STUDYING EXPECTATIONS

As argued in the Introduction, HRI has up until today primarily been concerned with designing robots in relation to users' preferences, not considering specifically how these preferences are formed. Gaining a richer understanding of users' interactions with robots and how they are shaped in relation to their expectations may provide an additional important dimension in the analysis in sHRI research. Specifically, being able to study users' expectations would allow for an understanding of how to manage these expectations. Disconfirmed expectations are not necessarily an indication of poor interaction design or individual preferences but rather it shows the need to adjust for the expectations in order to create a positive user experience.

In Paper II, the need for a more structured and systematic approach to study expectations in sHRI was identified and then realized. The motivations behind

this identified need are several. First, it is acknowledged that people tend to form their opinions of robots based on science fiction and other media, i.e., indirect experience, creating preconceived notions of robots capabilities (Sandoval, Mubin, and Obaid, 2014; Alves-Oliveira et al., 2015; Rosén et al., 2018; Oliveira and Yadollahi, 2023). Second, roboticists design social robots with the intentions of creating an interaction partner, blurring the line between a machine and a social agent (Alač, 2016), which in turn affects users' expectations. Third, because expectations change over time (Olson, Roese, and Zanna, 1996; Roese and Sherman, 2007), there is also a need to consider the temporal aspect of peoples' experiences with robots. The traditional view of studying behavior in research is introducing a stimuli to a group of people and observe the outcome of doing so. However, if the original group, before the stimuli, is not uniform, there might be an confounding variable that is affecting the results. In this case, I argue that the variable in sHRI is the expectations of social robots. Therefore, in Paper II, UX methods were introduced as a way to study how expectations may change over time. Lastly, even though UX is useful for studying expectations, UX usually focuses on technical artifacts that typically do not appear to be and act human-like; rather, UX is typically focused on users' experience with inanimate technological objects, although the intersection between sHRI and UX is growing. Thus, I identified the need to complement the UX approach to studying expectations with a theoretical grounding in social psychology.

Paper III builds on the ideas presented in Paper II by suggesting that the expectancy process by Olson et al. (1996) can be used to create a framework for studying expectations in sHRI. In this paper, the *Social Robot Expectation Gap Evaluation Framework* was developed. The framework takes inspiration from social psychology while deploying UX methodology in order to create an interaction where expectations will ideally align with the robot's capability. The framework is presented in table 5.1. There are slight modifications of the wording of the UX goals that has been updated for this thesis.

The proposal is that users' expectations of social robots can be assessed by considering three consequences of expectations, namely; affect, cognitive processing, and behavior & performance. Positive affect, low cognitive load, and smooth interactions are all indications that the robot meets a user's expectations. Conversely, negative affect, high cognitive load, and problems during interaction with social robots should be seen as indications of disconfirmed user expectations, not necessarily indications of poor interaction design or individual preferences.

The framework is intended to be modular, where certain metrics can be removed and new ones can be added, while still focusing on the three factors of expectations. Therefore, the metrics presented in table 5.1 should be viewed as a starting point, and researchers are encouraged to tailor their own metrics after the specific context, robot, and interaction that is used.

In Paper IV, the framework was applied in an empirical setting as an initial step towards validating the framework. The results from using the framework aligned with the expectancy process by Olson et al. (1996) including how the three consequences of expectations, i.e., affect, cognitive processing, and behavior & perfor-

Expectation Factors	UX Goal	Metric	Details	Qualities
Affect	The user should experience to have neutral to positive emotions towards the robot	The Negative Attitude Toward Robot Scale (NARS)	A questionnaire measuring user's negative attitudes towards robots	Hedonic
		The Robot Anxiety Scale (RAS)	A questionnaire measuring user's anxiety towards robots	Hedonic
		Facial Expressions	Observing the kinds of facial expressions made by the user	Hedonic
Cognitive Processing	The user should experience effortless cognitive processing during the interaction	Memory Recall	Asking the user to write down what they remember of the interaction	Pragmatic
		Reaction Time	Measuring the time it takes for the user to react accordingly to the robot's output	Pragmatic
Behavior and Performance	The user should experience a pleasant and smooth interaction	Gestures and body language	Observing the kind of gestures and body language the user expresses	Hedonic
		Choice of Conversation	Observing the kinds of conversations the user tend to focus on during the interaction	Hedonic
	The user should experience to have ease of conversation	Repeating Words	Measuring the number of repetions the user makes during the interaction	Pragmatic
		Interruptions of Interaction	Measuring the amount of times, and what kind of, interruptions occur for the user during the interaction	Pragmatic
		Duration of Interaction	Measuring the total time spent in the interaction as well as the total time spent on each conversation for the user during the interaction	Pragmatic

Table 5.1: The three factors of expectations and the proposed metrics to study each when interacting with a social robot.

mance) change over time. It became evident that, in terms of studying expectations, HHI models of interaction are transferable to sHRI research. I also identified the importance of conducting expectation research in-person because this appears to have a strong effect on expectations in a social human-robot interaction.

5.1.2 METHODOLOGICAL PRACTICES IN sHRI

While developing the Social Robot Expectation Gap Evaluation Framework, questions regarding the methodological practices in the sHRI field emerged. I believe this is important because the field is growing and various methods are still being developed at a fast pace, which causes the risk that certain practices become the norm without careful consideration (Lindblom and Andreasson, 2016; Dautenhahn, 2018; Jost et al., 2020). These concerns are highly relevant for the study of expectations in sHRI, but applies also more broadly to all sHRI research involving human participants.

In Paper I, ethical reporting practices in the field were investigated and analyzed. One of the issues raised in this thesis is whether or not building unrealistic expectations of social robots by designing them as human-like is a kind of deception. Whatever the stance any sHRI researcher may take, at the minimum, ethical dimensions need to be considered and reported on for transparency. I found through a literature study on ethical reporting that these ethical dimensions are drastically under-reported.

In Paper V, how NARS (Nomura et al., 2004) is used in sHRI was investigated. This paper was realized after the initial investigation of the potential of using NARS in the framework. Inconsistencies in how researchers within the field used the scale

were noticed, which made it hard to interpret and compare results between studies that used NARS. A literature review was conducted in order to more systematically gain a richer understanding of how NARS is being used in sHRI research. The results confirm our initial concerns, revealing highly inconsistent usage of NARS.

As mentioned in the background, the HRI field is a growing interdisciplinary research field with several different perspectives regarding the fields frameworks, terminologies, theories, models, methods, and tools (Baxter et al., 2016; Lagerstedt and Thill, 2023). Careful consideration needs to be put into the development of the methods in HRI research. The findings from this thesis highlights some of these issues, with the aim to stimulate a conversation on how to move forward as a field. One of the main takeaways is the call for the consistency in method use in the field. Specifically, future HRI researchers should follow ethical standards and report on these ethical aspects when investigating and presenting phenomena. Moreover, when HRI researchers use questionnaires like NARS, it should be used as originally intended. Any modification to questionnaires needs its own validation in order to ensure that the thing that is measured is actually measured. Being more consistent in use and reporting of questionnaires allows for meta-analysis, which makes it possible to examine more general phenomena. Ultimately, these results contribute to the continued development of the HRI field, specifically sHRI.

5.2 RQ2: WHAT IS THE ROLE AND RELEVANCE OF USERS' EXPECTATIONS IN sHRI?

Knowledge and insights into this research question is based on the empirical work implementing the proposed framework (RQ1), presented in Paper IV, Paper VI, and Paper VII.

With the developed framework as a basis, a systematic analysis of expectations in sHRI was conducted in an experimental setting. This experiment resulted in three publications. Paper IV revealed, through qualitative analysis and a UX focus, that participants (N=10) had very different experiences with the robot and that they interpreted the robot both as a machine and as a social agent. This caused confusion and negative user experience for participants in relation to the developed three UX factors of expectations; namely, affect, cognitive processing, and behavior & performance, which were derived from the consequences of expectations by Olsen et al. (1996). These results were extracted mainly from post-test interviews and observations during the interactions. There were overall improvements in the interaction quality between the first and second interaction in relation to cognitive processing and behavior & performance. For affect, the dominant experience was negative and stayed negative. The improvements in the interactions demonstrate how these users, to some degree, revised their expectations over time in order to align their expectations with the actual outcome of the interaction in order to improve the interaction quality. The improved interaction quality can also be tied to hypothesis testing, in line with the expectancy process by Olson et al. (1996), which is when users alter their behavior to test a hypothesis in order to verify their expectations. Four UX problems were identified in this analysis – (1) the

robot not being able to understand the user, by either not picking up on the users' speech or by mishearing the user and responding in strange ways, (2) the robot being interpreted as too simple and superficial in its dialogue despite the dialogue system being state-of-the-art, (3) users experiencing mixed emotions due to the anthropomorphic features, and (4) the robot failing to perform certain actions and commands when asked by the users.

In Paper VI, a quantitative analysis of the results from this empirical study (N=31) was presented, showing that previous experience was a substantial indicator of how their expectations changed over time in the interaction with the robot. Participants' responses were, to a large degree, decided before their first interaction with the robot. As a result, participants did not move towards group agreement and had a tendency to stick with their initial expectations. NARS, RAS, Closeness, and a question on perceived capability were used for the analysis. The analysis focused on the affect factor of expectations. Results showed that the variance of the responses did not decrease over the two interactions, which implies that users' expectations of robots were robust and did not revise when meeting and interacting twice with the robot in this study. There were changes in affect; however, as a group, the participants did not change their affect in the same direction. That is, affect changed both in positive and negative directions. Moreover, there was a statistically significant result for previous experience, which implies that participants' sources for their expectations seems to be a strong indicator for their expectations. These results would be in line with the expectancy process by Olson et al. (1996) which stresses how direct experience forms more certain and more accessible expectations. These results further strengthen the finding that participants' expectations were both robust and retained in the two interactions. Lastly, results revealed that some of the measures changed over time, while others did not. A possible explanation for these mixed results is that certain aspects of expectations stabilize at different times during an interaction, which would align with the results from Paper IV where larger revisions of user expectations were observed. The difference between these results observed in Paper IV and Paper VI indicates that affect is a relatively stable component of expectations, whereas cognitive processing and behavior are more frequently revised as a result of experience.

The results from Paper VI highlights the impact of previous experiences and how they can affect sHRI studies. Results found in sHRI studies might therefore not only reflect participants' affect, attitudes, opinions, and other subjective measures towards the robot *at hand*, but to a large degree also robots *in general*, predominantly based on an individual's sources of expectations. Considering that one single interaction for a shorter duration is common in the sHRI field, it is possible that much of the findings in this field have a strong confounding variable of users' expectations present in the obtained results.

In Paper VII, a reflexive thematic analysis was conducted to elaborate further on the qualitative dimensions of expectations. Findings from this analysis showed that user expectations are complex and often challenging to articulate explicitly. The results revealed how users adapt their approaches based on their interactions with the robot, highlighting positive user experience isn't exclusively governed

by interaction quality. It demonstrated a spectrum of confirmed expectations, emphasizing the significant role of past experiences and sources in shaping users' varied expectations of social robots.

The overall findings for this RQ are thus that expectations have a major role in sHRI. Due to the robot's anthropomorphic features, people can experience mixed expectations; people appear to expect social robots to be like humans while still viewing them as machines. There are revisions of expectations in relation to the expectation factors of cognitive processing and behavior & performance, but less for affect. People will also adapt their approaches based on their interactions with the robot, demonstrating how positive user experience isn't solely based on interaction quality. Furthermore, expectations can vary vastly based on their sources, direct versus indirect experiences, which means that participants will experience both high and low expectations due to what they have experienced previously. Individuals with previous first-hand experiences had significantly different expectations than those who did not have any first-hand experience. Participants with previous experiences of social robots seem to be more positive towards robots, whereas those who rely on other sources,e.g., social robots in media appear to be less certain and negative in their expectations towards robots. The implication of these results is that expectations are a strong confounding variable in sHRI research where participants may be primed by their previous, or lack of previous, experience with robots.

DISCUSSION

CHAPTER 6
DISCUSSION

This thesis started with an interest in understanding how people interpret social robots, which later formalized into an interest in peoples' *expectations* of social robots in human-robot interactions. Two research questions were constructed to approach the overall aim of investigating the role expectations play when interacting socially with social robots. The first research question (RQ1) asked how user expectations can be studied in sHRI, and the second research question (RQ2) asked what role and relevance users' expectations have in sHRI. The main contributions of this thesis, and the answers to the two research questions, are the theoretical development of the Social Robot Expectation Gap Evaluation Framework to study users' expectations of social robots and the empirical findings demonstrating how expectations play a major role in sHRI, including how previous experiences indicate what people expect of social robots.

In this chapter, I discuss these findings and their implications. I will also offer a set of guidelines for sHRI researchers.

6.1 IMPLICATIONS OF FINDINGS

There are several implications for the work presented in this thesis. First, understanding expectations of social robots has major implications for the sHRI field as it deepens the knowledge of how to study and use social robots within the field. Second, there is a societal implication as social robots are employed in several societal settings, and expectations affect how successful human-robot interactions may be. Lastly, there are practical implications for both designers of social robots because expectations can guide the development and design of social robots, and sHRI researchers because expectations can be managed by how we communicate about and present social robots.

At face value, it is quite obvious that social robots create different expectations than other technology. If we compare the expectations of social robots to phones, people know what to expect from a phone because we have plenty of direct experiences

with phones compared to social robots. Phones do not have artificial eyes or artificial body parts which might make the users believe that it is human-like. This is not to say that there are *no* social components to phones – there certainly are. We may talk to our phones at times, but we also know to what extent we can do so. Most likely, a person would not be afraid of a phone physically attacking them like they might with social robots (something one participant reported in our UX study, presented in Paper IV). People have, over time, gained enough direct experiences with phones to be able to know what is possible to do and what is not. Social robots, however, are not yet as common in our everyday life, with most people only having indirect experiences and inferences as their main source of their expectations (Olson, Roese, and Zanna, 1996; Oliveira and Yadollahi, 2023). From a user experience perspective, the way people experience and interact with social robots compared to other technologies is therefore vastly different.

What makes social robots particularly stand out from other technology, is the duality of viewing social robots as human-like *and* inanimate machines at the same time (this has been stressed previously, e.g., Alač, 2016 and Clark, 2023). Users' expectations of social robots are therefore quite complex and need to be further investigated and analyzed to be better understood in sHRI. This duality, of viewing social robots as human-like and machines, can even be found in our own research design. We use models of expectations from social psychology to study interaction design with machines. Viewing social robots as human-like and as machines lead to confusion, hesitation, bad user experience, and could ultimately cause the individual to stop using the robot. The outcome of not using the robot would be particularly detrimental in, for example, a care setting where a robot's task is to remind a patient of taking medication at certain intervals. Therefore, it becomes evident that the UX perspective is crucial in order to balance the duality of users' expectations. The expectations generated are highly individual and context-specific, for example, the type of robot, the environment, and the user's personal experience with robots.

Ultimately, because expectations are context-specific, the findings in this thesis highlight the importance of *how* and *what* HRI researchers communicate to the public because this information affects users' expectations. If we, for example, use anthropomorphic language when we talk about social robots we may build high expectations that cannot live up to the robot's actual capabilities, and disconfirmed expectations are then induced. This communication issue is not only directed to participants who partake in sHRI experiments but also of great relevance to the general public when we communicate our research or introduce our robots in media because the indirect sources of expectations also affect forthcoming human-robot interactions (Olson, Roese, and Zanna, 1996). From an ethical perspective, creating too high expectations of social robots by exaggerating robots' capabilities could be viewed as deception and feed the general fear of robots and AI depicted in the media (Cugurullo and Acheampong, 2023), which relates to the ongoing discussion on transparency in the HRI community (e.g., Felzmann et al., 2019; Wang et al., 2021; Winkle et al., 2021; van Straten, Peter, and Kühne, 2023; Zhong et al., 2023). Understanding user expectations allows for more efficient transparency in HRI as it can mitigate deception and create more ethical and

trustworthy interactions.

It is also important to highlight the designers' role when creating social robots. First of all, the designer needs to be aware of their own expectations which may not align with the intended users. The designer may think their robot meets the criteria of social interactions, but really it is merely technically impressive. If the designer does not understand their own expectations, it is hard to design a robot that will be employed successfully in another social context. The UX field has therefore developed procedures for grasping the intended users' preferences and needs in a certain context, identifying and formulating UX goals for the tasks to be achieved, and applying an iterative design and evaluation cycle (ISO: 9241-210:2010, 2.15; Hartson and Pyla, 2018). We could see this first-hand in our own empirical study. We used the OpenAI GPT-3 language model in our study right before the media's hype around this product occurred. We had, therefore, the opportunity to actually use state-of-the-art technology on users who did not know about it yet. Despite this, many of the participants were not impressed, which demonstrated our own bias of what we would expect in the interactions between the humans and the robot. Moreover, although users' had the possibility of achieving very complex interactions, many ended up with simple interactions and thus confirmed their own low expectations. The designers of social robots are faced with the difficult, but necessary task, of managing both their own expectations and their intended users' expectations.

Moreover, from an ethical point of view, design choices of social robots may uphold moral and social norms (Carlucci et al., 2015; Malle et al., 2015; Malle and Scheutz, 2019; Sparrow, 2020). There are legal and ethical considerations for norms because *"they are obeyed and enforced by communities"* (Malle, Bello, and Scheutz, 2019, p. 21). There are instances where a person is unable, or unwilling, to conform to norms that exist in the community. For example, certain members of the community may violate age norms (children and older people), autonomy norms (physically disabled people), and independence norms (care recipients). It is therefore important to consider the user's individual preferences and emotional and psychological needs when designing robots. Reflective design is one approach to combat held expectations that contribute to harmful norms (Dewey, 1933; Sengers et al., 2005; Bulman and Schutz, 2013; Nyhlén and Gidlund, 2019; Fronemann, Pollmann, and Loh, 2021). Reflective design is a norm-critical approach where the designers are constantly reflecting over the design process in order to avoid assumptions and models that are relied on. The goal of reflective design is to make such implicit assumptions (including norms) explicit.

With the social robot expectation gap as a foundational concept for understanding expectations in sHRI (figure 2.6), a key requirement for successful human-robot interactions is the alignment of user expectations with the actual capabilities of the social robot. As sHRI researchers, our goal is to bridge this expectation gap, and this involves addressing two essential aspects of user expectations.

Firstly, it's crucial to comprehend users' unique expectations, preferences, and emotional and psychological needs. These individual and context-specific expectations play a significant role in how users perceive and interact with social

robots. Secondly, once these expectations are understood, they should be explicitly identified and properly managed during the iterative development of the robot. Effective communication and design recommendations are vital for ensuring that user expectations are met and even exceeded.

However, the individual and context-specific nature of user expectations poses a challenge to the tools commonly used in sHRI research, such as questionnaires. Many questionnaires are designed with a specific platform in mind and may not capture the nuances of expectations that vary from one user to another and evolve over time. This thesis highlights the need to study the change and variability in individuals over time, which is often overlooked in current sHRI research. Our proposed framework offers a systematic approach to address this variability by assuming that people and robots differ, not only from study to study, but also from one point in time to another. By recognizing and addressing these dynamics, we can enhance our understanding of user expectations and ultimately improve the quality of human-robot interactions.

From a larger perspective, expectations and how they evolve can provide valuable insights about society at large. The societal landscape is continuously reshaped by technological advancements including the advancement of social robots. When a society begins to view social robots as integral components of everyday existence, it signifies a shift in expectations that reflects the ongoing societal maturity (Floridi, 2016). With each technological advancement and every interaction with social robots, users' expectations are redefined. Therefore, by studying expectations of social robots we can gain insight into the maturity of a society (Floridi, 2016). Moreover, as societies mature, our expectations of social robots will be to a larger degree met. When users' expectations of social robots are met, we can increase trust, acceptance, and positive user experience, which leads to successful integration of social robots into society.

We can only speculate about the future of social robots in society; however, it seems like the future may bring unforeseen transformations in line with our evolving expectations, where digital technologies like social robots become a norm. As articulated by Floridi (2016, p. 4):

> Information societies are maturing all over the world. More will appear in the future. In terms of expectations, similarities therefore will increase. To paraphrase Tolstoy, all mature information societies are alike in terms of people's expectations; each immature society is immature in its own way. So the next stage in the development of information societies, be this in ten or a hundred years, will not be a further maturation of their inhabitants' expectations about their digital affordances, it will be an unprecedented and unforeseen transformation altogether, for which the digital will have become an implicitly expected backdrop.

6.2 GUIDELINES FOR UNDERSTANDING AND REDUCING THE SOCIAL ROBOT EXPECTATION GAP IN sHRI

In the context of the findings identified in this thesis, I present a set of of guidelines that can be used to understand and reduce the social robot expectation gap. These guidelines relate to four key areas: being clear about the role of expectations, tracking expectations over time, the impact of expectations on designers and researchers, and transparency. While they are primarily formulated to support sHRI researchers, they can also be beneficial for various professionals, including designers, developers, and others involved in the research, development, and usage of social robots. The first three guidelines focus on research practices, and the last one focuses on the relationship between sHRI researchers and society.

6.2.1 GUIDELINE 1: UNDERSTAND AND CONTROL EXPECTATIONS AS VARIABLES IN STUDIES

To effectively handle user expectations in sHRI studies, it is essential to consider and control variables related to expectations. As presented in this thesis, user expectations are a factor that influences user responses and should be considered when interpreting sHRI results. There are three strategies that can be taken when considering expectations in sHRI.

Treating expectations as an independent variable

Account for users' expectations of social robots by measuring aspects of expectations. One way to do this is considering expectations as a demographic, distinguishing between participants with direct experiences and those with indirect experiences. It's also crucial to explore the nature of these experiences, including what kind of indirect and direct experience users have and how extensive these are. For example, the influence of media in indirect experiences. By doing so, user expectations can be incorporated into the context when analyzing and interpreting data.

Treating expectations as a dependent variable

Focusing on user expectations as the research topic deepens our understanding of expectations in sHRI. The Social Robot Expectation Gap Evaluation Framework can be used in this endeavor. The framework facilitates systematic research and evaluation of user expectations over time, helping identify gaps and patterns in these expectations.

Treating expectations as a confounding variable

User expectations are an inherent part of any research involving social robots. Therefore, at the very least, expectations should be considered as a confounding variable. This involves understanding how expectations can be influenced by factors such as the environment and the robot itself (design and behavior). There are strategies to account for expectations as a confounding variable. For

example, *restriction* is when participants with similar expectations are chosen for the study, and for between subject design studies, *matching* is when participants' expectations are matched with their counterparts in the other group.

6.2.2 GUIDELINE 2: TRACK EXPECTATIONS TEMPORALLY

Recognize that user expectations are not static but are constantly evolving, both outside and inside the lab. Essentially, this means that the group identity of the participants can change. A larger hypothesis presented in this thesis is that users interacting with a social robot will move toward a consensus. To study these dynamics, it's essential to view expectations as a variable that can change over time, influenced by various factors.

When considering expectations as a dependent variable, users' expectations of the robot's capabilities can shift and develop over the course of an interaction. This perspective allows researchers to assess whether these expectations are moving closer to being met or not. Factors such as the duration of the interaction, the design and behavior of the robot, and the specific context of the interaction all come into play when examining how and why expectations evolve. By systematically tracking these temporal changes in users' expectations, researchers gain valuable insights into the dynamic nature of sHRI and can adapt their approach to design for positive user experience.

6.2.3 GUIDELINE 3: FOSTER CRITICAL SELF-REFLECTION

Encourage critical self-reflection among researchers to recognize how their own expectations can influence the research process. sHRI researchers often have a predisposition toward technology and robots, which in itself can create specific expectations and might lead to unintentional biases. These expectations may impact what is considered obvious or not, and what aspects of the research process and design will and will not affect participants. To address this, it's crucial to introspect and consider how personal perspectives shape one's view. This is closely tied to expectations, because our inherent biases can influence the research process.

It's important to reinforce a social-cultural perspective within the research process itself, recognizing that there is no one-size-fits-all approach. This practice is valuable to both qualitative and quantitative research, especially when examining subjective measures. The willingness to revisit and challenge these preconceived ideas is fundamental to both producing well-rounded and unbiased research in the field of sHRI and to accurately study user expectations.

6.2.4 GUIDELINE 4: PRIORITIZE TRANSPARENCY IN DISSEMINATION

Be transparent when communicating the capabilities and limitations of social robots to users, recognizing the profound impact of communication on shaping user expectations. As a sHRI researcher, we communicate about social robots

in a diverse set of communications channels, ranging from scientific, technical, popular, to commercial domains. By fostering a transparent dissemination, users can receive consistent and reliable information about what to expect from social robots in certain situations. When communication is clear, detailed, and candid, users are better equipped to form accurate expectations about what the robot can and cannot do.

In this context, the pre-interaction phase is crucial, but it should not be limited to this alone. The practice of transparent communication should extend to post-interaction debriefing sessions, reinforcing the alignment of user expectations with actual robot capabilities. Sometimes, omitting information may be necessary for a research study, for example when the WoZ technique is used. Including a debriefing session afterwards is crucial in order to be transparent and to avoid deceptive practices that may affects users' expectations.

It is also important to tailor the communication for the specific user audience, adapting language and messaging to align with the user's level of technical understanding and familiarity with social robots. Ultimately, by prioritizing transparency, you create a foundation for users to develop realistic expectations, fostering a more positive user experience with social robots.

6.3 LIMITATIONS AND FUTURE WORK

Due to the Covid-19 pandemic, we were only able to conduct one empirical study using our framework. This means that the main answer to RQ2 is based on one study. More empirical studies would have both strengthened the findings of this RQ and further validated the framework. However, due to the time lost during the pandemic, we were able to construct a rigorous experiment where people interacted first-hand with a robot and where we could extract a substantial amount of data that could be analyzed with different methods. The pandemic also allowed for substantial theoretical work to be put toward the development of the framework.

In terms of the framework, there are many directions future work could go. First of all, although Paper IV is a start for the validation of the framework, more work needs to be put into this endeavor. In Paper IV, I analyzed only some of the metrics that were formulated for the framework and used for the empirical study. Therefore, future work will include analyzing all of the metrics proposed in Paper III with the specific purpose of validating the framework's use. There are also some future direction in terms of developing the framework further. For example, in Paper IV, we identified that there was overlap between the expectation factors. For example, RAS is a metric proposed for the affect expectation factor; however, RAS subscale 1 relates to behavioral characteristics, which makes it apt for the behavioral & performance factor as well. Therefore, further development in terms of the factors of expectations could be of interest. Indeed, affect, cognitive processing, and behavior & performance might not be as easy to separate as suggested by Olson et al. (1996). With this in mind, there are more aspects of expectations, that are not the factors of expectations, that could be added to the framework. For example, as

mentioned in the Background, there are four properties of expectations according to the expectancy process (Olson, Roese, and Zanna, 1996): certainty, accessibility, explicitness, and importance. We have seen connections between these properties and our results; for example, it seems like participants are more certain of their expectations when they have had previous experience with social robots (Paper VI). Future work could look into these properties further, including incorporating these more overtly to an updated framework. Moreover, the subjective measures in the framework do not further investigate the differences between expectations for robots in general and expectations of the particular robot under consideration. The subjective measures are designed to gauge expectations in a broader context, encompassing robots as a whole. It is worth noting that both NARS and RAS are commonly used as subjective measures after interacting with specific robots. In line with this, the framework's primary aim is to observe the evolution of these general expectations through interactions with specific robots. Future work could explore how the expectations of robots in general may differ from the expectations of the robot in the current study, making such differences distinct.

In terms of limitations in the empirical study, there a few drawbacks I would like to mention. Firstly, we only had one robot in the study. We can therefore not say that we can generalize these results to other social robots, however, this is also one of the main points of this thesis. In fact, in line with the expectancy process by Olson et al. (1996), it is likely that another robot would yield different results because they are not designed the same way and thus would create different expectations. For example, some robots may trigger uncanny valley (Mori, 1970; MacDorman, 2006) which may affect the results, such as higher scores on NARS and RAS. However, the major findings related to previous experiences would probably be similar to our findings because the participants had made up their minds before meeting the robot and kept those expectations even after interacting with the robot in the experiment.

Second, there are limitations that are inherent to most study designs. In general, people participating in studies are never truly "neutral" (Orne, 2009). Participants come into a study with their own understanding of a situation. This is something I have previously stressed as an important factor of human-robot interaction; in fact, it is a major point for why studying expectations is important. Despite the awareness of expectations, there are aspects of this non-neutral disposition from participants that affected the study. One such limitation is demand characteristics, where participants are aware of how their responses come off and alter their behavior accordingly. Although I refrained from mentioning before the interactions that we were studying expectations, I did inform the participants that I was investigating how people interact with robots in the home. It is therefore possible that the participants altered their behavior after what they thought was an appropriate way to act. Participants may have done this in order to produce "good data" that would be in line with what they would imagine a good participant would do (Orne, 2009). This altered behavior may have affected the interactions, the questionnaire responses, and the post-test interviews. It is hard to address this limitation in the study design; however, as already mentioned, we did not explicitly tell the participants that we were studying expectations beforehand (however,

participants were debriefed afterwards that we were investigating expectations). In addition, gathering both quantitative and qualitative measures allowed us to compare how people acted in the interactions, what people responded on the questionnaires, and what they actually said about the interaction afterwards. This way, we could screen for demand characteristics, and I did in fact see that what people said and what people did can differ, which is addressed in Paper IV. I also highlight the potential issues of the questionnaires we use in the field, addressed in Paper V and Paper VI.

Lastly, because we found that previous experiences were a strong indicator of people's expectations of the robot, I would have liked to include more in-depth questions about what sort of previous experiences the participants had before the study. Asking for details regarding the specific kind of previous experiences could have given key insights into if the differences in the kind of previous experiences would also cause differences in the expectations. Future work, therefore, includes investigating the different types of previous experiences, if any, and how they affect the interaction.

We still need to analyze certain parts of the collected data. First, the analysis we have done on the data has been without a linguistic perspective and there are many avenues in this direction that could be interesting. For example, I would like to look further into the kinds of questions asked by the participants to the robot. Based on what I saw as the test leader, it seems like participants with higher expectations asked more complex questions (e.g., if the robot was aware of the James Webb Telescope) whereas participants with lower expectations asked simpler questions (e.g., the robot's name and age). The findings from this kind of analysis could inform further how we can manage expectations to create more reciprocal and flowing interactions rather than one-sided interactions across participants.

Many participants expressed during the post-test interviews that they experienced feelings of awkwardness towards the robot, for example when Pepper did not respond to their questions. This is a strong indicator that they view the robot as a social being, which could be explored more in the data analysis.

There is also data related to speech recognition that could be analyzed further. In Paper XIII, we presented anecdotal results from the study where we saw indications that participants who deviated from the male voice and American accent had a harder time being "understood" and recognized by the robot. Some accents in languages are typically considered to have "no accent", which is a myth perpetuated by in-group biases based on geographical location and social statuses such as class, age, gender, and education (Chambers, 2009; Kinzler, Corriveau, and Harris, 2011). These biases can also show up in speech recognition systems used in robots based on the type of model that is used to train speech recognition. There is related work from HCI and HRI literature (e.g., Kennedy et al., 2017; Caliskan, Bryson, and Narayanan, 2017; Irfan et al., 2020; Ngueajio and Washington, 2022). However, there appears to be sparse reporting specifically on speech recognition. The secondary findings from this study were too few and too different from one another to perform any type of statistical analysis. I would therefore like to explore this further by performing a qualitative analysis by either deductively looking at

these cases through existing theories or categories, or inductively by investigating the themes or patterns that would emerge from the data (Patton, 2014).

Another related avenue I would have liked to explore outside of the empirical study is the ethical perspective of expectations, specifically the deception aspect of building high expectations of social robots. There are several directions this could take. Apart from building on the workshop contributions (Paper VIII, Paper XI, Paper XII, and Paper XIII), an empirical investigation could be performed with a focus on participants' opinions towards being deceived in this manner, including if they would view it as deception and, if so, how to mitigate it.

6.4 CONCLUDING REMARKS

In this thesis, I have theoretically and empirically characterized the role and relevance of users' expectations when interacting with social robots. The major outcome of this work is the Social Robot Expectation Gap Evaluation Framework. As social robots become more integrated in society, how people interact with them also becomes increasingly important. The results found in this thesis can inform social robot designers and sHRI researchers on how to manage expectations, both in the design and in the presentation of social robots. Increased attention to expectations can reduce the expectation gap, and by extension contribute to improved user experience. The expectations can be managed, first, by identifying what kind of expectations users have of a social robot, then designing the robot accordingly and communicating appropriately to the user. As demonstrated in this work, people expect social robots to be both inanimate machines and human-like interaction partners. The duality of these expectations set social robots apart from other artificial artifacts, and mitigating this confusion will create more successful human-robot interactions. Moreover, the results demonstrate how the source of the expectations have a direct effect on upcoming interactions, which stresses the unique experiences users' have, even in the same kind of interaction as other users. Therefore, understanding the individual needs of the users, and not relying on norms, is also of utmost importance in managing expectations. Social robots are intended to be useful in many societal settings and reducing the social robot expectation gap is a step in the right direction of this continued effort.

REFERENCES

REFERENCES

Ahmed, Imran (2021). “Dismantling the anti-vaxx industry”. In: *Nature Medicine* 27.3, pp. 366–366.

Alač, Morana (2016). “Social robots: Things or agents?” In: *AI & society* 31, pp. 519–535.

Aldebaran (2023). *http://www.aldebaran.com*.

Alenljung, Beatrice, Lindblom, Jessica, Andreasson, Rebecca, and Ziemke, Tom (2019). “User experience in social human-robot interaction”. In: *Rapid automation: Concepts, methodologies, tools, and applications*. IGI Global, pp. 1468–1490.

Alves-Oliveira, Patricia, Ribeiro, Tiago, Petisca, Sofia, Di Tullio, Eugenio, Melo, Francisco S, and Paiva, Ana (2015). “An empathic robotic tutor for school classrooms: Considering expectation and satisfaction of children as end-users”. In: *Social Robotics: 7th International Conference, ICSR 2015, Paris, France, October 26-30, 2015, Proceedings 7*. Springer, pp. 21–30.

Aly, Amir, Griffiths, Sascha, and Stramandinoli, Francesca (2017). “Metrics and benchmarks in human-robot interaction: Recent advances in cognitive robotics”. In: *Cognitive Systems Research* 43, pp. 313–323.

American Psychological Association (2017). *Ethical principles of psychologists and code of conduct*.

Andrews, Kristin (2020). *The animal mind: An introduction to the philosophy of animal cognition*. Routledge.

Aron, Arthur, Aron, Elaine N, and Smollan, Danny (1992). “Inclusion of other in the self scale and the structure of interpersonal closeness.” In: *Journal of personality and social psychology* 63.4, p. 596.

Bar, Moshe (2007). “The proactive brain: using analogies and associations to generate predictions”. In: *Trends in cognitive sciences* 11.7, pp. 280–289.

Bates, Joseph (1994). “The role of emotion in believable agents”. In: *Communications of the ACM* 37.7, pp. 122–125.

Baxter, Paul, Kennedy, James, Senft, Emmanuel, Lemaignan, Severin, and Belpaeme, Tony (2016). “From characterising three years of HRI to methodology and reporting recommendations”. In: *2016 11th acm/ieee*

international conference on human-robot interaction (hri). IEEE, pp. 391–398.
Beavis, Allan K and Thomas, A Ross (1996). "Metaphors as storehouses of expectation: Stabilizing the structures of organizational life in independent schools". In: *Educational Management & Administration* 24.1, pp. 93–106.
Bennett, Casey C (2021). "Evoking an intentional stance during human-agent social interaction: Appearances can be deceiving". In: *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*. IEEE, pp. 362–368.
Berger, Charles R and Calabrese, Richard J (1974). "Some explorations in initial interaction and beyond: Toward a developmental theory of interpersonal communication". In: *Human communication research* 1.2, pp. 99–112.
Berzuk, James M and Young, James E (2023). "Clarifying Social Robot Expectation Discrepancy: Developing a Framework for Understanding How Users Form Expectations of Social Robots". In: *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 231–233.
Billing, Erik, Rosén, Julia, and Lamb, Maurice (2023). "Language Models for Human-Robot Interaction". In: *Companion of the 18th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2023), Stockholm, Sweden, March 13–16, 2023*. Sweden, USA: ACM, pp. 905–906.
Billing, Erik, Rosén, Julia, and Lindblom, Jessica (2019). "Expectations of Robot Technology in Welfare". In: *The second workshop on social robots in therapy and care, in conjunction with the 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2019), Daegu, Korea, March 11–14 2019*, pp. 1–4.
Billings, Deborah R, Schaefer, Kristin E, Chen, Jessie YC, and Hancock, Peter A (2012). "Human-robot interaction: developing trust in robots". In: *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pp. 109–110.
Borg Jr, MB and Porter, Leroy (2010). "Following the life-course of an expectation: Examining the exchange of expectations in a homeless shelter in New York City". In: *The psychology of expectations*. Ed. by P Le-n and N. Tamez. New York: Nova Science Publishers Happague, pp. 1–47.
Braun, Virginia and Clarke, Victoria (2006). "Using thematic analysis in psychology". In: *Qualitative research in psychology* 3.2, pp. 77–101.
— (2021a). "Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches". In: *Counselling and psychotherapy research* 21.1, pp. 37–47.
— (2021b). "One size fits all? What counts as quality practice in (reflexive) thematic analysis?" In: *Qualitative research in psychology* 18.3, pp. 328–352.
Breazeal, Cynthia (2003). "Emotion and sociable humanoid robots". In: *International journal of human-computer studies* 59.1-2, pp. 119–155.
— (2004). "Social interactions in HRI: the robot view". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 34.2, pp. 181–186.

Bubic, Andreja, Von Cramon, D Yves, and Schubotz, Ricarda I (2010). "Prediction, cognition and the brain". In: *Frontiers in human neuroscience*, p. 25.

Bulman, Chris and Schutz, Sue (2013). *Reflective practice in nursing*. John Wiley & Sons.

Burgoon, Judee K and Jones, Stephen B (1976). "Toward a theory of personal space expectations and their violations". In: *Human communication research* 2.2, pp. 131–146.

Byrne, David (2022). "A worked example of Braun and Clarke's approach to reflexive thematic analysis". In: *Quality & quantity* 56.3, pp. 1391–1412.

Caliskan, Aylin, Bryson, Joanna J, and Narayanan, Arvind (2017). "Semantics derived automatically from language corpora contain human-like biases". In: *Science* 356.6334, pp. 183–186.

Carlucci, Fabio Maria, Nardi, Lorenzo, Iocchi, Luca, and Nardi, Daniele (2015). "Explicit representation of social norms for social robots". In: *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 4191–4196.

Chambers, Jack K. (2009). *Sociolinguistic Theory: Linguistic Variation and its Social Significance*. Revised. Wiley-Blackwell.

Clark, Andy (2013). *Mindware: An Introduction to the Philosophy of Cognitive Science*. OUP USA. ISBN: 9780199828159.

Clark, Herbert H and Fischer, Kerstin (2023). "Social robots as depictions of social agents". In: *Behavioral and Brain Sciences* 46, e21.

Coeckelbergh, Mark (2011). "Are emotional robots deceptive?" In: *IEEE transactions on affective computing* 3.4, pp. 388–393.

Cole, David (2023). "The Chinese Room Argument". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta and Uri Nodelman. Summer 2023. Metaphysics Research Lab, Stanford University.

Crick, Francis and Koch, Christof (2003). "A framework for consciousness". In: *Nature neuroscience* 6.2, pp. 119–126.

Cugurullo, Federico and Acheampong, Ransford A (2023). "Fear of AI: an inquiry into the adoption of autonomous cars in spite of fear, and a theoretical framework for the study of artificial intelligence technology acceptance". In: *AI & SOCIETY*, pp. 1–16.

Dautenhahn, Kerstin (2007a). "Methodology & themes of human-robot interaction". In: *International Journal of Advanced Robotic Systems* 4.1, p. 15.

— (2007b). "Socially intelligent robots: dimensions of human–robot interaction". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 362.14- 80, pp. 679–704.

— (2013). "Human-robot interaction". In: *The Encyclopedia of Human-Computer Interaction, 2nd Ed.*

— (2018). "Some Brief Thoughts on the Past and Future of Human-Robot Interaction". In: *ACM Transactions on Human-Robot Interaction (THRI)* 7.1, p. 4. ISSN: 2573-9522. DOI: 10.1145/3209769.

Dennett, Daniel C (1989). *The intentional stance*. MIT press.

Dewey, John (1933). *A restatement of the relation of reflective thinking to the educative process*. DC Heath.
De Wit, Jan, Pijpers, Laura, van den Berghe, Rianne, Krahmer, Emiel, and Vogt, Paul (2019). "Why ux research matters for hri: the case of tablets as mediators". In: *Workshop on the Challenges of Working on Social Robots that Collaborate with People, at the ACM CHI Conference on Human Factors in Computing Systems (CHI2019)*. ACM.
Duffy, Brian R (2002). "Anthropomorphism and robotics". In: *The society for the study of artificial intelligence and the simulation of behaviour* 20.
— (2003). "Anthropomorphism and the social robot". In: *Robotics and autonomous systems* 42.3-4, pp. 177–190.
Dumas, Joseph F and Redish, Janice C (1999). *A Practical Guide to Usability Testing*. Human/computer interaction. Intellect. ISBN: 9781841500201.
Edwards, Autumn, Edwards, Chad, Westerman, David, and Spence, Patric R (2019). "Initial expectations, interactions, and beyond with social robots". In: *Computers in Human Behavior* 90, pp. 308–314.
EuropeanComission (2018). *https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/h2020_ethics-soc-science-humanities_en.pdf*.
Felzmann, Heike, Fosch-Villaronga, Eduard, Lutz, Christoph, and Tamo-Larrieux, Aurelia (2019). "Robots and transparency: The multiple dimensions of transparency in the context of robot technologies". In: *IEEE Robotics & Automation Magazine* 26.2, pp. 71–78.
Festinger, Leon (1957). *A theory of cognitive dissonance*. Vol. 2. Stanford university press.
Fink, Julia (2012). "Anthropomorphism and human likeness in the design of robots and human-robot interaction". In: *International Conference on Social Robotics*. Springer, pp. 199–208.
Fischer, Kerstin (2011). "How people talk with robots: Designing dialog to reduce user uncertainty". In: *Ai Magazine* 32.4, pp. 31–38.
Floridi, Luciano (2016). "Mature information societies—a matter of expectations". In: *Philosophy & Technology* 29, pp. 1–4.
Fong, Terrence, Nourbakhsh, Illah, and Dautenhahn, Kerstin (2003). "A survey of socially interactive robots". In: *Robotics and autonomous systems* 42.3-4, pp. 143–166.
Fronemann, Nora, Pollmann, Kathrin, and Loh, Wulf (2021). "Should my robot know what's best for me? Human–robot interaction between user experience and ethical design". In: *AI & SOCIETY*, pp. 1–17.
FurhatRobotics (2023). *https://furhatrobotics.com*.
Gazziniga, Michael, Ivry, Richard, and Mangun, George (2014). *Cognitive neuroscience: the biology of the mind 4th ed*. WW Norton.
Goodrich, Michael A, Schultz, Alan C, et al. (2008). "Human–robot interaction: a survey". In: *Foundations and Trends® in Human–Computer Interaction* 1.3, pp. 203–275.
Hart, Chris (2018). *Doing a Literature Review: Releasing the Research Imagination*. SAGE Study Skills Series. SAGE Publications.
Hartson, Rex and Pyla, Pardha S (2018). *The UX book*. Morgan Kaufmann.

Hassenzahl, Marc and Tractinsky, Noam (2006). “User experience - a research agenda”. In: *Behaviour & Information Technology* 25.2, pp. 91–97.
Heider, Fritz and Simmel, Marianne (1944). “An experimental study of apparent behavior”. In: *The American journal of psychology* 57.2, pp. 243–259.
Heider, Fritz. (2015). *The Psychology of Interpersonal Relations*. Martino Fine Books. ISBN: 9781614277958.
Henschel, Anna, Hortensius, Ruud, and Cross, Emily S (2020). “Social cognition in the age of human–robot interaction”. In: *Trends in Neurosciences* 43.6, pp. 373–384.
Hohwy, Jakob (2013). *The predictive mind*. OUP Oxford.
Hone, Kate S and Graham, Robert (2000). “Towards a tool for the subjective assessment of speech system interfaces (SASSI)”. In: *Natural Language Engineering* 6.3-4, pp. 287–303.
Horstmann, Aike C and Krämer, Nicole C (2019). “Great expectations? Relation of previous experiences with social robots in real life or in the media and expectancies based on qualitative and quantitative assessment”. In: *Frontiers in psychology* 10, p. 939.
— (2020a). “Expectations vs. actual behavior of a social robot”. In: *Plos one* 15.8, e0238133.
— (2020b). “When a Robot Violates Expectations”. In: *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 254–256.
Irfan, Bahar, Hellou, Mehdi, Mazel, Alexandre, and Belpaeme, Tony (2020). “Challenges of a real-world HRI study with non-native english speakers: Can personalisation save the day?” In: *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 272–274.
Jokinen, Kristiina and Wilcock, Graham (2017). “Expectations and first experience with a social robot”. In: *Proceedings of the 5th International Conference on Human Agent Interaction*, pp. 511–515.
Jost, Céline, Le Pévédic, Brigitte, Belpaeme, Tony, Bethel, Cindy, Chrysostomou, Dimitrios, Crook, Nigel, Grandgeorge, Marine, and Mirnig, Nicole (2020). *Human-Robot Interaction*. Springer.
Kahn Jr, Peter H, Ishiguro, Hiroshi, Friedman, Batya, Kanda, Takayuki, Freier, Nathan G, Severson, Rachel L, and Miller, Jessica (2007). “What is a human?: Toward psychological benchmarks in the field of human–robot interaction”. In: *Interaction Studies* 8.3, pp. 363–390.
Kennedy, James, Lemaignan, Séverin, Montassier, Caroline, Lavalade, Pauline, Irfan, Bahar, Papadopoulos, Fotios, Senft, Emmanuel, and Belpaeme, Tony (2017). “Child speech recognition in human-robot interaction: evaluations and recommendations”. In: *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 82–90.
Khan, Sara and Germak, Claudio (2018). “Reframing HRI design opportunities for social robots: Lessons learnt from a service robotics case study approach using UX for HRI”. In: *Future internet* 10.10, p. 101.
Kinzler, Katherine D, Corriveau, Kathleen H, and Harris, Paul L (2011). “Children’s selective trust in native-accented speakers”. In: *Developmental science* 14.1, pp. 106–111.

Kveraga, Kestutis, Ghuman, Avniel S, and Bar, Moshe (2007). “Top-down predictions in the cognitive brain”. In: *Brain and cognition* 65.2, pp. 145–168.

Kwon, Minae, Jung, Malte F, and Knepper, Ross A (2016). “Human expectations of social robots”. In: *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, pp. 463–464.

Lagerstedt, Erik and Thill, Serge (2020). “Benchmarks for evaluating human-robot interaction”. In: *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, pp. 137–143.

— (2023). “Multiple roles of multimodality among interacting agents”. In: *ACM Transactions on Human-Robot Interaction* 12.2, pp. 1–13.

Lakoff, George and Johnson, Mark (2008). *Metaphors we live by*. University of Chicago press.

Lee, John D and See, Katrina A (2004). “Trust in automation: Designing for appropriate reliance”. In: *Human factors* 46.1, pp. 50–80.

Lee, Jong-Eun Roselyn and Nass, Clifford I (2010). “Trust in computers: The computers-are-social-actors (CASA) paradigm and trustworthiness perception in human-computer communication”. In: *Trust and technology in a ubiquitous modern environment: Theoretical and methodological perspectives*. IGI Global, pp. 1–15.

Lewis, Michael, Sycara, Katia, and Walker, Phillip (2018). “The role of trust in human-robot interaction”. In: Springer International Publishing, pp. 135–159.

Lindblom, J. (2015). *Embodied social cognition. Cognitive systems monographs (COSMOS)*. Springer International Publishing Switzerland.

Lindblom, Jessica, Alenljung, Beatrice, and Billing, Erik (2020). “Evaluating the user experience of human–robot interaction”. In: *Human-Robot Interaction*. Springer, pp. 231–256.

Lindblom, Jessica and Andreasson, Rebecca (2016). “Current challenges for UX evaluation of human-robot interaction”. In: *Advances in ergonomics of manufacturing: Managing the enterprise of the future*. Springer, pp. 267–277.

Lindblom, Jessica, Rosén, Julia, Lamb, Maurice, and Billing, Erik (Manuscript). “Disentangling People’s Experiences and Expectations when Interacting with the Social Robot Pepper: A Qualitative Analysis”. In: *Manuscript for scientific journal*, pp. 1–41.

Lohse, Manja (2009). “The role of expectations in HRI”. In: *New Frontiers in Human-Robot Interaction*, pp. 35–56.

— (2010). *Investigating the influence of situations and expectations on user behavior: empirical analyses in human-robot interaction*. Phd Thesis.

Lyons, Joseph B and Guznov, Svyatoslav Y (2019). “Individual differences in human–machine trust: A multi-study look at the perfect automation schema”. In: *Theoretical Issues in Ergonomics Science* 20.4, pp. 440–458.

MacDorman, Karl F. (2006). “Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley”. In:

ICCS/ CogSci-2006 long symposium: Toward social mechanisms of android science, pp. 26–29.
Malle, Bertram F, Bello, Paul, and Scheutz, Matthias (2019). "Requirements for an artificial agent with norm competence". In: *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 21–27.
Malle, Bertram F, Fischer, K, Young, J, Moon, A, and Collins, E (2020). "Trust and the discrepancy between expectations and actual capabilities". In: *Human-robot interaction: Control, analysis, and design*, pp. 1–23.
Malle, Bertram F and Scheutz, Matthias (2019). "Learning how to behave: Moral competence for social robots". In: *Handbuch Maschinenethik [Handbook of Machine Ethics]. Springer Reference Geisteswissenschaften*. 10, pp. 978–3.
Malle, Bertram F, Scheutz, Matthias, Arnold, Thomas, Voiklis, John, and Cusimano, Corey (2015). "Sacrifice one for the good of many? People apply different moral norms to human and robot agents". In: *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, pp. 117–124.
Manzi, Federico, Massaro, Davide, Di Lernia, Daniele, Maggioni, Mario A, Riva, Giuseppe, and Marchetti, Antonella (2021). "Robots are not all the same: young adults' expectations, attitudes, and mental attribution to two humanoid social robots". In: *Cyberpsychology, Behavior, and Social Networking* 24.5, pp. 307–314.
Marchesi, Serena, Spatola, Nicolas, Perez-Osorio, Jairo, and Wykowska, Agnieszka (2021). "Human vs humanoid. A behavioral investigation of the individual tendency to adopt the intentional stance". In: *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 332–340.
Mascolo, Michael F and Bidell, Thomas R (2020). *Handbook of integrative developmental science: Essays in honor of Kurt W. Fischer*. Routledge.
McDaniel, Ellen and Gong, Gwendolyn (1982). "The language of robotics: Use and abuse of personification". In: *IEEE Transactions on professional communication* 4, pp. 178–181.
Meister, Martin (2014). "When is a robot really social? An outline of the robot sociologicus". In: *Science, Technology & Innovation Studies* 10.1, pp. 107–134.
Metta, Giorgio, Natale, Lorenzo, Nori, Francesco, Sandini, Giulio, Vernon, David, Fadiga, Luciano, Von Hofsten, Claes, Rosander, Kerstin, Lopes, Manuel, Santos-Victor, José, et al. (2010). "The iCub humanoid robot: An open-systems platform for research in cognitive development". In: *Neural networks* 23.8-9, pp. 1125–1134.
Moetesum, Momina and Siddiqi, Imran (2018). "Socially believable robots". In: *Human-Robot Interaction: Theory and Application* 1.
Moore, Roger K (2017). "Is spoken language all-or-nothing? Implications for future speech-based human-machine interaction". In: *Dialogues with Social Robots*. Springer, pp. 281–291.
Mori, Masahiro (1970). "Bukimi no tani (The uncanny valley)". In: *Energy* 7.4, pp. 33–35.

Nasr, Nancy (2021). “Overcoming the discourse of science mistrust: how science education can be used to develop competent consumers and communicators of science information”. In: *Cultural Studies of Science Education*, pp. 1–12.

Nass, Clifford, Steuer, Jonathan, Tauber, Ellen, and Reeder, Heidi (1993). “Anthropomorphism, agency, and ethopoeia: computers as social actors”. In: *INTERACT’93 and CHI’93 conference companion on Human factors in computing systems*, pp. 111–112.

Natarajan, M. and Gombolay, M. (2020). “Effects of anthropomorphism and accountability on trust in human robot interaction”. In: *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 33–42.

Ngueajio, Mikel K and Washington, Gloria (2022). “Hey ASR System! Why Aren’t You More Inclusive? Automatic Speech Recognition Systems’ Bias and Proposed Bias Mitigation Techniques. A Literature Review”. In: *HCI International 2022–Late Breaking Papers: Interacting with eXtended Reality and Artificial Intelligence: 24th International Conference on Human-Computer Interaction, HCII 2022, Virtual Event, June 26–July 1, 2022, Proceedings*. Springer, pp. 421–440.

Nomura, Tatsuya, Kanda, Takayuki, Suzuki, Tomohiro, and Kato, Kennsuke (2004). “Psychology in human-robot communication: An attempt through investigation of negative attitudes and anxiety toward robots”. In: *RO-MAN 2004. 13th IEEE international workshop on robot and human interactive communication (IEEE catalog No. 04TH8759)*. IEEE, pp. 35–40.

Nomura, Tatsuya, Kanda, Takayuki, Suzuki, Tomohiro, and Kato, Kensuke (2008). “Prediction of human behavior in human–robot interaction using psychological scales for anxiety and negative attitudes toward robots”. In: *IEEE transactions on robotics* 24.2, pp. 442–451.

Nyhlén, Sara and Gidlund, Katarina L (2019). “‘Everything’ disappears… reflexive design and norm-critical intervention in the digitalization of cultural heritage”. In: *Information, Communication & Society* 22.10, pp. 1361–1375.

Oliveira, Raquel and Yadollahi, Elmira (2023). “Robots in movies: a content analysis of the portrayal of fictional social robots”. In: *Behaviour & Information Technology*, pp. 1–18.

Olson, James M, Roese, Neal J, and Zanna, Mark P (1996). “Expectancies”. In: *Social psychology: Handbook of basic processes*. Ed. by E.T. Higgins and A.W. Kruglanski. Guilford Press, pp. 211–238.

OpenAI (2023). *https://openai.com/*.

Orne, Martin T (2009). “Demand characteristics and the concept of quasi-controls”. In: *Artifacts in behavioral research: Robert Rosenthal and Ralph L. Rosnow’s classic books* 110, pp. 110–137.

Paetzel, Maike, Perugia, Giulia, and Castellano, Ginevra (2020). “The persistence of first impressions: The effect of repeated interactions on the perception of a social robot”. In: *Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction*, pp. 73–82.

Patton, Michael Quinn (2014). *Qualitative research & evaluation methods: Integrating theory and practice*. Sage publications.

Pereira, Roberto, Baranauskas, M Cecilia C, and Liu, Kecheng (2015). "On the relationships between norms, values and culture: preliminary thoughts in HCI". In: *Information and Knowledge Management in Complex Systems: 16th IFIP WG 8.1 International Conference on Informatics and Semiotics in Organisations, ICISO 2015, Toulouse, France, March 19-20, 2015. Proceedings 16*. Springer, pp. 30–40.

Premack, David and Woodruff, Guy (1978). "Does the chimpanzee have a theory of mind?" In: *Behavioral and brain sciences* 1.4, pp. 515–526.

Reeves, Byron and Nass, Clifford (1996). "The media equation: How people treat computers, television, and new media like real people". In: *Cambridge, UK* 10.10.

Riek, Laurel D (2012). "Wizard of oz studies in hri: a systematic review and new reporting guidelines". In: *Journal of Human-Robot Interaction* 1.1, pp. 119–136.

Robins, Ben, Dautenhahn, Kerstin, and Dubowski, Janek (2004). "Investigating Autistic children's attitudes towards strangers with the theatrical robot-A new experimental paradigm in human-robot interaction studies". In: *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No. 04TH8759)*. IEEE, pp. 557–562.

Roese, Neal J and Sherman, Jeffrey W (2007). "Expectancy". In: *Social psychology: Handbook of basic processes*. Ed. by A.W. Kruglanski and E.T. Higgins. Guilford Press.

Rosén, Julia (2021). "Expectations in Human-Robot Interaction". In: *Advances in Neuroergonomics and Cognitive Engineering*. Ed. by Hasan Ayaz, Umer Asgher, and Lucas Paletta. Springer International Publishing, pp. 98–105.

Rosén, Julia, Billing, Erik, and Lindblom, Jessica (2023). "Applying the Social Robot Expectation Gap Evaluation Framework". In: *Human-Computer Interaction*. Ed. by Masaaki Kurosu and Ayako Hashizume. Springer International Publishing, pp. 169–188.

Rosén, Julia and Lagerstedt, Erik (2023). "Speaking Properly with Robots". In: *HRI'23 Workshop—Inclusive HRI II, Equity and Diversity in Design, Application, Methods, and Community, in conjunction with the 18th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2023), Stockholm, Sweden, March 13–16, 2023*, pp. 1–3.

Rosén, Julia, Lagerstedt, Erik, and Lamb, Maurice (2022). "Is Human-Like Speech in Robots Deception?" In: *HRI'22 Workshop—Robo-Identity 2, in conjunction with the 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2022), online, March 7–10, 2022*, pp. 1–3.

— (2023). "Investigating NARS: Inconsistent Practice of Application and Reporting". In: *The 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2023), Busan, South Korea, 2023*. IEEE, pp. 922–927.

Rosén, Julia, Lindblom, Jessica, and Billing, Erik (2021). "Reporting of Ethical Conduct in Human-Robot Interaction Research". In: *Advances in Human Factors in Robots, Unmanned Systems and Cybersecurity*. Ed. by

Matteo Zallio, Carlos Raymundo Ibañez, and Jesus Hechavarria Hernandez. Springer International Publishing, pp. 87–94.
Rosén, Julia, Lindblom, Jessica, and Billing, Erik (2022). "The Social Robot Expectation Gap Evaluation Framework". In: *Human-Computer Interaction. Technological Innovation*. Ed. by Masaaki Kurosu. Springer International Publishing, pp. 590–610.
Rosén, Julia, Lindblom, Jessica, Billing, Erik, and Lamb, Maurice (2021). "Ethical Challenges in the Human-Robot Interaction Field". In: *TRAITS Workshop, in conjunction with the 16th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2021), Boulder, USA, March 8–12, 2021*, pp. 1–2.
Rosén, Julia, Lindblom, Jessica, Lamb, Maurice, and Billing, Erik (2020). "Digital Human Modeling Technology in Virtual Reality–Studying Aspects of Users' Experiences". In: *DHM2020*. IOS Press, pp. 330–341.
— (Under review). "Previous Experience Matters: An In-Person Investigation of Expectations in Human-Robot Interaction". In: *Under review for scientific journal*, pp. 1–19.
Rosén, Julia, Richardson, Kathleen, Lindblom, Jessica, and Billing, Erik (2018). "The Robot Illusion: Facts and Fiction". In: *Workshop in Explainable Robotics System, in conjunction with 13th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2018), Chicago, USA, March, 5–8, 2018*, pp. 1–2.
Roto, Virpi, Law, EL-C, Vermeeren, Arnold POS, and Hoonhout, Jettie (2011). "User Experience White Paper – Bringing clarity to the concept of user experience. In Dagstuhl Seminar on Demarcating User Experience." In.
Rudling, Maja (2023). *Facilitators of communication and the development of autism: From responsiveness to basic communicative cues, to emerging pragmatic language use*.
Sandoval, Eduardo Benitez, Mubin, Omar, and Obaid, Mohammad (2014). "Human robot interaction and fiction: A contradiction". In: *Social Robotics: 6th International Conference, ICSR 2014, Sydney, NSW, Australia, October 27-29, 2014. Proceedings 6*. Springer, pp. 54–63.
Scassellati, Brian (2002). "Theory of mind for a humanoid robot". In: *Autonomous Robots* 12, pp. 13–24.
Schaefer, Kristin E (2016). "Measuring trust in human robot interactions: Development of the "trust perception scale-HRI"". In: *Robust intelligence and trust in autonomous systems*. Springer, pp. 191–218.
Schramm, Lena T, Dufault, Derek, and Young, James E (2020). "Warning: This robot is not what it seems! exploring expectation discrepancy resulting from robot design". In: *Companion of the 2020 ACM/IEEE international conference on human-robot interaction*, pp. 439–441.
Schwitzgebel, Eric (2019). "Belief". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Fall 2019. Metaphysics Research Lab, Stanford University.
Seligman, Martin EP (1972). "Learned helplessness". In: *Annual review of medicine* 23.1, pp. 407–412.

Sengers, Phoebe, Boehner, Kirsten, David, Shay, and Kaye, Joseph'Jofish' (2005). "Reflective design". In: *Proceedings of the 4th decennial conference on Critical computing: between sense and sensibility*, pp. 49–58.
Sharkey, Amanda and Sharkey, Noel (2021). "We need to talk about deception in social robotics!" In: *Ethics and Information Technology* 23, pp. 309–316.
Shourmasti, Elaheh Shahmir, Colomo-Palacios, Ricardo, Holone, Harald, and Demi, Selina (2021). "User experience in social robots". In: *Sensors* 21.15, p. 5052.
Simmons, Reid, Makatchev, Maxim, Kirby, Rachel, Lee, Min Kyung, Fanaswala, Imran, Browning, Brett, Forlizzi, Jodi, and Sakr, Majd (2011). "Believable robot characters". In: *AI Magazine* 32.4, pp. 39–52.
Sparrow, Robert (2020). "Robotics has a race problem". In: *Science, Technology, & Human Values* 45.3, pp. 538–560.
Steup, Matthias and Neta, Ram (2020). "Epistemology". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Fall 2020. Metaphysics Research Lab, Stanford University.
Van Straten, Caroline L, Peter, Jochen, and Kühne, Rinaldo (2023). "Transparent robots: How children perceive and relate to a social robot that acknowledges its lack of human psychological capacities and machine status". In: *International Journal of Human-Computer Studies* 177, p. 103063.
Suchman, Lucille Alice (2007). *Human-machine reconfigurations: Plans and situated actions*. Cambridge university press.
Thellman, Sam, Silvervarg, Annika, and Ziemke, Tom (2017). "Folk-psychological interpretation of human vs. humanoid robot behavior: Exploring the intentional stance toward robots". In: *Frontiers in Psychology* 8.NOV.
Thrun, Sebastian (2004). "Toward a framework for human-robot interaction". In: *Human–Computer Interaction* 19.1-2, pp. 9–24.
Tonkin, Meg, Vitale, Jonathan, Herse, Sarita, Williams, Mary-Anne, Judge, William, and Wang, Xun (2018). "Design methodology for the ux of hri: A field study of a commercial social robot at an airport". In: *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 407–415.
Ullman, Daniel and Malle, Bertram F (2018). "What does it mean to trust a robot? Steps toward a multidimensional measure of trust". In: *Companion of the 2018 acm/ieee international conference on human-robot interaction*, pp. 263–264.
Vernon, David (2014). *Artificial Cognitive Systems: A Primer*. London: The MIT Press, p. 265. ISBN: 0262028387.
Wang, Jianmin, Liu, Yujia, Yue, Tianyang, Wang, Chengji, Mao, Jinjing, Wang, Yuxi, and You, Fang (2021). "Robot Transparency and Anthropomorphic Attribute Effects on Human–Robot Interactions". In: *Sensors* 21.17, p. 5722.
Ward, Jamie (2015). *The Student's Guide to Cognitive Neuroscience*. Taylor & Francis.
Weiss, Astrid (2016). "Creating service robots for and with people: A user-centered reflection on the interdisciplinary research field of

human-robot interaction". In: *15th Annual STS Conference Graz, Critical Issues in Science, Technology, and Society Studies*.

Winkle, K., Caleb-Solly, P., Leonards, U., Turton, A., and Bremner, P. (2021). "Assessing and Addressing Ethical Risk from Anthropomorphism and Deception in Socially Assistive Robots". In: *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 101–109.

Winkle, Katie, Jackson, Ryan Blake, Melsión, Gaspar Isaac, Brščić, Dražen, Leite, Iolanda, and Williams, Tom (2022). "Norm-breaking responses to sexist abuse: A cross-cultural human robot interaction study". In: *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, pp. 120–129.

Winkle, Katie, Lagerstedt, Erik, Torre, Ilaria, and Offenwanger, Anna (2023). "15 Years of (Who) man Robot Interaction: Reviewing the H in Human-Robot Interaction". In: *ACM Transactions on Human-Robot Interaction* 12.3, pp. 1–28.

World Medical Association, Declaration of Helsinki (2018). *https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/*.

Zhang, Dan and Wei, Bin (2020). *Human-Robot Interaction*. Cambridge Scholars Publisher.

Zhong, Mengyu, Fraile, Marc, Castellano, Ginevra, and Winkle, Katie (2023). "A case study in designing trustworthy interactions: implications for socially assistive robotics". In: *Frontiers in Computer Science* 5.1152532.

PUBLISHED AND SUBMITTED PAPERS

PAPER I

REPORTING OF ETHICAL CONDUCT IN HUMAN-ROBOT INTERACTION RESEARCH

Reprinted from Julia Rosén, Jessica Lindblom, and Erik Billing (2021). “Reporting of Ethical Conduct in Human-Robot Interaction Research”. In: *Advances in Human Factors in Robots, Unmanned Systems and Cybersecurity*. Ed. by Matteo Zallio, Carlos Raymundo Ibañez, and Jesus Hechavarria Hernandez. Springer International Publishing, pp. 87–94 with permission from Springer International.

Reporting of Ethical Conduct in Human-Robot Interaction Research

Julia Rosén(✉), Jessica Lindblom, and Erik Billing

Interaction Lab, University of Skövde, Högskolevägen 1, 541 28 Skövde, Sweden
julia.rosen@his.se

Abstract. The field of Human-Robot Interaction (HRI) is progressively maturing into a distinct discipline with its own research practices and traditions. Aiming to support this development, we analyzed how ethical conduct was reported and discussed in HRI research involving human participants. A literature study of 73 papers from three major HRI publication outlets was performed. The analysis considered how often the following five principles of ethical conduct were reported: ethical board approval, informed consent, data protection and privacy, deception, and debriefing. These five principles were selected as they belong to all major and relevant ethical guidelines for the HRI field. The results show that overall, ethical conduct is rarely reported, with four out of five principles mentioned in less than one third of all papers. The most frequently mentioned aspect was informed consent, which was reported in 49% of the articles. In this work, we aim to stimulate increased acknowledgment and discussion of ethical conduct reporting within the HRI field.

Keywords: Human-Robot Interaction · Ethics · Methodology

1 Introduction

There is an ongoing dialogue on how to shape the future of HRI [1, 2]. Baxter et al. [3] emphasize the importance of keeping the HRI field interdisciplinary while finding a common ground to employ research. It should be emphasized that a truly interdisciplinary perspective on HRI will require researchers to adopt a wider set of concepts, theories, and methods in their own research, which implies the need to read a broader spectrum of literature as well as correctly applying the methods therein [2]. Combining several different disciplines and research areas, as is necessary in the case of research within HRI, leaves the risk of misinterpretations of underlying epistemological, theoretical, and methodological foundations that may not be explicitly articulated among the different disciplines, and erroneously considered as common knowledge within a community. Therefore, such endeavors ought to be discussed from a methodological perspective, including its ethical practices.

The focus in this paper is ethical reporting when conducting HRI research with human participants. While there are already well established protocols concerning the ethical principles, consensus has still not been reached on when and how these guidelines

M. Zallio et al. (Eds.): AHFE 2021, LNNS 268, pp. 87–94, 2021.
https://doi.org/10.1007/978-3-030-79997-7_11

should be applied within the HRI field. Proper ethical conduct ought to be considered in any research field and is certainly not a controversial, nor new, claim. It is crucial in fields where human participants are involved as the well-being and interests of the participants must be considered. Controversial experiments such as the obedience experiment by Milgram from 1963 [4] and the Stanford prison simulation by Haney et al. [5] created a need to standardize how to protect participants [6]. Of course, these kinds of unethical studies are not being executed in HRI research today; however, it might sometimes be difficult to foresee how the participants will be impacted by an experiment, which makes proper ethical conduct crucial for any research field.

There are numerous ethical best practice documents created by e.g. foundations, associations, and lawmakers, to ensure empirical research is done ethically. In this work, we have found three common ethical guidelines that apply to empirical HRI research: *Ethical Principles of Psychologists and Code of Conduct* by the American Psychological Association (APA) [7], *Ethics in Social Science and Humanities* by the European Commission (EC) [8], and *Declaration of Helsinki* by the World Medical Association (WMA) [9]. In short, APA's set of ethical principles is relevant for any experimental psychology research involving participants, and aims to ensure that the empirical research being done is ethical and well-reasoned. These principles are standardized across the field and are considered crucial when doing any psychology research. EC has several documents for ethical conduct depending on the area of research. For this paper, we have chosen to include their document on *Ethics in Social Science and Humanities* since much of the empirical HRI research being done today usually falls in this category; any research conducted within the European Union needs to adhere to these guidelines. Lastly, WMA created the *Declaration of Helsinki* with ethical principles for any medical research involving participants. This document is used in many fields that deals with human participants, including the HRI field.

Within these guidelines, we have identified five recurring ethical principles. First, **ethical board approval** refers to a board or committee that is responsible for approving certain empirical research before the study is executed. An appropriate board depends on the regulations and laws that exist in the researchers' residing country and where the study is intended to take place. Usually, the researcher must provide information to the ethical board about the intended study and how it is intended to be carried out. Second, **informed consent** is a document provided to the participants covering the nature of the study that they are asked to participate in. The participant should be informed of key elements of their participation, including but not limited to: how long the study takes, that they can end their participation at any point, any limits of confidentiality, and how to get in touch with the researchers if any questions or concerns should arise. Third, **data protection and privacy** refers to researchers' responsibility to ensure that the participants' information and data are kept with integrity and treated as confidential. For example, studies involving EU citizens, the participants' data protection and privacy needs to be compliant with the General Data Protection Regulation Union [10]. Fourth, **deception** refers to when the participants are deceived in any way while participating in the study. Deception is sometimes necessary to attain certain results that would not be revealed otherwise [7]. For deception to be justified, the researchers must have ruled out all other options that do not involve deception, agree that the empirical research is

important enough to risk deceiving participants, and they must explicitly explain to the participants afterwards what part of the study was deceptive. Fifth, **debriefing** refers to a session after the study where participants are made aware of information that is deemed appropriate to disclose. One purpose of debriefing is to correct any misconceptions the participants may have regarding what they experienced during the study, including revealing any forms of deception and why it was necessary.

With the aim of stimulating an increased discussion and reporting of ethical principles, a literature study was conducted to gauge how ethical conduct is reported in empirical HRI research. We based our ethical reporting investigation on the five aforementioned ethical principles as these were developed as a way to protect participants and are relevant for HRI research.

2 Method

To reach an overview of how ethics are reported in the HRI field, especially more recently, three major publication outlets from the HRI field were analyzed - the ACM/IEEE International Conference on Human-Robot Interaction (HRI) [11], the IEEE International Conference on Robot & Human Interactive Communication (RO-MAN) [12], and the ACM Transactions on Human-Robot Interaction (THRI) [13]. The years analyzed were 2018 for the HRI conference, 2019 for the RO-MAN conference, and all articles/papers published from 2018 and 2019 in the THRI journal. For both the HRI conference and the THRI journal, all full-length papers were considered. HRI had 49 full length papers total and THRI had 31 full length papers total (editorials excluded). For the RO-MAN conference, however, a random selection of 40 papers was used to reduce the number to a similar amount as the two other outlets. In total, 120 papers were considered.

The sample used here can be seen as a form of data triangulation. Data triangulation is the usage of a variety of data sources in a study. Through the use of data triangulation, one explores whether the inferences from the empirical data are valid, and estimates consistency. Since the aim was to review how ethics are reported when participants are involved, we applied an inclusion criterion defined as studies that comprised human participants involved in an experimental setting. An experiment is defined as a study where certain variables are manipulated which "investigates cause and effect relationships, seeking to prove or disprove a causal link between a factor and an observed outcome" [p. 127, 14]. The literature review was an iterative process; after the inclusion criterion was applied, remaining papers were read in more detail and analyzed in terms of ethical conduct, considering the five common principles of APA [7], EC [8] and WMA [9], specifically: ethical board approval, informed consent, data and privacy, deception, and debriefing.

The purpose was to detect any explicit mentions of these five principles in relation to the reported experiments. If an aspect was mentioned, the paper was marked with a yes for that principle, otherwise, it was marked with a no. These principles were reported in numerous ways and to varying degrees. For example, we did not differentiate between a consent form that was reported in close detail (e.g. how the consent was given, what was included in it), and a consent form that was mentioned briefly (sometimes at other places than in the method section). For the scope of our literature study, we decided to not judge

the level of detail for each aspect in these papers, but only to note if there was any mention of them or not. Also, if the article discussed more than one experiment, we assigned a yes if the principle as mentioned in relation to at least one of these experiments. Thus, a no indicates that the principle was never mentioned for any experiment reported in the paper.

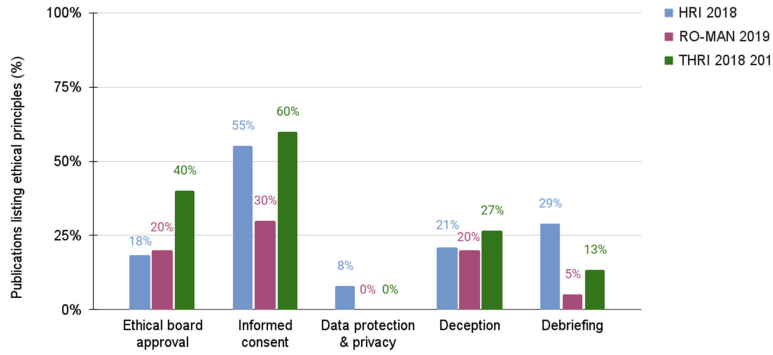

Fig. 1. The amount of ethical principles reported in 73 experiments from three major publication outlets in HRI.

Deception and debriefing were also chosen to be reported in the same matter: that is, they were included if the authors explicitly mentioned them. The exception was Wizard of Oz studies (WoZ), i.e. robots being remotely controlled by a human, where deception is necessary; unless mentioned that the participants were aware of the staging. Any publication that explicitly mentioned how the privacy of the participants was considered was marked with yes. As the phrasing can vary when it comes to data and privacy, a closer analysis was made in order to detect if this issue was mentioned. This was again mentioned to varying degrees, but we focused broadly and included any mention of it.

3 Results

A total of 73 publications were analyzed in relation to ethical principles. We would like to stress that these results cannot say whether the authors had considered these ethical principles in their experiments or not, but rather if the authors explicitly mentioned them in their publication. Below is a summary of the primary findings (Fig. 1).

Ethical board approval was explicitly mentioned in 23% of all the publications (7 in HRI, 4 in RO-MAN, and 6 in THRI). Thus a total of 17 out of the 73 publications explicitly mentioned that they had an ethical board approval to conduct the study.

Informed consent was explicitly mentioned in 49% of all the publications (21 in HRI, 6 in RO-MAN, and 9 in THRI). Thus, a total of 36 out of 73 publications explicitly mentioned the use of informed consent in their study.

Data protection and privacy was explicitly mentioned in 4% of all the publications. Three publications in HRI mentioned this aspect; however, none of the analyzed papers from RO-MAN and THRI discussed data protection and privacy. Thus, a total of 3 out of 73 publications explicitly mentioned that they considered data protection and privacy in their study.

Deception was explicitly mentioned as a method in 22% of all the publications (8 in HRI, 4 in RO-MAN, and 4 in THRI). Thus, a total of 16 out of 73 publications explicitly mentioned that deception was used in their study. Out of these 16 publications, a WoZ-set up was used as a method in 50% of those (8 out of 16 publications).

Debriefing was explicitly mentioned in 19% of all publications (11 in HRI, 1 in RO-MAN, and 2 in THRI). Thus, a total of 14 out of 73 publications explicitly mentioned the use of debriefing in their study. As discussed in the background, debriefing is critical in studies involving deception and when considering the publications that involved deception, 44% publications mentioned debriefing in relation to deception (7 out of 16 publications).

None of the five identified ethical principles were reported in 36% of the papers (26 out of the 73 publications). A follow up email was sent to corresponding authors of the 2018 HRI conference that had publications lacking information about both ethical board and informed consent. Out of 16 contacted authors, 7 responded that they did get ethical approval for their study and that all participants signed a consent form before participating in the study. One of these authors responded that the study was only conducted on lab members and therefore did not deem it appropriate to include informed consent and ethical board approval.

4 Discussion

The results from our literature study show that ethical conduct was rarely reported in the three publication outlets chosen. THRI reports ethical board approval more frequently; 40% of the analyzed publications mentioned board approval, compared to HRI and RO-MAN where about 20% of the analyzed publications mentioned this aspect. The HRI conference requires that the corresponding author checks a box that the study has been approved by relevant ethics committees if participants were involved. Due to this, it could be argued that it is not necessary to explicitly write that the study has been ethically approved and might be the reason for the low number reported in the reviewed papers. Despite this, it would be of value to list this in the publications, making it more obvious for the reader when, and how, ethical principles were considered. Readers that have not submitted to this conference before would not know that an ethical board approval box was checked before submission.

This argument does not, however, explain the low rate of reporting informed consent. Informed consent is the most commonly reported aspect of ethical conduct that we identified; however, it is still missing in half of all papers examined. HRI and THRI report informed consent more frequently than RO-MAN in this regard, but all outlets neglected to mention this aspect in at least 40% of the analyzed papers.

As indicated by the responses to the emails sent to corresponding authors of some HRI publications, both board approval and informed consent are probably used more

frequently than reported. Still, proper ethical conduct deserves to be properly emphasized in the literature.

Surprisingly, only 4% of the papers from this literature study explicitly mentioned data and privacy. By data and privacy, we mean where the author addresses how the data are handled and how the privacy of the participants is kept. This issue is relevant to HRI research since some personal data are usually gathered in experimental studies, e.g., through video recordings. As other ethical principles, data and privacy policies are likely used more frequently than explicitly reported in the papers. Nevertheless, it is an important ethical concern that deserves more attention.

Deception and debriefing are well rounded and customary practices in fields like psychology, and similar strategies can be seen in empirical HRI research. When presenting robots in HRI studies, they may appear to be more intelligent than what they actually are, oftentimes with intentional deception, e.g. WoZ [15]. In our literature study, there were eight publications that used WoZ, which is a common method used in this field [15]. Other than WoZ, deception was explicitly mentioned in 14 publications. Interestingly, there were 14 publications in total where debriefing was explicitly mentioned, but only 9 coming from the publications with deception. Our interpretation is that debriefing also is used in studies not including deception to inform the participants on the nature, purpose, and conclusions of the empirical study. This is a positive practice that could perhaps be adopted more frequently. Moreover, we found some themes that could broach on deception. For example, from the 2018 HRI conference two papers used a robot with emulated emotions. Although these publications used a consent form, each robot's capabilities do not seem to be addressed explicitly. Another publication deceived participants into thinking the robot was making errors in a card game when in fact the robot's behavior was programmed. It could be argued that studies like these should include debriefing to make sure that participants do not leave the study with any misconceptions.

Based on the obtained findings in our literature study, debriefing might not always be considered before sending participants away after the data collection step is conducted. Though it might not always be needed, it could be a very useful tool, not only to fulfill ethical guidelines, but to gain a deeper understanding of participants involved in HRI studies. In addition to this, it is also possible that by not being truthful to participants and by not debriefing them, the view that researchers are not truthful could be more common which could cause undesirable effects on future studies.

Although many publications neglected to report some, or all, of the five ethical principles in their papers, there were several authors that pursued proper ethical conduct in their work. These are, of course, worth noting and could perhaps be seen as best practice when conducting empirical HRI research. For example, Oliveira et al. [16] explicitly mentioned informed consent ("After signing an informed consent, participants were asked to provide information regarding their sex and age"), data and privacy ("The anonymity and confidentiality of the individual data was guaranteed"), and debriefing ("At the end, participants were thanked for their collaboration, received a movie ticket for participating in the study, and were debriefed"). Rea and Young [17] explicitly mentioned ethical board approval ("Our university's research ethics board approved both studies"), informed consent ("Participants were first given a briefing of the experiment and signed an informed consent form"), and deception with debriefing ("after all three conditions,

the participants were debriefed about the deception in the obstacle course room with the robot – it was, in fact, always the exact same robot").

The publications presented in the above paragraph show some of the unique ethical issues the field of HRI is facing, and to our knowledge there are not yet any established guidelines on how this type of deception, involving (human-like) robots, should be treated. As WoZ studies are more explicit deception, these studies touch on uncharted territory and need to be addressed further in this field. Future studies could explore this topic further.

One possible reason for excluding ethical conduct from the published articles may be space limitation. As many technical conferences, both HRI and RO-MAN put a hard page limit on published papers. Some conferences, including HRI, already exclude references from this page limit. We suggest that acknowledgments of ethical conduct could be treated in a similar way. If references to board approval, informed consent, and other ethical conduct could be included outside the page limit, it might help shape a practice where ethical conduct is a standard piece included in HRI publications to a larger degree than today.

It is worth noting that underreporting of ethical principles is not a unique issue for HRI. A literature study by Schroter et al. [18] found that ethical approval and consent form is underreported in medical journals, with failure to mention these in 31% and 47% of the papers, respectively. The above authors emphasize the importance of being transparent in publications. We hope that our paper's contribution can be one step towards a similar discussion in the HRI field.

The issue of how to handle ethical issues, like deception, deserves a wider acknowledgment and discussion within the HRI field. One concrete example could be the four criteria put forward by Matthias [19] providing a framework for when and why misleading and deceiving robots are morally permissible. According to the author, deception when using robots in healthcare is only morally acceptable when (1) it is in the patient's best interest, (2) it results in increased autonomy for the patient (e.g. being able to make more choices and being able to control the robot), (3) it is transparent or suggestive that deception is occurring and that the patient can chose to stop the deception, and (4) no harm could come to the patient, directly or indirectly. The latter also means that if the patient relies on a specific service from the robot (e.g. reminders to take medication) it must also be informed of the actual capabilities and services of the robot to be informed of what they can expect from it.

The underreporting of ethical principles in HRI research may have several ethical implications at the societal level. On the one hand, uninformed participants' expectations of robots may result in misunderstandings of current robots' capabilities and functionality. On the other hand, HRI researchers often demonstrate their robots to the public with pre-scripted lines and behaviors, where the robot interacts "naturally" with a human, without explaining the robot's functionality and how the interaction was set up beforehand. Although this is outside the scope of this study, we want to raise the question—what ethical responsibilities do researchers have towards not only participants, but to the public when presenting robots? We hope that this work contributes to stimulate the conversation of proper ethical conduct in the interdisciplinary field of HRI, both *methodologically* and *ethically*.

Acknowledgements. Special thanks to Oskar MacGregor for his valuable insight on proper ethical conduct. We would also like to thank Erik Lagerstedt and Kajsa Nalin for their support and help on parts of the analysis of the literature study.

References

1. Dautenhahn, K.: Socially intelligent robots: dimensions of human-robot interaction. Philos. Trans. Roy. Soc. B Biol. Sci. **362**(1480), 679–704 (2007)
2. Lindblom, J., Andreasson, R.: Current challenges for UX evaluation of human-robot interaction. In: Schlick, C., Trzcieliński, S. (eds.) Advances in Ergonomics of Manufacturing: Managing the Enterprise of the Future, pp. 267–277. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-41697-7_24
3. Baxter, P., Kennedy, J., Senft, E., Lemaignan, S., Belpaeme, T.: From characterising three years of HRI to methodology and reporting recommendations. In: The Eleventh ACM/IEEE International Conference on Human Robot Interaction, pp. 391–398. IEEE Press (2016)
4. Milgram, S.: Behavioral study of obedience. Psychol. Sci. Public Interest **67**(4), 371–378 (1963)
5. Haney, C., Banks, C., Zimbardo, P.: Study of prisoners and guards in a simulated prison. Naval Res. Rev. **26**(9), 1–17 (1973)
6. Weiten, W.: Psychology: Themes and Variations. Cengage Learning (2007)
7. American Psychological Association: Ethical Principles of Psychologists and Code of Conduct. American Psychological Association. https://www.apa.org/ethics/code
8. European Commission: Ethics in Social Science and Humanities. European Commission. https://ec.europa.eu/info/sites/info/files/6._h2020_ethics-soc-science-humanities_en.pdf
9. World Medical Association: Declaration of Helsinki. World Medical Association. https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/
10. European Union: General Data Protection Regulation. EU. https://gdpr-info.eu
11. ACM/IEEE International Conference on Human-Robot Interaction. https://dl.acm.org/conference/hri
12. IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). https://www.ieee-ras.org/conferences-workshops/financially-co-sponsored/ro-man
13. ACM Transactions on Human-Robot Interaction. https://dl.acm.org/journal/thri
14. Oates, B.J.: Researching Information Systems and Computing. Sage (2005)
15. Riek, L., Howard, D.: A code of ethics for the human-robot interaction profession. In: Proceedings of We Robot (2014)
16. Oliveira, R., et al.: Friends or foes? Socioemotional support and gaze behaviors in mixed groups of humans and robots. In: Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, pp. 279–288. ACM (2018)
17. Rea, D.J., Young, J.E.: It's all in your head: using priming to shape an operator's perceptions and behavior during teleoperation. In: Proceedings of the 2018 ACM/IEEE International Conference on Human Robot Interaction, pp. 32–40. ACM (2018)
18. Schroter, S., Plowman, R., Hutchings, A., Gonzalez. A.: Reporting ethics committee approval and patient consent by study design in five general medical journals. J. Med. Ethics **32**(12), 718–723 (2006)
19. Matthias, A.: Robot lies in health care: when is deception morally permissible? Kennedy Inst. Ethics J. **25**(2), 169–162 (2015)

PAPER II

EXPECTATIONS IN HUMAN-ROBOT INTERACTION

Reprinted from Julia Rosén (2021). “Expectations in Human-Robot Interaction”. In: *Advances in Neuroergonomics and Cognitive Engineering*. Ed. by Hasan Ayaz, Umer Asgher, and Lucas Paletta. Springer International Publishing, pp. 98–105 with permission from Springer International.

Expectations in Human-Robot Interaction

Julia Rosén(✉)

Interaction Lab, University of Skövde, Högskolevägen 1, 541 28 Skövde, Sweden
julia.rosen@his.se

Abstract. It is acknowledged that humans expect social robots to interact in a similar way as in human-human interaction. To create successful interactions between humans and social robots, it is envisioned that the social robot should be viewed as an interaction partner rather than an inanimate thing. This implies that the robot should act autonomously, being able to 'perceive' and 'anticipate' the human's actions as well as its own actions 'here and now'. Two crucial aspects that affect the quality of social human-robot interaction is the social robot's physical embodiment and its performed behaviors. In any interaction, before, during or after, there are certain expectations of what the social robot is capable of. The role of expectations is a key research topic in the field of Human-Robot Interaction (HRI); if a social robot does not meet the expectations during interaction, the human (user) may shift from viewing the robot as an interaction partner to an inanimate thing. The aim of this work is to unravel the role and relevance of humans' expectations of social robots and why it is important area of study in HRI research. Moreover, I argue that the field of HRI can greatly benefit from incorporating approaches and methods from the field of User Experience (UX) in its efforts to gain a deeper understanding of human users' expectations of social robots and making sure that the matching of these expectations and reality is better aligned.

Keywords: User experience · Human-robot interaction · Expectations

1 Introduction

The field of Human-Robot Interaction (HRI) has emerged, in part, to gain a deeper understanding of how humans interact 'naturally' with social robots [1]. Social robots are physical artifacts that are created for the purpose of behaving as an interaction partner [2]. Social robots should be able to (1) interact with its user, (2) serve different functions, and (3) possess social skills [5]. Its embodiment in the world, as well as its engagement with the world, are key components when defining socially interactive robots [3]. Social robots can therefore be created for the sole purpose of being companions, partners, and assistants to humans, while being implemented in applications areas such as hospitals, health care, education, and entertainment. Subsequently, there is a need to investigate and analyze what kinds of expectations humans have of social robots and how these expectations influence the interaction quality.

It is not an easy task to answer why some robots trigger social responses from humans, but there has been great progress to deepen this understanding. For example,

H. Ayaz et al. (Eds.): AHFE 2021, LNNS 259, pp. 98–105, 2021.
https://doi.org/10.1007/978-3-030-80285-1_12

the study of anthropomorphism addresses how a robot with human physical features, e.g., eyes and mouth, allows for expectations of behavior similar to a human being [2]. In fact, anthropomorphism is specifically used as a mechanism to evoke expectations of social competence. Expectations, thus, play a crucial role in HRI as they set the tone for the interaction quality. Humans tend to attribute mind, or agency, in other humans, animals, and artifacts [4, 5]. It has been shown that robots do not need to have the same level of intelligence as humans to be perceived as intelligent [2, 6]. The field of HRI aims to understand and develop social robots that take advantage of this tendency. It is therefore important to understand what the main goal of the robot is. If this is not considered, skewed expectations might occur which can lead to seeing the robot as an inanimate object.

While the general goal for HRI designers is to create robots that interact with humans, the toolbox and methodologies to reach this goal differ [1]. The User Experience (UX) field aims to analyze, design, evaluate, and implement artifacts with the user's experience in mind [7]. From a UX perspective, a social robot is rather seen as a tool to achieve a certain goal in a certain context of use. The goal, therefore, defines what the role of the social robot should be and what kind of tasks it should carry out to achieve that goal based on its end-users and the particular usage context. On the one hand, if the aim of a robot is to serve as a companion for older adults with the goal for these users to experience being less alone; the robot should, in the best of worlds, fulfill some identified social needs and would be expected to exhibit social behaviors. On the other hand, if the aim of a robot is to vacuum floors in a home with the goal for the users to experience less stress over cleaning, it would not be expected to behave socially to the same extent as a robot that is supposed to aid in feeling less alone.

In this work, I will discuss why humans' *expectations* of social robots play a crucial role in the user experience of social robots. Applying UX methodology [7] provides a viable approach to systematically develop and evaluate/assess/study expectations in order to create social robots with a positive UX. This is of importance as expectations of social robots may function as a confounding variable that threatens the internal validity of any HRI research.

2 Expectations

Expectations can be defined as believed probabilities of future events that sets the stage for the human belief system, which guides our behavior, hopes, and intentions [8, 9]. Humans can vividly conjure images of possible outcomes, even if the situation has not yet occurred, allowing regulation of behavior [9]. In real-time interactions with other beings, expectations serve to orchestrate anticipations of possible actions. As expectations are predictions of the future, expectations are often aligned with wishful thinking, and consequently can result in disappointment [8]. When an expectation turns out to be correct, it is confirmed, whereas when an expectation turns out to be false, it is disconfirmed. A constant pattern matching is unfolding between previous outcome, expected outcome, and actual outcome; sometimes called fluency processing [8]. This process of expectations analysis is mostly carried out implicitly and happens swiftly with low cognitive effort. It is acknowledged that once expectations are set on a specific outcome,

they are difficult to disrupt [9]. Therefore, it is important to understand how expectations affect UX before, during, and after interaction with social robots. Expectations are studied in the field of UX as it is an important aspect of user experience [10]. If an interaction with an artifact does not meet expectations, a user can experience negative UX which affects the user's emotions as well as the acceptance of the artifact [7].

2.1 Expectations of Social Robots

Social robots have not yet fully emerged in society and most people lack first-hand experience and technical knowledge of these artifacts. Most humans rely on expectations based on what they have seen in the media and it is therefore hard to foresee what kind of expectations humans have of social robots [11]. An obvious source of exposure of robots in media is science fiction. Movies like 'Ex Machina' and shows like 'Westworld' portray robots as super humans with a very high degree of artificial intelligence. Although these examples are obviously not real, they still constitute a significant basis for people's exposure to social robots.

Perhaps more noteworthy is what kinds of expectations exist for real social robots. For example, a study by Billing et al. [12], found that assistant nurses had unrealistic expectations of what social robots will be able to perform in welfare, both now and within 10 years. Moreover, part of the FP7 project DREAM carried out by my research group, the effects of robot-enhanced therapy for children with autism spectrum disorders was studied [13]. This study was made with a clinical protocol where written consent was obtained from caregivers to the participating children. Still, both children and their parents arrived with certain expectations, possibly very different from the experience of interacting with a real robot. These expectations do affect the study itself, especially in situations where we are interested in users' attitudes towards robots. After reading about the DREAM project in a newspaper, a parent to a disabled child, not involved in the study, contacted us with the hope of being able to purchase such a robot for her child. While the engagement of the public is of course very positive, the parent had in this case clearly formed unrealistic expectations of the robot's cognitive and social abilities, assuming that such a robot is able to function as a social companion on an everyday basis.

Designing robots with social cues, e.g., eyes, mouth, and nose, is in its nature deceptive as the robots are not able to utilize these features like humans while users will infer that this is the case [11]. Deception is also a common tactic in HRI studies as it allows for insight into human behavior. Wizard-of-Oz (WoZ) is a popular technique where the robot is remotely controlled by a human, making it appear as if the robot is able to behave in certain ways that is not possible today [14]. If participants are not made aware of this deception, expectations of future interactions with robots might be affected by these deceptive experiences.

Moreover, it is not uncommon to advertise social robots as companions with human feelings, thoughts, and empathy, showing robots that do not exist yet. Artifacts are often sold with unrealistic advertisements, like a car that transforms into a robot that dances; however, although commercials for social robots and cars have many similarities from a marketing point of view, I stress an important difference in that cars are commonly known to the customers and the expectations are based on previous exposure to them.

That is, it is apparent for customers' what part of the car is real and what is not. The same line of argument does not hold for social robots; they are not yet a part of our daily lives and the expectations are not built on first-hand experience. Therefore, negative UX can occur in HRI which may result in the robot not being used altogether [11].

3 Studying Expectations with UX Methodology

In recent years, there has been an increased interest in incorporating UX methodology in HRI research [15–17]. People have higher expectations on social robots than most other artifacts studied in UX [15, 18]. Hassenzahl and Tractinsky [10] stress three main factors that make up UX; (1) the internal states of the user (2) the designed systems characteristics, such as its purpose, complexity, and usability; and (3) the context/environment for the interaction. Being aware of these factors allows designing for a positive UX. These main factors can be applied in HRI research; for example, user's expectations (internal state) of a social robot (designed system characteristic) in an assisted living facility (context).

The UX design lifecycle, or UX wheel, is a model of core activities in UX [7]. The wheel consists of four iterative steps: *Analyze, Design, Prototype,* and *Evaluate*. The main purpose of the UX wheel is to ensure that the goal of any artifact in its context is fulfilled. The UX wheel provides support to systematically study how user expectations have an impact on the experience of social robots – before, during, and after the interaction. Below, each step of the UX wheel is presented along with an example of how it can be applied in HRI, drawing inspiration from the example mentioned in the previous paragraph.

3.1 Analyze: Understanding User Work and Needs

This step refers to when field data is elicited and analyzed through interviews and observations by studying users' work practices or daily habits [7]. The overall aim is to identify and formulate the needs of the end-users and gain an understanding of the bigger picture. This step is beneficial for HRI as it can help us understand the context in which the robot will operate and what expectations users will have of the interaction. If, for example, a robot is designed to be a companion at an assisted living facility, the users are the older adults living in that home. First steps would include understanding what aspects of companionship the robot should fulfill, including users' expectations of the robot, and the context it will operate in. If the goal is to have an effect on the user's well-being, we can start by studying what the user's might need in their everyday life and what they expect from the interaction. The user might feel lonely and forget to take medication. Perhaps the robot should be able to interact with as many user as possible.

3.2 Design: Creating Design Concepts

This step refers to when the gathered information is realized in UX goals and conceptual design concepts to actualize the user's needs and expectations [7]. When designing for emotional needs, the designer aims to make the design meaningful. If the analysis of

the gathered data reveals that the users are feeling lonely, robot features could include face-tracking, communication abilities, and having a fuzzy exterior. All of which could possibly fulfill the need to feeling less lonely. From this, it could be deduced that the users would benefit from interacting with a social robot (viewed as an interaction partner) to fulfill these needs. Therefore, the social robot should be designed with this in mind. If the social robot is expected to go between rooms to talk to several users, having wheels could be beneficial in order to reach the different users. If it is revealed that the users expect that the robot will be able to hug, features could be implemented in order to meet these expectations. If, however, it is not feasible to implement such changes, changing the design to lower expectations could be an option. Perhaps a robot with immobile arms, showing clearly it cannot lift its arms for a hug, or designing the social robot as a zoomorphic animal, thus lowering the expectation of communicative skills but still being able to fulfill other social needs, such as having a soft fur to pet.

3.3 Prototype: Realizing Design Alternatives

This step refers to when different kinds of prototypes are created from conceptual designs in order to evaluate an artifact before committing to the final design [7]. This is beneficial as changing the artifact after it's created can be costly and time consuming. Prototypes can be useful in HRI as it can help manage expectations by investigating what features fulfill the user needs. For example, if we want the social robot to appear talkative as a way to invite interaction, different low-fi prototype options could be presented in order to assure aligned expectations and reality. This could include comparing an anthropomorphic robot and a zoomorphic robot to have the possibility to investigate which design option is the most suitable for the identified needs and formulated UX goals. However, it can be hard to create functional hi-fi prototypes as social robots are usually complex and need to be completed before it can be properly used [19]. There are ways to go around this; for example, a WoZ set-up allows for testing some aspects of the robot while it is still not implemented.

If we go back to the social robot in an assisted living facility, we could study what kind of expectations the user has before, during, and after the interaction by having the user interact with a prototype. The users could expect that the social robot would respond to being touched, and experienced disappointment when this did not occur. More prototype testing could therefore include the robot, still being remotely controlled by the designer, giving auditory feedback to being touched without having to actually spend time programming this feature. If it has a positive outcome, implementing this feature would be deemed worth the resources.

3.4 Evaluate: Verifying and Refining Designs

This step refers to when the work is evaluated with various methods to assess how well the artifact fulfills its goal. The goal is to identify and improve UX problems (e.g. design flaws). There are two major approaches to UX evaluation; formative and summative evaluations [9]. Formative evaluation occurs during the development process and a summative evaluation occurs on the final social robot. The purpose of formative evaluation is to receive feedback on design ideas in the earlier steps in the UX wheel,

e.g., via sketches of interaction flow or physical mock-ups of the envisioned robot. Summative evaluation is used to measure the UX of a high-fidelity prototype or the final artifact. It could also be used to gain an understanding of its usage in practice, i.e., in an ecologically valid environment [16].

This step evaluates how well the social robot is fulfilling its goals and to what extent user expectations are met. If there are still severe UX problems left and major expectations are not met, the UX designer will perform another iteration in the UX wheel and continue until UX goals and expectations are met. By a sequence of formative and summative evaluations it can be determined what needs to be refined with the social robot in the assisted living facility. For example, if the social robot does not meet the expectations of being socially competent despite previous steps, the designer can go back and refine the issue by updating certain features like designing the robot to ask the user to repeat themselves when the question or command is not understood by the robot. By summative evaluation, it is determined whether all the UX goals are fulfilled and the social robot is meeting the expectations of being a companion and has improved the well-being of the users living in the assisted living facility.

4 Conclusions

Expectations could be a severe confounding variable that ought to be considered in any HRI research as exposure to social robots are rare and assumptions are made mostly from media; ultimately threatening the internal validity [11]. If disconfirmed expectations of social robots cause negative UX, there is a need to discuss how to prevent it.

Expectations can be combated by, for example, revealing the actual capabilities of the social robot to people, e.g., showing the mechanical parts that make up the robot in order to grasp its lifeless nature [11]. A study by Sun and Sundar [20] found that participants had a more positive experience with robots when they assembled the robot themselves, demonstrating that a good interaction can still occur even if they participants knows what is ‘under the hood’.

What I argue for in this paper is to include UX methodology when designing and evaluating social robots. Being able to assess expectations at an early stage, following the UX design lifecycle, is crucial if we want to study and evaluate how users are experiencing social robots. However, employing the UX design lifecycle is not always feasible since some of the social robots that are being used in HRI today are often bought from robotics companies and are rarely self-built in the lab. There is therefore very little a HRI researcher can do in terms of designing them from scratch, as the UX design lifecycle promotes. Bringing in UX after the fact has traditionally little use after the design process [7].

Moreover, HRI research comprise also many technical topics of social robots, far from a specific UX design. Still, I believe that incorporating some aspects from UX design could be useful in any stage of HRI, to help researchers gain a deeper understanding of how well a robot suits a particular context of use. Although many robots are bought ready-to use, it is still possible to design some robot behaviors, such as social behaviors that are suitable for a specific context. For a social robot situated in an assisted living facility, adding features like asking if the user has taken their medicine, or having conversations,

can still be added to most social robots. Similar efforts have been made by Tonkin et al. [17], proposing a UX methodology in robotic applications for commercially available social robots.

In addition, as people seem to have solid views of robots and expectations are hard to disrupt, it might be a challenge to adjust/change these expectations. It is therefore of major importance to study whether and to what extent expectations can be changed over time – before, during, and after interaction. According to Manzi et al. [3], expectations of the mechanical properties of a robot are higher before an interaction with a robot. Edwards et al. [20], showed that, not only do participants have expectations of interactions, but that this can also change after exposure to robots. After the interaction, participants demonstrated less uncertainty and greater social presence.

From an ethical point of view, though there are many kinds of social robots, there are even more kinds of humans. Not all humans fit the'norm' and some might benefit from social robots because of it. This is especially important in health care as there will be patients with specific needs that fall outside the norm. A UX designer has a responsibility to make sure that the design of a social robot is ethically justified which might sometimes go against other needs such as stakeholders [22].

Indeed, UX methodology can be beneficial to predict and prevent disconfirmed expectations in robots; whether if it's to view them as a thing or as an interaction partner. In the field of HRI, it can be assumed that the underlying goal of an interaction is for the human to expect to interact with an interaction partner. From the vast HRI literature, it is evident that this is indeed possible. Nonetheless, each interaction is unique and some social robots are able to trigger a higher level of expectations than others. According to Alač [6], some components of social robots evoke life-like expectations, such as mirroring head movements, whereas other components don't, such as a touch screen designs on the robot. Users can therefore expect and view a social robot as an interaction partner, while handling the robot as a thing. As Alač eloquently put it, "The robot is a special thing because of its agential characteristics" [6, p. 533]. There are many dimensions to expectations and more research is needed in order to gain a deeper understanding of how it affects HRI research. UX methodology, including the UX design lifecycle, could be a useful tool to reach this goal.

Acknowledgments. I would like to express my gratitude to Jessica Lindblom and Erik Billing for their valuable input and guidance for this project.

References

1. Dautenhahn, K.: Methodology & themes of human-robot interaction. Int. J. Adv. Rob. Syst. **4**(1), 103–108 (2007)
2. Duffy, B.R.: Anthropomorphism and the social robot. Robot. Auton. Syst. **42**(3–4), 177–190 (2003)
3. Manzi, F., Massaro, D., Di Lernia, D., Maggioni, M., Riva, G., Marchetti, A.: Robots are not all the same. Cyberpsychol. Behav. Soc. Network. **24**, 1–8 (2020)
4. Dennett, D.: The Intentional Stance. MIT press, Cambridge (1989)
5. Heider, F., Simmel, M.: An experimental study of apparent behavior. Am. J. Psychol. **57**(2), 243–259 (1944)

6. Alač, M.: Social robots: things or agents? AI Soc. **31**(4), 519–535 (2016)
7. Hartson, R., Pyla, P.S.: The UX Book. Morgan Kaufmann, San Diego (2018)
8. Borg Jr., M., Porter, L.: Following the life-course of an expectation. In: Leon, P., Tamez, N. (eds.) The Psychology of Expectations, pp. 1–47. Nova Science Publishers (2010)
9. Roese, N.J., Sherman, J.W.: Expectancy. In: Kruglanski, A.W., Higgins, E.T. (eds.) Social Psychology: Handbook of Basic Principles, pp. 91–115 (2007)
10. Hassenzahl, M., Tractinsky, N.: User experience — A research agenda. Behav. Inf. Technol. **25**(2), 91–97 (2006)
11. Malle, B.F., Fischer, K., Young, J.E., Moon, A., Colling, E.: Trust and the Discrepancy Between Expectations and Actual Capabilities of Social Robots. Cambridge Scholars Publishing (2020)
12. Billing, E., Rosén, J., Lindblom, J.: Expectations of robot technology in welfare. In: The Second Workshop on Social Robots in Therapy and Care in Conjunction With the 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2019), Daegu, Korea (2019)
13. Billing, E., et al.: The DREAM dataset. PLOS ONE **15**(8), 1–15 (2020)
14. Riek, L.D.: Wizard of Oz studies in HRI. J. Hum.-Robot Interact. **1**(1), 119–136 (2012)
15. Huang, W.: When HCI Meets HRI. Virginia Polytechnic Institute and State University (2015)
16. Lindblom, J., Alenljung, B., Billing, E.: Evaluating the user experience of human robot interaction. In: Jost, C., et al. (eds.) Evaluation Methods Standardization for Human-Robot Interaction, pp. 231–256 (2020)
17. Tonkin, M., Vitale, J., Herse, S., Williams, M.A., Judge, W., Wang, X.: Design methodology for the UX of HRI. In: Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, pp. 407–415 (2018)
18. Wallström, J., Lindblom, J.: UX Evaluation of social robots with the USUS goals framework. In: Jost, C., et al. (eds.) Evaluation Methods Standardization for Human-Robot Interaction, pp. 177–201. Springer (2020)
19. Weiss, A., Bernhaupt, R., Schwaiger, D., Altmaninger, M., Buchner, R., Tscheligi, M.: User experience evaluation with a wizard of Oz approach. In: 2009 9th IEEE RAS International Conference on Humanoid Robots, pp. 303–308. IEEE (2009)
20. Sun, Y., Sundar, S.S.: Psychological importance of human agency how self-assembly affects user experience of robots. In: 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 189–196. IEEE (2016)
21. Edwards, A., Edwards, C., Westerman, D., Spence, P.R.: Initial expectations, interactions, and beyond with social robots. Comput. Hum. Behav. **90**, 308–314 (2019)
22. Bardzell, S., Bardzell, J.: Towards a feminist HCI methodology. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 675–684 (2011)

PAPER III

THE SOCIAL ROBOT EXPECTATION GAP EVALUATION FRAMEWORK

Reprinted from Julia Rosén, Jessica Lindblom, and Erik Billing (2022). “The Social Robot Expectation Gap Evaluation Framework”. In: *Human-Computer Interaction. Technological Innovation*. Ed. by Masaaki Kurosu. Springer International Publishing, pp. 590–610 with permission from Springer International.

The Social Robot Expectation Gap Evaluation Framework

Julia Rosén[1(✉)], Jessica Lindblom[1,2], and Erik Billing[1]

[1] Interaction Lab, University of Skövde, Box 408, 541 28 Skövde, Sweden
julia.rosen@his.se
[2] Department of Information Technology, Uppsala University, Box 337, 751 05 Uppsala, Sweden

Abstract. Social robots are designed in manners that encourage users to interact and communicate with them in socially appropriate ways, which implies that these robots should copy many social human behaviors to succeed in social settings. However, this approach has implications for what humans subsequently expect from these robots. There is a mismatch between expected capabilities and actual capabilities of social robots. Expectations of social robots are thus of high relevance for the field of Human-Robot Interaction (HRI). While there is recent interest of expectations in the HRI field there is no widely adapted or well formulated evaluation framework that offers a deeper understanding of how these expectations affect the success of the interaction. With basis in social psychology, user experience, and HRI, we have developed an evaluation framework for studying users' expectations of social robots. We have identified three main factors of expectations for assessing HRI: affect, cognitive processing, and behavior and performance. In our framework, we propose several data collection techniques and specific metrics for assessing these factors. The framework and its procedure enables analysis of the collected data via triangulation to identify problems and insights, which can grant us a richer understanding of the complex facets of expectations, including if the expectations were confirmed or disconfirmed in the interaction. Ultimately, by gaining a richer understanding of how expectations affect HRI, we can narrow the social robot expectation gap and create more successful interactions between humans and social robots in society.

Keywords: Human-robot interaction · Social robots · Evaluation framework · User experience · Expectations

1 Introduction

The ongoing increase of new interactive technologies in our daily lives results in growing expectations of these technologies from their intended users [15]. One such interactive technology is social robots, and their degree of participation in everyday activities in society are becoming more sophisticated [2,5,20]. Social

M. Kurosu (Ed.): HCII 2022, LNCS 13303, pp. 590–610, 2022.
https://doi.org/10.1007/978-3-031-05409-9_43

robots have several characteristics that separate them from other kinds of more traditional technologies; they occupy space, act rather autonomously, and users have to respond to them "here and now." In addition, they are designed in a manner that encourage users to interact and communicate with them in socially appropriate ways. Given that social robots involve the intersection of a digital artifact and of agency as found in living creatures, social robots are often either positively experienced as fascinating and interesting or negatively experienced as frightening and scary [1]. Hence, social robots tend to blur the distinction between a thing and an agent, which results in several challenges for users regarding what to expect when interacting with social robots. To complicate the issue further, the majority of humans have either no first-hand experience or limited experience of interacting with social robots. Humans' main exposure to social robots are predominantly from movies and the media, which may result in misleading and inaccurate expectations of social robots [1,6,11,31].

We argue, it is of utmost importance to gain a deeper understanding of the roles expectations play when users are interacting with robots as well as the impact of expectations on user experience (UX) before, during, and after the interaction. Some recent but limited research has studied humans' expectations of social robots [6,11,21,23], but research conducted in the intersection of the UX, social robotics, and human-robot interaction (HRI) fields seems to be limited. Although there is an interest on this topic, there is no widely adapted or well formulated evaluation framework. Expectations are especially important to understand when they are disconfirmed as this can guide and inform the future design of social robots, which will ultimately lead to more successful interactions. We propose to narrow these knowledge gaps within HRI with inspiration from social psychology and UX. UX goes beyond usability and pragmatic factors to include hedonic factors, in which emotional and experiential aspects are emphasized enabling evaluation of users' expectations of social robots in a systematic way [8,15]. In the present work, we provide an evaluation framework that considers the role of expectations when interacting with social robots. We will both describe key insights from the development of the evaluation framework and the proposed version of the evaluation framework, including the procedure for applying it when performing a minor empirical UX evaluation when interacting first-hand with a social robot [8,20].

2 Background

2.1 Expectations of Social Robots: Related Work

The concept *expectation* is defined as believed probabilities of the future that sets the stage for the human belief system which guides our behavior, hopes, and intentions [29,30]. As expectations relate to predictions of the future, expectations are often aligned with wishful thinking, and can thus result in disappointment and negative affect [29,30]. Confirmed expectations occur when expectations and actual outcome is aligned; in contrast, disconfirmed expectations are when expectations and actual outcome is not aligned. Our work is inspired by

the research done on *expectancy* in social psychology [29,30]; however, we will use the concept *expectation* which we view as similar when used in this context as expectations is another state of expectancy. Expectations play a crucial role in social interactions. In real-time interactions with other beings, expectations serve to orchestrate and predict possible actions. Expectations allow humans to co-exist and handle the complexity of the social world, and provide a way to decrease how complex the world is through stereotyping and norms. Stereotypes are expectations of groups or individuals that are exaggerated, biased, and overgeneralized [29]. Social robots are often stereotyped, commonly based on science fiction and limited actual interaction with social robots [3,34]. Not only do users who are intended to interact with social robots have expectations of social robots, but also the designers who create these robots have expectations while designing for these anticipated interactions.

Lohse [21] explicitly addressed the role of expectations in HRI, offering a starting point when introducing some assumptions on user expectations which needed to be studied and explored more systematically considering the influence of expectations on HRI research. She claimed that users' expectations can be inferred from data collected in actual HRI situations and analysis of the collected data can be advantageous by considering users' expectations as well as their views on the interaction situation. She pointed out that knowledge about users' expectations can support the design and development process of robots because expectations seem to be dependent on the robot's actual behavior, and as a result could be formed to improve the quality of the HRI. She also suggested that future work should not only consider research on users' expectations, but also to study robot's expectations, looking at both sides of the HRI coin. Finally, she questioned how expectations could be accurately measured. Since this publication, there has, to our knowledge, not been any attempts to develop a well formulated evaluation framework for expectations.

There has also been recent but limited research considering expectations of social robots. Manzi et al. [23] examined how the physical appearance and behaviors performed by the social robots affected the quality of interaction with humans. It was revealed that the participants' expectations were influenced by the interaction per se, and were surprisingly independent of the particular kind of social robot. Edwards et al. [6] examined how initial expectations and impressions may be modified and confirmed through short first-hand experience of communicating with a social robot based on HHI models of social interaction. Expectations were assessed by altered levels of (i) uncertainty, (ii) social attraction, and (iii) social presence. The results revealed that many participants reported feelings of affinity and connectedness, whereas a nearly identical encounter with a human experimenter resulted in opposite outcomes. The participants' modified expectations toward the robot may result in a so-called robot conformation, i.e., the human tendency to magnify the robot's confirmation responses and its limited ability to offer behavioral feedback contrasted to how humans act. Jokinen and Wilcock [14] studied expectations of social robots using the Expectations and Experience (EE) method. The EE model was used to investigate and analyze

the quality of interaction, i.e., what the users experience in regard to what they desire or expect. The results confirm that expectations in general were rated higher than the actual experience, showing that a majority of the participants perceived a positive experience, and indicating that the participants perceived the interaction with the robot enjoyable and interesting. However, there are indications of a negative relationship between expectations of the robot's behavior and the extent to which the participants perceived that they were 'understood' by the robot. Interestingly, the authors note that the most experienced participants seemed to be the most critical ones. The authors concluded that insights from user evaluations of social robots should not be limited to increasing positive UX but could also be used to understand how to reduce the difference between the users' expectations and actual experiences. They argued that reducing expectations and experience mismatch is of major importance to cultivate long-term relations such as trust between social robots and their end-users.

Horstmann and Krämer [11] investigated which kinds of expectations humans have concerning social robots as well as the bases for forming these expectations. Their results indicate that humans' experiences of media regarding social robots lead to increased expectations of robots' abilities and capabilities, in turn successively enhancing humans' expectations of social robots. Moreover, humans' awareness and acquired knowledge of negatively perceived fictional social robots enlarges negative expectations of robots being threats to humans. Conversely, those humans who have more non-fiction knowledge about the capacities and limitations of robot technology show reduced levels of anxiety towards social robots. They pointed out that their findings are mostly based on subjective ratings and no first-hand encounters with social robots, which mainly revealed what their participants hypothesize what they should expect. They suggested that future work should examine what kinds of expectations and preconceptions humans hold towards robots, and in what ways these influence their behavior when interacting first-hand with a social robot. Later, Horstmann and Krämer [13] examined participants' expectations versus their actual behavior when interacting with the a social robot. The authors concluded that in general, during first-hand interaction with a social robot, the robot's perceived behavior is more influential for participants' evaluations of it than their formulated expectations. Hence, the main insight from their study, which also confirms prior research, indicates that the robot's behavior during the actual interaction is the key variable influencing how the participants evaluate the actual robot as well as the interaction quality with it. Moreover, Horstmann and Krämer [12] conducted an experiment aiming to examine how a negative expectancy violation caused by a social robot and its reward valence could demonstrate how desirable it is for the participants to interact with the social robot, and how this affected the evaluation of HRI quality after the interacting. The results indicate that when the robot negatively violated expectations, participants evaluated the robot competence, sociability, and interaction skills more negatively.

Moore [25] introduced the so-called "habitability gap", i.e., the perceived mismatch between the capabilities and expectations addressed by users and the actual features and intended benefits provided by social robots. Moore assumed that given the recent and rapid development of interactive technology, it should be expected that the capabilities of such digital artifacts would gradually develop, but there is a habitability gap in which UX drops when the robots' flexibility is enhanced. Schramm et al. [35] presented an initial conceptual framework for expectations of social robots, consisting of two aspects. The first aspect is the design (i.e., appearance and behavior) of the robot that "emits capability signals" [35, p. 439], which is divided into the life-likeness by the robot, the functional design of the robot (e.g., cameras means the robot is able to see), and how the robot is introduced. The second aspect is the mental model constructs of the robot held by the user, based on the design of robot, which are formed by the observation of the robot's mechanical (e.g., physical ability) and life-like capability (e.g., emotional system). The authors stressed how a simple robot behavior can lead to complex constructions of what to expect of the robot, thus creating a gap between expected and actual capabilities. Although the authors mentioned that they have started to develop measurement tools to evaluate and create a deeper understanding of expectations based on robot design, they did not present any measurements or evaluation framework. Hence, our framework complements the work done by the authors.

In summary, three main conclusions for future research can be drawn. First, there is a need to study users' expectations in actual HRI beyond the mere encounter with a social robot and to systematically study how users' expectations may be altered if their initial expectations are not confirmed (e.g., [6,11]). Second, the close relationship between expectations and UX provides a well-aligned approach to include the various time spans in UX. Third, there is a need to develop an evaluation framework that encompasses several main factors that systematically could be evaluated in actual HRI to better cope with mismatches of expected capabilities and actual capabilities of the robot.

2.2 Users' Experience and Expectations from a HRI Perspective

A relevant HRI angle is the intersection with the study of UX, which allows for a user-centered perspective, including studying users' expectations. Although this intersection of UX within HRI exists, and is getting more traction in the literature [2,17,31,36], it is often overlooked [2]. There are also overlaps between users' experience and users' expectations when interacting with technology [33]. User expectations could indicate the anticipated behavior, direct the focus of attention, serve as a source of reference for the actual UX and how it is interpreted, and subsequently has an impact on the users' overall perception of the technology [15]. As such, expectations can have a critical influence on the formation of the actual UX of the social robot. Research on expectations can reveal a deeper understanding of the central aspects of UX. One of the main components of UX, that is sometimes missing in HRI research, is the temporal aspect of an interaction [31].

A key aspect of UX occurs during the actual interaction with a system, but is not the only relevant UX aspect to consider [33]. Users are also affected by indirect experiences before their first encounter with an interactive system. Such indirect experiences are rooted in previous experiences and thoughts related to advertisements, presentations, and demonstrations of the system or related systems. Indirect experience may also include exposure to various media and movies, or other people's opinions. In the same way, users' indirect experiences may occur after the actual usage situation, such as when they are reflecting on previous usage and prior expectations, or through the impact from other users' assessment of the system which may retroactively alter the actual UX [33]. A key point in UX research is to study the temporal aspects of using a system which makes it highly relevant for expectations. Not only are expectations about future events, they are also dynamic and can change over time [29], making temporal aspects crucial in order to understand expectations. Roto et al. [33] identified four temporal aspects to consider in UX, including anticipated UX, momentary UX, episodic UX, and cumulative UX. These temporal aspects are dynamic and can vary depending on the situation. Anticipated UX refers to the period before an interaction, whether it is for the first time or repeated interaction. It is the imagined experience of future interaction. Momentary UX refers to the time span during the interaction of a system. Episodic UX refers to the appraisal of a particular episode after interaction with a system. Lastly, cumulative UX refers to opinions of a system as a whole, after interaction over time [33]. The temporal aspects are particularly relevant when it comes to expectations and subsequent interaction of a system, especially the interaction with social robots. The momentary use of social robots is typically the only temporal aspect studied in HRI research. Although not explicitly mentioned by Roto et al. [33], expectations are also important to consider, and can affect the interaction across all temporal aspects including before the interaction with the robot (anticipated use), during the interaction with the robot (momentary use), after the interaction with the robot (episodic use), and the interaction over time with the robot (cumulative use).

2.3 Olson's et al.'s Expectation Model

One of the psychological theories of expectations that Lohse [21] took a closer look at in relation to robots was the expectation process by Olson et al. [29]. We consider this model to be of major interest, and it serves as the foundation for our framework. However, we have made several adjustments to the original model. First, we modified its main focus on HHI to HRI. Second, we altered its social psychological perspective to a more user-centered one, applying concept commonly used within UX to better fit the purpose of our evaluation framework. In Fig. 1, we present this modified version of the expectation process model by Olson et al. [29, p. 231]. This model is adjusted to accommodate the interaction with social robots. There are also other models of expectations e.g., [24], though this model is, to our knowledge, one of the most comprehensive presentations of expectations, and is thus of major importance for our work.

As the model shows (Fig. 1), *direct experience, other*, and *beliefs* form expectations. Direct experience is the expectations built on first-hand experience. Although not as common, some have their expectations built on actual first-hand experience of robots. Direct experience is more common with robots already used in society at large, such as vacuum cleaner robots or lawn mower robots. Expectations built on direct experience are typically held with greater certainty and are a stronger predictor of future behavior [29]. *Other* refers to the expectations built on indirect experience, including from other people (e.g., personal connections and from media) or from exposure to social robots in different ways. In the original model, the "other" bubble refers to "other people", and is in the present version expanded to include not only other people ("humans"), but also social robots ("robots"). These two exist on a gradient scale as humans tend to interpret social robots in an anthropomorphic manner [1]. This gradient does not refer to how human-like a given social robots actually is, but rather how much (by design or accident) a given social robot is expected to behave like humans, i.e., the social robot may be perceived as a human-like creature, even if it is not "human-like" in some objective way. *Beliefs* are sources of expectations that can be inferred from other beliefs (e.g., "robots are intelligent in movies, thus the robot I am interacting with is likely intelligent too"). Together with direct experience and others, beliefs are built, thus are interrelated with each other.

Expectations can vary along four different dimensions: certainty, accessibility, explicitness, and importance [29]. *Certainty* refers to the subjective estimated probability of how likely it is that the outcome will occur. *Accessibility* refers to how easy it is to activate and use a certain expectation. *Explicitness* refers to what degree expectations are consciously generated. Some expectations are implicitly assumed, usually related to the degree of certainty (e.g., "the sun will rise in the morning"), whereas other expectations are consciously thought about (e.g., how an interaction with a social robot will be like). *Importance* refers to the expectation's significance, the higher the importance, the higher is the impact.

The rest of the model relates to the consequences of expectations [29]. These factors can be divided into three categories: affective, cognitive, and behavioral. *Affective* refers to emotional consequences, such as attitudes and anxiety. *Cognitive* refers to factors that has an effect on cognitive processes, such as interpretation and memory. *Behavioral* refers to consequences that causes choice of actions due to the expectations, such as forming intentions to act, hypothesis testing, and self-fulfilling prophesies. These factors occur when expectations are confirmed or disconfirmed [29]. Confirmed expectations often lead to positive affects, are often handled implicitly (and with ease), and results in expectations that are upheld with greater certainty. In UX terms, confirmed expectations of the interaction quality with social robots results in a positive UX. Confirmed expectations may also, on rare instances, produce secondary affect (positive or negative), given the inferences made after confirming the expectation. In contrast, disconfirmed expectations often lead to negative affects and are handled explicitly as they are surprising and need heavy cognitive processing in order to make sense of what went wrong. Again in terms of UX, disconfirmed expectations of interacting with social robots produces a negative UX. Sometimes there is a desire to retain (i.e.,

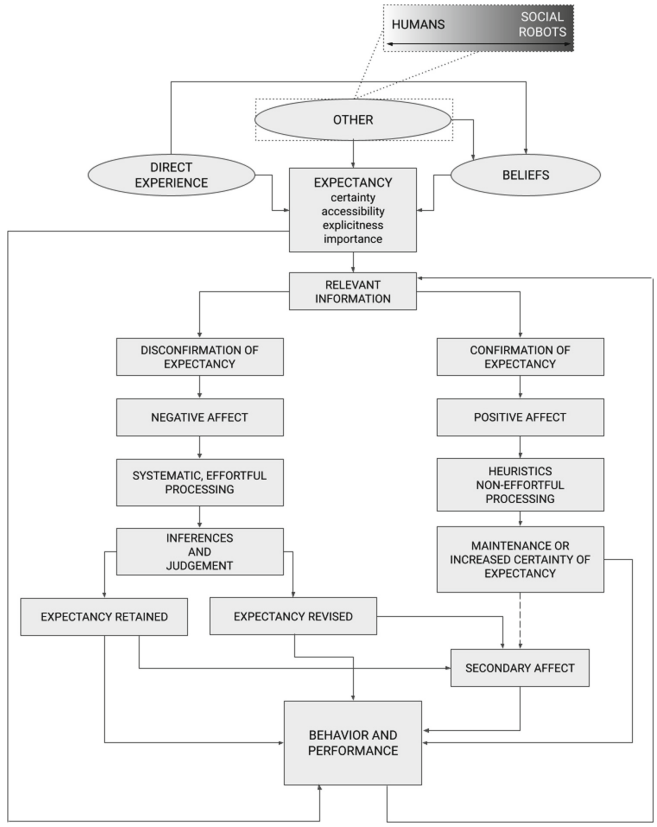

Fig. 1. A modified model of the expectation (expectancy) process by Olson et al. [29, p. 231]. The modified version is extended to include expectations of social robots.

uphold) the initial expectations and the event may be explained away. Other times there is a need to carefully consider what went wrong in order to revise the expectations for future instances. Future expectations of the instance are held with a higher level of uncertainty. Both confirmed and disconfirmed expectations may lead to altered behavior and performance [29].

When considering a social robot, a user may enter an interaction with high expectations of the robot's ability to interact socially. These beliefs may be built on the exposure to science fiction robots in the media. When the social robot is unable to uphold a complex interaction, the expectations are disconfirmed, resulting in negative affect, or negative UX, (e.g., disappointment) and effortful cognitive processing in order to rationalize what went wrong. This cognitive processing turns into inferences and judgment, and the expectations may either

be revised to match the interaction experience with the robot or retained due to the robust beliefs of social robots (i.e., the action will be explained away or ignored). Even if the expectations are revised, they will be held with less certainty [29]. Effortful cognitive processing of disconfirmed expectation can also introduce secondary affects relating to discomfort, dissonance, or frustration. Lastly, expectations will ultimately have an effect on behavior and performance. The user may lose interest and stop the interaction due to negative affective experiences, or may try to figure out how to make the interaction more successful. Hence, these affects contribute to a corresponding positive or negative UX.

In Fig. 2, we have illustrated how the mismatch between expected and actual capacities of a social robot can occur, which we call *the social robot expectation gap*. We stress that disconfirmed expectations can happen when expectations are both too high and too low. The social robot expectation gap (Fig. 2) is illustrated with its two spaces of disconfirmed low as well as disconfirmed high expectations. These two spaces of the social robot expectation gap can be viewed as the two different outcomes when the user's expectations and the actual capabilities of the robot are not aligned. The diagonal line is when expectations and robot capabilities are confirmed. On the one hand, if a human interacts with a social robot and expects, based on prior exposure to science fiction movies, that it is capable of feeling pain when falling but does not express any distress when such occurrences occur, the expectation will be disconfirmed in the form of high expectations (falling in the blue space of disconfirmed expectations). On the other hand, if the user does not expect that the robot is capable of any verbal communications and then it does strike up a conversation, the expectation will be disconfirmed in the form of low expectations (falling in the green space of disconfirmed expectations). A consequence of this presentation of expectations is that we can achieve high interaction quality with robots both with high and low capabilities, given that expectations are confirmed on the diagonal line. An interaction can go smooth given that expectations are confirmed, regardless of actual capabilities of the social robot. Moreover, expectations are not static but dynamic across an interaction [33]. Therefore, this social robot expectation gap is also dynamic and may change before, during, and after the interaction.

Disconfirmed high expectations have been proposed previously by Kwon et al. [18], called the "expectation gap"; however, the authors do not account for disconfirmed low expectations. The figure presented (Fig. 2) here includes both too high and too low expectations of social robots. As mentioned previously, Moore [25] presented the "habitability gap" which is the perceived mismatch between the capabilities and expectations, and although this gap relates to the social robot expectation gap, it focuses on different aspects (e.g., flexibility and usability, voice-based systems) of interacting with new technology. Besides the obvious central theme of expectations, the social robot expectation gap, the expectation gap [18], and the habitability gap [25] demonstrate how the interaction is affected by human expectations of artifacts, in particular social robots in the case of this work. In addition, when expectations are low, there will be less interaction and

thus the robot's capabilities will be less discovered, ultimately affecting the interaction quality. The social robot expectation gap serves, together with the model by Olson et al. [29], as the foundation for the design and development of the framework.

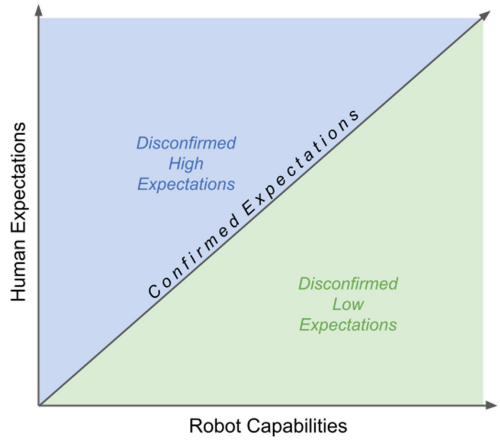

Fig. 2. The social robot expectation gap, with the two spaces of disconfirmed expectations that occurs when a robot's capabilities does not align with the expected capabilities.

3 Design and Development of the Framework

Drawing from the model by Olson et al. [29], we propose an evaluation framework for expectations in order to investigate the consequences of high and low expectations of social robots when a user interacts with a social robot, prior, during, and after the interaction. Based on the model (Fig. 1), we have identified three main factors in which expectations can be adequately evaluated: affect, cognitive processing, and behavior and performance. The data should be analyzed together in order to get a full picture of expectations; one metric alone may not be able to point at expectations, but together they are building blocks to extract information regarding expectations.

3.1 Affect

Affects are the factors that cause any sort of emotional reaction when expectations are either confirmed or disconfirmed. This is the first step after confirming or disconfirming expectations, following the Olson et al. model [29]. Attitudes is one such affect. Attitudes are defined as mental states that occur before the behavior, and are regarded in psychology research as one of the key elements for

expectations [26,29]. Moreover, attitudes can be viewed as a user's belief of an object and its characteristics in relation to the user's perception of those characteristics. Anxiety is another affect, and expectations can both increase and decrease levels of anxiety. Expectations in themselves cannot cause anxiety, but certain contents of expectations can cause anxiety. Anxiety can be elicited as an anticipation of an event, such as fear of failure.

For this factor of expectations, we propose to evaluate negative attitudes and anxiety towards robot by using the two questionnaires by Nomura et al. [26] that cover these topics. These questionnaires are popular in the HRI field, and have been used in various ways to measure attitudes and emotions towards robots. de Graaf and Allouch [7] found that, despite some flaws, these are valid techniques to evaluate users' emotions towards robots. In addition, we suggest observations of users' facial expressions to assess non-verbal emotions during the interaction.

The negative attitude toward robot scale (NARS) is a scale made for measuring people's attitudes toward robots in interaction and in daily-life [26]. NARS consists of 14 questionnaire items and are classified into three sub-scales. The first sub-scale consists of 6 items on the theme of "negative attitude toward situations of interaction with robots". The second sub-scale consists of 5 items on the theme of "negative attitude toward social influence of robots". The third sub-scale consists of 3 items on the theme of "negative attitude toward emotions in interaction with robots". The scale is on a 5-point Likert scale, and the individual score is calculated by summing the scores for each sub-scale.

The robot anxiety scale (RAS) measures the altered behavior participants may have towards robots based on their anxiety towards robots. Nomura et al. [26,27] argued that negative attitudes may not lead to different behaviors toward robots. Anxiety is explained as emotions (anxiety or fear) that inhibit interaction with robots. RAS consists of 11 questionnaire items and are classified into three sub-scales. The first sub-scale consists of three items on the theme of "anxiety toward communication capability of robots" The second sub-scale consists of four items of the theme of "anxiety toward behavioral characteristics of robots" The third sub-scale consists of four items on the theme of "anxiety toward discourse with robots" The scale is on a 6-point Likert scale, and like NARS, the individual score is calculated by summing the scores for each sub-scale [26,27].

Facial Expressions is a complementary way to assess emotions by observation users' facial expressions during the interaction. The observations offer additional information of users' emotions during the interaction with the robot, compared to the questionnaires that are distributed after the interaction. Facial expressions provide non-verbal cues that may interpret the user's emotional states. Relevant facial expressions comprise, but are not limited to, anger, frustration, happiness, confusion, and surprise. Although there is an ongoing discussion on how reliable facial expression are for studying emotions [4], we have chosen to include this metric as these emotions should be interpreted as indicators together with the other metrics rather than inferred alone.

3.2 Cognitive Processing

Cognitive processing is the factor that causes any sort of cognitive strain, which is resource demanding for the users when expectations are either confirmed or disconfirmed [29,30]. Expectations are the basis of attention and have a direct influence on perception. When expectations are disconfirmed, attention is drawn to the event (due to the surprising outcome) and the event is consequently processed. This process happens because the information challenges current beliefs [30]. Moreover, expectations drive interpretations of that particular event. Users try to interpret perceptual events so they are aligned with their present expectations. This phenomenon is commonly recognized as stereotyping but has prolonged effects on how we perceive our surroundings even when we do not construct explicit stereotypes. For example, disconfirmed low expectations (c.f., Fig. 2) when the robot outperforms our expectations, is commonly treated by users as a fluke or luck [29]. Conversely, it is also likely that users may attribute disappointing performance from a social robot as a temporary fluke, stemming from disconfirmed high expectations. However, as pointed out by Olson et al. [29], disconfirmed expectations of behavior can also be attributed to deception. Thus, a social robot (or their designers) may be viewed as having ulterior motives, which can be related to trust in robots [22]. Cognitive processing affects memory, and disconfirmed expectations have been shown to cause better memory recall as more effortful cognitive processing occurs in these cases [9]. Recall is also related to how much effort the user puts on making sense of disconfirmed expectations. For this factor of expectations, we identified two aspects that can be tested during the interaction: memory recall and reaction time.

Memory recall can be used to measure cognitive processing. Expectations have an effect on memory [29,30], and with inspiration from Hashtroudi et al. [9], we propose to measure users' recall ability after the interactions with a social robot. Hashtroudi et al. [9] found that memory works better with disconfirmed expectations. The authors pointed out that irrelevant information has the lowest recall ability. Although the authors' research had a more complex experimental set-up, we propose a simple set-up where users are asked to write down what the remember from the interaction afterwards to gain insight into their cognitive processing. Tying the recall ability to NARS and RAS, we argue that is possible to investigate whether disconfirmed expectations may lead to better memory recall of the interaction with the robot. This is especially interesting if certain characteristics are ascribed to the robot that are related to greater fear of the robot. Thus, if the users view robots as having scary characteristics in the surveys, and this turns out to be disconfirmed, it is likely that the users remember more of the interaction. Alternatively, users with strongly held expectations which are then disconfirmed, may have created false memories that align with their expectations (e.g., stereotyping may lead to seeing certain behaviors that are false) [30]. Therefore, when collecting data through memory recall after the interactions, it is important to consider how the users' memory may be affected by their expectations. This effect has been formalized in the peak-end rule [16].

It is possible that users remember certain characteristics of the robot that are aligned with their view of the robot, although their behavior is similar. For this reason, we stress how triangulation can be used to understand the data further (e.g., users may score high on fear of robots via NARS, and also describe scary characteristics of the robot).

Reaction time can be used to measure cognitive processing, inspired by the study by Hashtroudi et al. [9]. Reaction time has been seen to be longer when dealing with demanding cognitive processing [9,29]. Reaction time, in our framework, refers to the reaction time for the users to interact with the robot's output, both verbal (e.g., conversation) and non-verbal (e.g., touch). Reaction time needs to be tailored for the robot and its capabilities. Studying whether reaction time has an effect on expectations can be of value in itself, but we also propose that we can tie memory recall and reaction time together, similar to the study by Hashtroudi et al. [9].

3.3 Behavior and Performance

Behavior and performance are the factors that cause changes in an user's deliberate actions. Expectations are the basis for basically any behavior because expectations drive our intentions and actions [29,30]. One specific kind of behavior that may be of relevance for assessing expectations is hypothesis testing. Because expectations involve believed probabilities users may test various hypotheses relating to the expected likelihood of various kinds of interactions. This tends to happen with expectations that do not have a 100% certainty and expectations that are explicit. Hypothesis testing is thus a behavior exhibited by a user when trying to make sense of an event. In UX, hypothesis testing can be characterized in terms of the seven stages of action model by Norman [20,28]. Similar to hypothesis testing, the seven stages of action model is about forming the intention to act by specifying and then executing a sequence of actions, followed by evaluating the outcome in relation to the intention.

For the evaluation of expectations of social robots, we have identified five behaviors and performance measures, which are presented below. Although behavior and performance have more aspects than the previous factors of expectations, they are all rather simple to measure as they are collected by the test leader. According to Nomura et al. [26], the results from NARS may not lead to behavioral changes which could mean that it wont be possible to measure NARS and the below behavioral changes together. We are not claiming fear of robots does not cause behavioral changes, but rather that this specific NARS scale may not catch subsequent behavior, according to the authors. NARS may have effect on the affect factor of expectations but not the behavior and performance factor. We propose, however, to look at the below measures alone and in relation to RAS. Moreover, it should be mentioned that reaction time does also fit the factor of behavior and performance, and we have chosen to present it under cognitive processing because it can be tied to memory recall as well. Therefore, reaction time can be used in this factor as well.

Gestures and body language done by the users can be used to measure behavior. Gestures and body language are forms of non-verbal behaviors [19]. Body language could be head nods by the user to confirm they understood the robot, or leaning away from the robot to show discomfort. Gesturing could include pointing as a way to communicate direction for the robot, or waving hands in front of the robot in attempts to get the robot's attention. For a given interaction, there could be many gestures and communicative body motions, and the aim is to observe any behavior that is noticeable and tie these to the expectation factors. Gestures and language could thus be compared to, for example, interruptions of interaction. Perhaps the user may wave their hands when interruptions occur as an attempt to get back on track.

Choice of conversation may be dependent on what kinds of expectations users have towards the social robot and can be used to measure behavior. By choice of conversation, we mean what users will choose to say to the robot during the interaction. Of course, if the robot communicates non-verbally this aspect is not relevant. However, since many social robots are able to uphold simple conversations, it is likely that users try to figure out what conversations are possible. This behavior relates therefore to the seven stages of action model by Norman [28]; that is, the users may try to figure out what kind of conversations are possible. If a user's expectations of a social robot is disconfirmed, it is possible that the user will spend more time trying new ways to discuss things with the robot, rather than having one single conversation. If a user's expectations of a social robot is confirmed, is possible the user will show an excited facial expression and uphold the same topic longer.

Repeating words refer to the number of repetitions the user makes during the interaction and can be used to measure performance. Yet again, the seven stages of action model [28] could be of relevance here. By repeating words, we mean if users need to repeat themselves to be understood by the robot. Perhaps the user chooses other similar words but different words as a way to test which words actually can be understood by the robot. This may be related to affect such as frustration with not being understood, or have an effect on cognitive processing. Therefore, this should be triangulated with the collected data from that section in order to form the expectations. The user may expect that the robot will be able to uphold complex conversations, and when this does not occur (i.e., disconfirmed expectations), the user shows a frustrated facial expression and repeats themselves several times. Subsequently, the user may try to correct (or revise) their expectations which causes behavioral changes. If the user expects certain conversation, and the robot is able to confirm this expectation, there might not be repeating of words and the interaction goes smooth.

Interruptions of interaction refer to the number of interruptions during the interaction and can be used to measure performance. Interruptions could be caused by the user when the interaction is not going as expected. Interruptions could also imply that the conversation is not flowing. It is possible that disconfirmed expectations may occur due to these interruptions. Is also possible that confirmed expectations may occur due to the user expecting a bad

interaction. Interruptions may also be related to attempts to correct (revise) their expectations. Not only should the amount of interruptions be measured, but also under what circumstances they happen. Again, these could be tied to the other metrics in various ways, for example if choice of conversation changes after the interruptions and having better memory recall when interruptions occur.

Duration of interaction refer to the total time the interaction is taking place, and can be used to measure performance. As a rule of thumb, people behave in ways that are consistent with their expectations [29]. For example, people tend to choose tasks where they expect to be successful, and they will also put more effort and time into such tasks than ones they expect to fail [29]. The duration of the interaction, when applicable, could therefore be of relevance when users are interacting with the social robot; it is possible that users will interact longer with a robot if they expect to succeed in having a good interaction with the robot. Even more interesting would it be to compare duration of interaction in the first and second episodes of interaction. It is possible that users who had disconfirmed expectations of the interaction with the robot will have shorter duration of interaction in the second interaction episode as they know they will fail at having a successful interaction. This could be of interest in relation to the number of interruptions too. A lot of interruptions in the first interaction episode may lead to shorter duration of interaction in the second interaction episode since the user will expect these interruptions and will tend to avoid them in the second interaction. Perhaps scoring low on fear of robots may lead to longer duration of interaction as well.

4 Result

In this work, we have highlighted the importance of studying expectations in HRI. We have also started to specify how expectations could be studied and evaluated, and here we present the result: the Social Robot Expectation Gap Evaluation Framework. We base this framework on the model by Olson et al. [29] from the social psychology field, our Social Robot Expectation gap (Fig. 2) from the HRI field, and evaluation methods from the UX field [8,33]. This framework is intended for studying and evaluating expectations of social robots or other robots that act in a social manner in real interactions. The overall UX goal is to have expectations confirmed.

In this section we present the current version of our framework as illustrated in the matrix in Table 1. Included in this framework are UX goals and what metrics (data collection techniques) that are used to assess the three main factors. Each metric relates to either hedonic or pragmatic qualities [10]. Moreover, we present the proposed evaluation procedure (for further details on UX procedures, see [20].

Table 1. The three factors of expectations and the proposed metrics to study each when interacting with a social robot.

Expectation Factors	UX Goal	Metric	Details	Qualities
Affect	The user should expect to have neutral to positive emotions towards the robot	The Negative Attitude Toward Robot Scale (NARS)	A questionnaire measuring user's negative attitudes towards robots	Hedonic
		The Robot Anxiety Scale (RAS)	A questionnaire measuring user's anxiety towards robots	Hedonic
		Facial Expressions	Observing the kinds of facial expressions made by the user	Hedonic
Cognitive Processing	The user should experience effortless cognitive processing during the interaction	Memory Recall	Asking the user to write down what they remember of the interaction	Pragmatic
		Reaction Time	Measuring the time it takes for the user to react accordingly to the robot's output	Pragmatic
Behavior and Performance	The user should expect a pleasant and smooth interaction	Gestures and body language	Observing the kind of gestures and body language the user expresses	Hedonic
		Choice of Conversation	Observing the kinds of conversations the user tend to focus on during the interaction	Hedonic
	The user should expect to have ease of conversation	Repeating Words	Measuring the number of repetions the user makes during the interaction	Pragmatic
		Interruptions of Interaction	Measuring the amount of times, and what kind of, interruptions occur for the user during the interaction	Pragmatic
		Duration of Interaction	Measuring the total time spent in the interaction as well as the total time spent on each conversation for the user during the interaction	Pragmatic

4.1 Procedure

Phase 1: Identify the scenario. As a first step before carrying out the evaluation, one has to identify what kind of interaction to study, including identifying and creating scenarios between the human and the robot. We suggest two possible ways these scenarios can unfold, depending on what is being evaluated. First, the scenarios could be identical in order to compare them. If expectations are not confirmed in the first interaction, measuring changes in affect, cognitive processing, and behavior and performance is relevant as they would likely change. Another option is to have a scenario change in the second interaction, thus adding expectancy violation, similar to the study by Horstmann and Krämer [12]. Both these options offer different angles of studying and evaluating expectations of social robots. Once the scenario is chosen, baseline and max levels relating to the UX goals and metrics need to be set to fit the scenario [20]. We have chosen not to include baseline and max level in Table 1, as these need to be tailored to the actual scenario. This phase also includes recruiting participants, 5–8 is recommended for an empirical UX evaluation [20]. The different data collection techniques and metrics need to be prepared and tested in advance as well as informed consent to the participating users. The study needs to be in accordance with relevant ethical guidelines before being conducted [32].

Phase 2: Collect data. The proposed data collection techniques make it possible to collect complementary data for the factors of expectations (Table 1). In Fig. 3, we present a timeline of the step-wise procedure of this phase that aligns these aspects together. In particular, we highlight the temporal aspect of expectations, before, during, and after the interaction. We therefore urge the

investigators to allow participants to repeat the same scenario at least twice or modify the second one to investigate and analyze how the temporal aspect affects the expectations of the interaction with the robot. The data collection phase is divided into the following steps 1) before the interaction, 2) during first interaction, 3) after first interaction, 4) during second interaction, and 5) after interaction (Fig. 3). The data is collected via questionnaires, observation (field notes or recordings), and interviews.

Before the first interaction, it is important to understand what kind of expectations participants may have of social robots. Therefore, we suggest a pre-questionnaire where previous experience with robots is asked for. Questions regarding what they expect of the robot in the interaction could also be of importance, in order to study participants' explicit expectations. We urge investigators to tailor the pre-questionnaire so it suits the scenario and the actual robot. It is important to avoid generating expectations through the pre-questionnaire process; we urge researchers to put extra care into this process when selecting questions. As the final step of the data collection phase, we suggest to collect data from doing a post-interview. Here, open-ended questions regarding their expectations and if they were confirmed are valuable for the overall analysis of the participants' expectations.

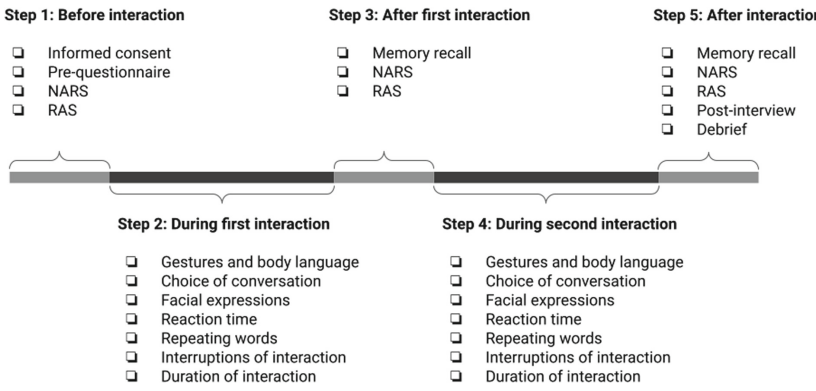

Fig. 3. The steps of phase 2: collect data

Phase 3: Analyzing the collected data. When the data collection has been carried out, the collected quantitative as well as qualitative data should be brought together and then analyzed, with a focus on identifying UX problems of coping with disconfirmed expectations. Triangulation is used to reach more reliable findings. Triangulation means that multiple data sources are used to compare and contrast the data in order to gain a deeper understanding of the obtained findings. Several findings that are pointing in the same direction imply that there is an identified UX problem that needs to be considered. The identified problems are then arranged into scope and severity. The scope can be either

global or local, where global problems entail the interaction with the robot as a whole, and local problems only entail certain moment(s) of the interaction. The severity of problems provides insights of which kinds of re-design that should be prioritized or what aspects need to be studied in more depth. Degrees of severity is ranging from high to low, where higher degrees include severe mismatches between users' expectations and the actual interaction between a human and a robot. Lower degrees of severity include problems that have a minor negative effect on expectations in the HRI, as in situations when it is easy for the user to find an effortless workaround.

Phase 4: Reporting on findings and recommendations. A major outcome of the findings is to what extent the users had their expectations confirmed or disconfirmed, the underlying reasons for these findings, and where these are situated on the social robot expectation gap (Fig. 2), with the overall UX goal is to have users' expectations confirmed. The scope and severity dimensions provide some recommendations for how to reduce the disconfirmed expectations in the chosen scenario as well as some insights of how and why the participants altered or changed their expectations.

5 Discussion

With this work, we aim to contribute to the field of HRI by addressing and studying in further detail the role and relevance of expectations in HRI, including how to narrow the gap between expected capabilities and actual capabilities of social robots (i.e., the social robot expectation gap, Fig. 2) thus achieving confirmed expectations of an interaction with a social robot. By doing so, we have consequently addressed what role as well as impact expectations may play in HRI, especially for social robots in interactions with humans. The developed evaluation framework, which is based on prior research on expectations, and its procedure serves as the initial steps to contribute to the above aim. This framework takes inspiration from social psychology, UX, and HRI, and we stress the temporal aspect of expectations. That is, expectations are dynamic over time and can change before, during, and after the interaction. Our framework offers ways to assess expectations from the different factors (i.e., affect, cognitive processing, and behavior and performance), how they may change over time, and if the expectations are confirmed or disconfirmed. We envision the framework to be tailored and adapted to the specific situation that is being studied. Therefore, as these metrics are extensive, it can be scaled down to a few selected metrics that are suitable for the chosen situation. Triangulation should be used for the analysis of these metrics to reach more reliable findings

Future work includes applying the framework in practice and collect empirical data, which has been hindered due to the current covid-19 pandemic situation. By conducting an empirical evaluation based on the framework, its potential could be validated and relevant improvements could be made on the current version of the framework. Additional implications of the framework is that the obtained findings based on severity and scope could offer significant insights for

future more experimental studies on certain aspects of humans' expectations of interacting with social robots. In the long run, this work will contribute to the inclusion of social robots in society.

References

1. Alač, M.: Social robots: things or agents? AI Soc. **31**(4), 519–535 (2015). https://doi.org/10.1007/s00146-015-0631-6
2. Alenljung, B., Lindblom, J., Andreasson, R., Ziemke, T.: User experience in social human-robot interaction. In: Rapid Automation: Concepts, Methodologies, Tools, and Applications, pp. 1468–1490. IGI Global (2019)
3. Alves-Oliveira, P., Ribeiro, T., Petisca, S., Di Tullio, E., Melo, F., Paiva, A.: An empathic robotic tutor for school classrooms. In: International Conference on Social Robotics. pp. 21–30. Springer (2015). https://doi.org/10.1007/978-3-319-25554-5_3
4. Barrett, L., Adolphs, R., Marsella, S., Martinez, A., Pollak, S.: Emotional expressions reconsidered. Psychol. Sci. Pub. Interest **20**(1), 1–68 (2019)
5. Dautenhahn, K.: Methodology themes of human-robot interaction. Int. J. Adv. Rob. Syst. **4**(1), 15 (2007)
6. Edwards, A., Edwards, C., Westerman, D., Spence, P.: Initial expectations, interactions, and beyond with social robots. Comput. Hum. Behav. **90**, 308–314 (2019)
7. de Graaf, M.M., Allouch, S.B.: The relation between people's attitude and anxiety towards robots in human-robot interaction. In: 2013 IEEE RO-MAN, pp. 632–637. IEEE (2013)
8. Hartson, H., Pyla, P.: The UX Book. Morgan Kaufmann, Burlington (2018)
9. Hashtroudi, S., Parker, E.S., DeLisi, L.E., Wyatt, R.J., Mutter, S.A.: Intact retention in acute alcohol amnesia. J. Exp. Psychol. Learn. Mem. Cogn. **10**(1), 156 (1984)
10. Hassenzahl, M., Tractinsky, N.: User experience - a research agenda. Behav. Inf. Technol. **25**(2), 91–97 (2006)
11. Horstmann, A., Krämer, N.: Great expectations? Relation of previous experiences with social robots in real life or in the media and expectancies based on qualitative and quantitative assessment. Front. Psychol. **10**, 939 (2019)
12. Horstmann, A., Krämer, N.: When a robot violates expectations. In: Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, pp. 254–256 (2020)
13. Horstmann, A.C., Krämer, N.C.: Expectations vs. actual behavior of a social robot. Plos One **15**(8), e0238133 (2020)
14. Jokinen, K., Wilcock, G.: Expectations and first experience with a social robot. In: Proceedings of the 5th International Conference on Human Agent Interaction, pp. 511–515 (2017)
15. Kaasinen, E., Kymäläinen, T., Niemelä, M., Olsson, T., Kanerva, M., Ikonen, V.: A user-centric view of intelligent environments. Computers **2**(1), 1–33 (2013)
16. Kahneman, D., Fredrickson, B.L., Schreiber, C.A., Redelmeier, D.A.: When more pain is preferred to less. Psychol. Sci. **4**(6), 401–405 (1993)
17. Khan, S., Germak, C.: Reframing HRI design opportunities for social robots. Fut. Internet **10**(10), 101 (2018)
18. Kwon, M., Jung, M., Knepper, R.: Human expectations of social robots. In: 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 463–464. IEEE (2016)

19. Lindblom, J.: Embodied social cognition. Cognitive systems monographs (COSMOS). Springer International Publishing Switzerland (2015). https://doi.org/10.1007/978-3-319-20315-7_3
20. Lindblom, J., Alenljung, B.: The anemone: theoretical foundations for UX evaluation of action and intention recognition in human-robot interaction. Sensors **20**(15), 4284 (2020)
21. Lohse, M.: The role of expectations in HRI. New Frontiers in Human-Robot Interaction, pp. 35–56 (2009)
22. Malle, B., Fischer, K., Young, J., Moon, A., Collins, E.: Trust and the discrepancy between expectations and actual capabilities. In: Zhang, D., Wei, B. (eds.) Human-Robot Interaction: Control, Analysis, and Design, chap. 1, pp. 1–23. Cambridge Scholars Publishing (2020)
23. Manzi, F., Massaro, D., Di Lernia, D., Maggioni, M.A., Riva, G., Marchetti, A.: Robots are not all the same. Cyberpsychol. Behav. Soc. Netw. **24**(5), 307–314 (2021)
24. Meister, M.: When is a robot really social? An outline of the robot sociologicus. Sci. Technol. Innov. Stud. **10**(1), 107–134 (2014)
25. Moore, R.K.: Is spoken language all-or-nothing? Implications for future speech-based human-machine interaction. In: Dialogues with Social Robots, pp. 281–291. Springer (2017). https://doi.org/10.1007/978-981-10-2585-3_22
26. Nomura, T., Kanda, T., Suzuki, T., Kato, K.: Psychology in human-robot communication. In: RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No. 04TH8759), pp. 35–40. IEEE (2004)
27. Nomura, T., Suzuki, T., Kanda, T., Kato, K.: Measurement of anxiety toward robots. In: ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication, pp. 372–377. IEEE (2006)
28. Norman, D.: The Design of Everyday Things. Basic Books, New York (2013)
29. Olson, J., Roese, N., Zanna, M.: Expectancies, pp. 211–238. Guilford Press (1996)
30. Roese, N., Sherman, J.: Expectancy. In: Kruglanski, A.W., Higgins, E.T. (eds.) Social Psychology: Handbook of Basic Principles. The Guilford Press, New York (2007)
31. Rosén, J.: Expectations in human-robot interaction. In: Neuroergonomics and Cognitive Engineering: Proceedings of the International Conference on Applied Human Factors and Ergonomics, pp. 98–105. Springer (2021). https://doi.org/10.1007/978-3-030-80285-1_12
32. Rosén, J., Lindblom, J., Billing, E.: Reporting of ethical conduct in human-robot interaction research. In: Zallio, M., Raymundo Ibañez, C., Hernandez, J.H. (eds.) AHFE 2021. LNNS, vol. 268, pp. 87–94. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-79997-7_11
33. Roto, V., Law, E., Vermeeren, A., Hoonhout, J.: User experience white paper - bringing clarity to the concept of user experience. In: Dagstuhl Seminar on Demarcating user Experience (2011)
34. Sandoval, E., Mubin, O., Obaid, M.: Human robot interaction and fiction. In: International Conference on Social Robotics. pp. 54–63. Springer (2014). https://doi.org/10.1007/978-3-319-11973-1_6

35. Schramm, L.T., Dufault, D., Young, J.E.: Warning: this robot is not what it seems! exploring expectation discrepancy resulting from robot design. In: Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, pp. 439–441 (2020)
36. Tonkin, M., Vitale, J., Herse, S., Williams, M., Judge, W., Wang, X.: Design methodology for the UX of HRI. In: Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, pp. 407–415 (2018)

PAPER IV

APPLYING THE SOCIAL ROBOT EXPECTATION GAP EVALUATION FRAMEWORK

Reprinted from Julia Rosén, Erik Billing, and Jessica Lindblom (2023). “Applying the Social Robot Expectation Gap Evaluation Framework”. In: *Human-Computer Interaction*. Ed. by Masaaki Kurosu and Ayako Hashizume. Springer International Publishing, pp. 169–188 with permission from Springer International.

Applying the Social Robot Expectation Gap Evaluation Framework

Julia Rosén[1], Erik Billing[1], and Jessica Lindblom[1,2(✉)]

[1] Interaction Lab, University of Skövde, Box 408, 541 28 Skövde, Sweden
{julia.rosen,jessica.lindblom}@his.se
[2] Department of Information Technology, Uppsala University, Box 337, 751 05 Uppsala, Sweden

Abstract. Expectations shape our experience with the world, including our interaction with technology. There is a mismatch between what humans expect of social robots and what they are actually capable of. Expectations are dynamic and can change over time. We have previously developed a framework for studying these expectations over time in human-robot interaction (HRI). In this work, we applied the social robot expectation gap evaluation framework in an HRI scenario from a UX evaluation perspective, by analyzing a subset of data collected from a larger experiment. The framework is based on three factors of expectation: affect, cognitive processing, as well as behavior and performance. Four UX goals related to a human-robot interaction scenario were evaluated. Results show that expectations change over time with an overall improved UX in the second interaction. Moreover, even though some UX goals were partly fulfilled, there are severe issues with the conversation between the user and the robot, ranging from the quality of the interaction to the users' utterances not being recognized by the robot. This work takes the initial steps towards disentangling how expectations work and change over time in HRI. Future work includes expanding the metrics to study expectations and to further validate the framework.

Keywords: Human-robot interaction · Social robots · Expectations · User experience · Evaluation · Expectation gap

1 Introduction

Although still on the brink to the general public, social robots and their extent of being situated in our everyday activities in society are becoming more sophisticated and common [3,6,17] which increases the expectations of such robots [7,10,12,14,16,24]. Social robots need to be able to communicate and act 'naturally' with their human users, not only on the socio-cognitive level but on user experience (UX) level. Social robots also need to achieve their intended benefits and support for human users [3,5,15,23,24]. The majority of human users have either no first-hand experience or very limited experience of interacting first-hand with social robots [12,14,16,24]. Human users' main exposure to social

M. Kurosu and A. Hashizume (Eds.): HCII 2023, LNCS 14013, pp. 169–188, 2023.
https://doi.org/10.1007/978-3-031-35602-5_13

robots is predominantly from movies and the media, which may result in false and incorrect expectations of social robots [7,10,11,26].

As pointed out by Stephanidis et al. [29], there is an identified need for the evaluation of interactive artificial systems to go beyond mere performance-based approaches that focus mainly on pragmatic qualities to embrace the overall user experience that also considers hedonic qualities. The authors argue that more traditional usability evaluation approaches in human-computer interaction (HCI) are rather insufficient for new interactive artificial systems, including social robots. Social robots are equipped with new perception and sensing possibilities, which enable them to shift initiatives via mutual action and intention recognition in conversations, displaying variations in morphology via having human-like attributes, and endowed with perceived socio-cognitive abilities. Moreover, there is an ongoing shift in application purposes, going from digital systems as being task-oriented tools to being considered as social companions or peers per se [2]. The identified challenges with these artificial systems include the need to interpret signals from multiple communication channels [29]. These channels could be eye and gaze following, pointing, body language, speech, and conversation in social robots that are conducted in a more natural way for the sake of social interaction. The unsuitability of task-specific measures in social robots, which predominantly are rather 'taskless', and therefore more focus should be on social interaction for the sake of companionship and creating a relationship. Stephanidis et al. [29] emphasized that if one should consider the enormous number of quality characteristics that should be evaluated in such human-artificial intelligent environments, like social human-robot interaction (HRI), it has become evident that new assessment methods are required. Hence, new frameworks and models are needed in order to provide holistic and systematic approaches for the evaluation of UX in human-artificial intelligent environments.

Roto et al. [25] noted that there are several overlaps between users' expectations as well as users' experiences when interacting with advanced artificial intelligence systems. User expectations could indicate the anticipated behavior, direct the focus of attention, serve as a source of reference for the actual UX and how it is interpreted, and subsequently has an impact on the user's overall perception of artificially intelligent systems [13]. Therefore, expectations often have a serious influence on the formation of the actual user experience of the social robot. This statement stresses that performing research on expectations can reveal a deeper understanding of the central aspects of user experience. One of the main components of UX, which is often missing in social robotics and HRI research, is the temporal aspect of the interaction [25].

Although there is limited research on the temporal aspect of expectations in HRI, there is some related research conducted on the changes in expectations. Paetzel et al. [21] investigated how persistent the first impression of a robot is, with different perceptual dimensions stabilizing at different points over three interactions. In the study, competence was stabilized after the initial two minutes in the first interaction, anthropomorphism and likeability were set after the sec-

ond interaction, and perceived threat and discomfort were unstable until the last interaction. Serholt and Barendrgt [27] found that children's social engagement towards a tutoring robot decreased over time, suggesting that the human-human interaction (HHI) model of engagement faded out, or were used more seldom, as the robots did not meet children's expectations for engagement. Edwards et al. [7] found expectations of a robot have an effect on first impressions, displaying more certainty and greater social presence after the brief interaction.

Despite the recent interest in expectations in social robots in the HRI field, there was a lack of an evaluation framework that offers a deeper understanding of how these expectations affect the success of the human-robot interaction from a first-hand perspective. Therefore, we developed the social robot expectation gap evaluation framework [24]. The framework provides a methodological approach for investigating and analyzing users' expectations before, during, and after interaction with a social robot from a human-centered perspective [24]. The framework has its foundation in the social psychological expectation process developed by Olson et al. [20] and user experience (UX) evaluation methodology [8,15,23]. The framework contains three main factors in which users' expectations can be evaluated: affect, cognitive processing, as well as behavior and performance. Several UX goals and related metrics were formulated for each factor, which relates to either pragmatic or hedonic qualities of the social human-robot interaction experience [8,9,15]. Moreover, the framework contains a four-phased evaluation procedure that consists of 1) identifying the scenario, 2) collecting data, 3) analyzing the collected data, and 4) reporting on findings and recommendations (for details, see Rosén et al., [24]).

In this paper, we report on how parts of the social robot expectation gap evaluation framework [24] are applied in an empirical UX evaluation. Our study is exemplified by a subset of data collected from a larger study on the role and relevance of expectations in social HRI, which examined how expectations may affect the forthcoming interaction quality and how expectations may alter user experience over time. The main contributions of this paper are two-fold; first, it provides an initial validation and testing of a sub-part of the framework which systematically studies how users' expectations may be altered if their initial expectations are not confirmed and its impact on user experience. Second, the implications of the findings highlight that expectations are especially important to understand when they are disconfirmed as this can guide and inform the future design of social robots, which will ultimately lead to more successful interactions.

2 The Social Robot Expectation Gap Evaluation Framework

The social robot expectation gap evaluation framework has its foundation in the social psychological expectation process developed by Olson et al. [20] and user experience (UX) evaluation methodology [8,15]. The framework is shown in Table 1. Drawing from the model by Olson et al. [20], we here briefly present the main characteristics of expectations as described in Rosén et al. [24]. In our

modified model, we present expectations as formed by i) direct experience, ii) other people or robots, and iii) beliefs. Direct experience is the expectations built on first-hand experience, i.e., from an actual encounter with a social robot, which is usually lacking in HRI research on expectations. Expectations that are built on direct experience are a stronger predictor of future behavior [20]. Other people or robots are expectations from indirect experiences, such as views from peers and friends or from exposure to social robots in various ways in movies and other media. Beliefs are sources of expectations that can be inferred from other beliefs, e.g., a robot may be evil because they usually are depicted as such in science fiction.

Regardless of its origins, expectations can vary in four dimensions: certainty, accessibility, explicitness, and importance [20]. Certainty is the subjective estimated probability of how likely it is that the outcome will occur. Accessibility is how easy it is to activate and use a certain expectation. Explicitness is what degree expectations are consciously generated. Some expectations are implicitly assumed, usually related to the degree of certainty, whereas other expectations are consciously reflected upon. Importance is the expectation's significance, the higher the importance, the higher the impact.

As explained by the model, there are consequences of expectations, which is the main focus for our framework [20,24]. These factors can be divided into three categories: i) affective, ii) cognitive processing and iii) behavioral and performance. Affective refers to emotions and feelings, such as anxiety. Cognitive processing refers to factors that have an effect on our cognitive processes, such as interpretation and memory abilities. Behavioral and performance refer to consequences that cause a course of actions due to expectations, like forming intentions to act. Behavior and performance are the factors that cause changes in the user's deliberate actions. Expectations serve as the basis for essentially any behavior since expectations initiate our intentions and actions.

The above factors are the consequence of when expectations are confirmed or disconfirmed [20]. Confirmed expectations often result in positive affect and feelings that are frequently performed implicitly and effortlessly, resulting in expectations that are maintained with greater certainty. Confirmed expectations may also, in rare situations, produce secondary (positive or negative) affect since the inferences are made after confirming the expectations. On the contrary, disconfirmed expectations often end up with negative effects and are usually considered explicitly since they are perceived as surprising and therefore need heavy cognitive processing in order to interpret and understand the contradiction. We further illustrate confirmed and disconfirmed expectations with the social robot expectation gap [24] depicted in Fig. 1. The social robot expectation gap demonstrates the two spaces of disconfirmed expectations, which either is the result of too high or too low expectations, balancing on the thin line of confirmed expectations. The underlying cause of disconfirmed expectations, either too high or too low, is the perceived or experienced mismatch between the social robot's capabilities and the user's expectations of the social robot which positively or negatively affects the overall interaction quality, and hence the perceived user experience of

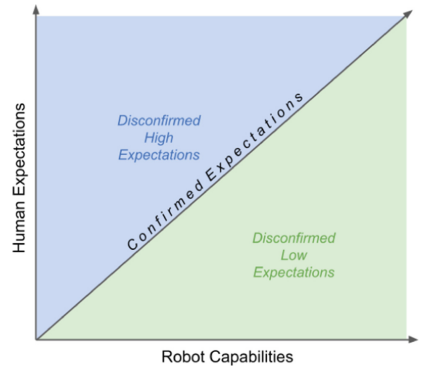

Fig. 1. The social robot expectation gap, with the two spaces of disconfirmed expectations that occurs when a robot's capabilities do not align with the user's expected capabilities [24, p. 599].

the social HRI. It should be pointed out that in the case that the user's expectations are low, there is a risk that the user underestimates the robot's actual social interaction capabilities which may result in less curious inquiry and exploration, resulting in less discovery of its interaction capabilities – resulting in a poor user experience. Similarly, in the case when the user's expectations are very high and the user overestimates the robot's interaction capabilities, assuming that the robot is equipped with advanced socio-cognitive interaction abilities, the outcome may be disappointment and frustration because the user feels deceived – resulting in a poor user experience. A user may also have rather low or moderate expectations of the social robot, but through inquiry during the interaction with the robot that develops into an ongoing and mutual dialogue, the user may be positively surprised by the robot's social interaction capabilities, feeling satisfied and noticed by the robot, resulting in positive user experience.

The current version of our developed evaluation framework for expectations was created in order to investigate the consequences of users' high and low expectations of social robots when interacting with a social robot – before, during, and after the interaction. The evaluation procedure consists of four phases: 1) identifying the scenario, 2) collecting data, 3) analyzing the collected data, and 4) reporting on findings and recommendations (for details, see Rosén et al. [24]). The framework should be viewed as modular, where UX goals and metrics can be added or removed in order to target the different expectations factors. Included in the framework are the formulated UX goals and the metrics (data collection techniques) used to assess the three main factors. These UX goals and metrics can be found in Table 1. Each metric relates to either hedonic or pragmatic qualities [9].

Table 1. The social robot expectation gap evaluation framework

Expectation Factors	UX Goal	Metric	Details	Qualities
Affect	The user should expect to have neutral to positive emotions towards the robot	The Negative Attitude Toward Robot Scale (NARS)	A questionnaire measuring user's negative attitudes towards robots	Hedonic
		The Robot Anxiety Scale (RAS)	A questionnaire measuring user's anxiety towards robots	Hedonic
		Facial Expressions	Observing the kinds of facial expressions made by the user	Hedonic
Cognitive Processing	The user should experience effortless cognitive processing during the interaction	Memory Recall	Asking the user to write down what they remember of the interaction	Pragmatic
		Reaction Time	Measuring the time it takes for the user to react accordingly to the robot's output	Pragmatic
Behavior and Performance	The user should expect a pleasant and smooth interaction	Gestures and body language	Observing the kind of gestures and body language the user expresses	Hedonic
		Choice of Conversation	Observing the kinds of conversations the user tend to focus on during the interaction	Hedonic
	The user should expect to have ease of conversation	Repeating Words	Measuring the number of repetions the user makes during the interaction	Pragmatic
		Interruptions of Interaction	Measuring the amount of times, and what kind of, interruptions occur for the user during the interaction	Pragmatic
		Duration of Interaction	Measuring the total time spent in the interaction as well as the total time spent on each conversation for the user during the interaction	Pragmatic

3 Method

In this paper, we report on how parts of the social robot expectation gap evaluation framework can be applied from a UX evaluation perspective. This is exemplified by a subset of data collected from a larger experiment on the role and relevance of expectations in social HRI (for a full list of metrics for the full experiment, see Table 1). The structure of this section follows the phases of the framework to illustrate how our testing was performed: scenario, data collection, and data analysis. The final phase, results, is presented in Sect. 4.

Fig. 2. The set-up of the interaction

3.1 Phase 1: Scenario

The developed scenario enabled the human user to interact first-hand with the Pepper robot in our lab. The physical layout of the lab is a 60 m^2 room where half of the open space was dedicated to the human-robot interaction setup. The underlying idea was that the user and the robot should be able to get to know each other in a more exploratory way, being engaged in dialogue. There were

two interaction sessions in total, lasting 2.5 min each. The users were informed that the study's aim was *to investigate how people interact with a social robot that is intended to be used in the home and that they could ask the robot about anything*. A pilot was conducted to try out the scenario. Once the scenario was chosen, baseline and max levels relating to the UX goals and metrics were set to fit the scenario. The baseline level considered the minimum acceptable level for the interaction and target levels and the desired level for the interaction. These levels are presented in the Subsect. 3.2.

The Robot and the Dialogue System: To study users' expectations of robots from a first-hand perspective, we created a scenario with the social robot Pepper manufactured by Aldebaran [28]. Pepper is a 120 cm tall social robot designed to interact with humans. The *autonomous life* functionality built into the Pepper robots was used, including simulated breathing and awareness (head turns) towards the user [28]. The robot was equipped with a customized dialogue system powered the OpenAI GPT-3 large language model [1], developed by the second author. Users' speech was recorded using the robot's microphones and translated into text using Google's speech-to-text cloud service. Recognized text was sent to the GPT-3 *text-davinci-002* model for computation of suitable dialogue responses. Resulting text responses were transformed into spoken language using the *ALAnimatedSpeech* module, built into the default NaoQi middle-ware delivered with the Pepper robot. The animated speech module produced synthetic speech accompanied by head and arm gestures. A more detailed description of the dialogue system is available in [4]. The motivation for developing the above dialogue system was to enable a more natural, smooth, and intuitive dialogue between the user and the Pepper robot.

3.2 Phase 2: Data Collection

We purposely sampled [22] a subset of ten ($N = 10$) users who were recruited for a larger empirical study in our lab. The inclusion criteria were 1) that they had no prior experience of interacting with social robots, and 2) that they were either Swedish or English native speakers. There were 6 women, 4 men (no-one self-described or chose non-binary), with ages between 20–49 years ($M = 31$). The users were informed of the purpose of the study and informed consent was distributed, read, and signed by the users before the interactions. The evaluation was conducted in accordance with the Declaration of Helsinki.

Once entering the lab, the users filled in a questionnaire with background information about their age, gender, first language, previous experience with robots, and interest in robots. The users were asked to sit on a chair approximately one meter in front of the Pepper robot and were then instructed to engage in open conversation with the robot, with sessions for filling in questionnaires between the two interactions. The two interactions were video recorded. At the end of the two interactions, an open-ended post-test interview was conducted that focused on the users' experience of interacting with the robot. Afterward,

the users were debriefed, encouraged to raise any questions, and thanked for their participation. They were informed how the Pepper robot and its speech system functioned. The first author did the data collection. We focused on the following metrics:

The Robot Anxiety Scale (RAS): RAS was collected before the interaction, after the first interaction, and after the second interaction. RAS consists of three subscales with different themes relating to anxiety: S1: anxiety toward communication capability of robots, S2: anxiety toward behavioral characteristics of robots, and S3: anxiety toward discourse with robots. The scale is on a 6-point Likert-scale, with low scores meaning less anxiety (1: I do not feel anxiety at all, 6: I feel anxiety very strongly), and the individual score is calculated by summing the scores for each sub-scale [18,19]. The base levels for RAS before the interaction were for S1: 9, and for S2 and S3: 12; baselines after the first interaction were for S1: 8, and for S2 and S3: 11; and baselines after the second interaction were for S1: 7, and for S2 and S3: 10. Target levels for RAS before the interaction were for S1: 7, S2 and S3: 10; target levels after the first interaction were for S1: 6, S2 and S3: 9; and target levels after the second interaction were for S1: 5, and for S2 and S3: 8.

Interruptions: Video recordings from the interaction for the amount of time the users were interrupted by the robot while talking were collected. The baseline levels for interruptions for the first interaction were 2 and for the second interaction was 1. The target level for the first interaction was 1 and for the second interaction was 0.

Post-test Interview: These five questions were asked after the second interaction: (1) How did you feel the interactions went?, (2) Did you experience any difference between the first and second interaction?, (3) Did you have any expectations of how the interaction would go?, (4) Was anything surprising about the interaction or the robot?, and (5) Did you have any specific emotion during the interaction? The interviews were video recorded and then briefly transcribed.

Observations: Field notes were taken by the first author and video recordings were done for upcoming analysis.

The baseline and target levels were the same for all of the qualitative measures (post-test interviews and observations). The baseline for the first interaction was negative, and the baseline for the second interaction was neutral. The target level for the first interaction was neutral, and the target level for the second interaction was positive.

3.3 Phase 3: Analysis of the Data

The collected data were analyzed via triangulation [8,15,22], in which the collected quantitative and qualitative data were compared and contrasted, focusing

on identifying UX problems of coping with disconfirmed expectations. The analysis centered on the four UX goals. Of special interest were if anything required more cognitive processing, affect, or any altered behavior during the interaction. After the data collection, the video recordings of the interaction sessions with the Pepper robot and the post-interviews were analyzed and briefly transcribed by the last author. The transcripts from the video recordings were then analyzed, zooming in on interruptions, repetitions of questions, or hesitations from the users while interacting with the robot. We also analyzed and interpreted the characteristics and content of the human-robot conversations, users' facial expressions, and body language when interacting with the Pepper robot. We particularly looked for evidence and findings in the data that pointed in the same direction, implying that there we identified relevant UX problems that needed to be considered.

4 Results

In this section, the results from the collected quantitative and qualitative data via triangulation [8,15,22] of the RAS questionnaire, number of interruptions as well as the analysis of the observations and interviews are presented. For an overview of the results in relation to the UX goals, see Table 3.

First, the overall findings regarding users' experiences interacting with the Pepper robot and the expectations of interacting with the robot before, during, and after are presented and described. We present whether, or to what extent, the four UX goals are not fulfilled, partly fulfilled or fulfilled in Sect. 4.1. The outcome of the triangulation is then presented for each of the four UX goals, in which the most relevant positive or negative UX aspects related to how users' expectations are presented. The presentation of these findings consists of descriptions of most important identified aspects of expectations and UX problems combined with quotes from the users. In Sect. 4.2, we present some recommendations based on the scope and severity ratings of the identified UX problems.

4.1 Aspects Related to the Four UX Goals

We here portray a more nuanced picture of this particular UX goal and identified UX problems.

UX Goal 1: The User Should Expect to Have Neutral to Positive Emotions Towards the Robot. The findings show that on a general level, this UX goal was not fulfilled. The RAS scores strongly support the claim that this UX goal is fulfilled because it reaches the target levels in all instances, except for RAS subscale 3: anxiety toward discourse with robots, by one point in 'before the interaction'. The data for RAS were collected for each subscale and before the interaction, after the first interaction, and after the second interaction. A high RAS score means higher level of anxiety. The scores are presented for each data collection point: The overall scores for RAS S1 (range: 3–18) were 6

Expectation Factors	UX Goal	Metric	Details	Qualities	Baseline before interacion	Baseline after first interaction	Baseline after second interaction	Target level before interacion	Target level after first interaction	Target level after second interaction	Observed results before interacion	Observed results after first interaction	Observed results after second interaction	Meet target before interacion	Meet target after first interaction	Meet target second interaction
Affect	The user should expect to have neutral to positive emotions towards the robot	The Robot Anxiety Scale (RAS)	A questionnaire measuring user's anxiety towards robots	Hedonic	S1: 9 S2: 12 S3: 12	S1: 8 S2: 11 S3: 11	S1: 7 S2: 10 S3: 10	S1: 7 S2: 10 S3: 10	S1: 6 S2: 9 S3: 9	S1: 5 S2: 8 S3: 8	S1: 6 S2: 10 S3: 11	S1: 5 S2: 8 S3: 8	S1: 5 S2: 7 S3: 8	S1: yes S2: yes S3: no	S1: yes S2: yes S3: yes	S1: yes S2: yes S3: yes
		Facial expressions	Observing the kind of facial expressions made by the user			Negative	Neutral		Neutral	Positive		Negative	Negative		No	No
		Post-test interview	Asking the user if the felt any emotions during the interaction													
Cognitive Processing	The user should experience effortless cognitive processing during the interaction	Observations and post-test interview	Observing the interaction and asking the user what was surprising about the interaction	Hedonic		Negative	Neutral		Neutral	Positive		Negative	Positive		No	Yes
Behavior and Performance	The user should expect a pleasant and smooth conversation	Observations and post-test interview	Observing the interaction and asking users about their behavior during the interaction	Hedonic		Negative	Neutral		Neutral	Positive		Negative	Neutral		No	No
	The user should expect to have ease of conversation	Interruptions of Interaction	Measuring the amount of times, and what kind of, interruptions occur for the user during the interaction	Pragmatic		2	1		1	0		0.3	0.3		Yes	No
		Observations and post-test interview	Observing the interaction and asking users about their interaction	Hedonic		Negative	Neutral		Neutral	Positive		Negative	Neutral		No	No

Fig. 3. The applied social robot expectation gap evaluation framework, with set levels and results

($SD = 2.26$), 5 ($SD = 2.98$), 5 ($SD = 2.10$). The overall scores for RAS S2 (range: 4–24) were 10 ($SD = 4.75$), 8 ($SD = 4.24$), 7 ($SD = 3.66$). The overall scores for RAS S3 (range 4–24) were 11 ($SD = 3.25$), 8 ($SD = 4.76$), 8 ($SD = 3.29$).

For the qualitative data, there is a mixture of experiences among the users moving toward the negative, and there are mixed feelings and facial expressions within an individual user from the qualitative data. For example, four of the users displayed rather hesitant or reluctant behaviors toward the robot during the interactions, such as sitting in front of the robot in more defensive positions. Examples of postures were leaning a little backward on the chair, crossing their arms in front of the chest, or putting their arms on top of the crossed legs. One user displayed several signs of stress or anxiety, frequently scratching one of her legs intensively. The users having a more reluctant position quite often squeezed or played with their hands or fingers, especially when the interaction did not proceed well or when the robot did not respond. Two of the users displayed a more neutral position, looking interested but still a bit reserved. Two users were leaning forwards toward the robot and looked interested and seemed to invite a closer interaction space with the robot.

One user was very frightened by the robot initially and thought that the robot actually would attack or punch him while the robot raised its arms. One of the more reluctant users giggled repeatedly during the interactions, and although the users most of the time focused their gaze on the robot, several users looked away warily from time to time. However, one of the users who was rather reluctant from the very beginning actually moved the chair closer to the robot after a while and leaned more forward. Once when the robot's arms reached out towards the user, the same user immediately leaned slightly backward, but then leaned forward when the robot's arms were put to a more natural position. Many users displayed rather curious or interested facial expressions, albeit a bit reluctant. One user displayed a fearless facial expression and was very active in the interactions. Two users looked more neutral although still interested. Many users displayed rather confused or puzzled facial expressions, albeit usually looking more interested and even smiled a lot when the interaction went well or when the robot responded to them directly. Only two users looked very amused or amazed and continued to be in a positive state throughout the interactions. Most displayed behaviors were rather stable during the interactions, with the general impression that they were more relaxed during the second interaction and then focused more on engaging in conversations with the robot than considering emotions towards the robot itself.

Several statements by the users at the post-test interviews showed rather mixed or negative emotions towards the robot. For example, one user who displayed a rather reluctant behavior stated: *"I didn't have any emotions toward the robot."* Many users expressed that the experience was very interesting and fascinating and that they did not really know what to expect in general, and consequently did not know what emotions to experience from the robot explicitly. It seems that many different feelings and emotions occurred at once which caused them to be unable to categorize or verbalize their emotions. This can indicate

a tentative UX problem due to a lack of prior experience and that they have not yet formed any precise emotions towards the robot. Therefore, the majority displayed and expressed rather puzzled or hesitant emotions towards the robot although being interested or fascinated at the same time. In sum, the dominant experience is rather negative primarily in relation to expecting to have neutral to positive emotions towards the robot.

UX Goal 2: The User Should Experience Effortless Cognitive Processing During the Interaction. The findings show that on a general level, this UX goal was partly fulfilled. The dominant user experience is slightly negative initially and alters into a more positive one in relation to effortless cognitive processing between the first and second interaction. There is a mixture of how effortless the cognitive processing was during the interactions among the users, since the ways the interactions with the robot unfolded affected cognitive processing. The human-robot interaction proceeded very smoothly for two of the users, and they did not experience any cognitive strain. For the rest of the users, the picture is more puzzling. One user expressed: *"I was rather unprepared, so it was a bit difficult to know what to talk about, but my impression is that he [the robot] tries to answer in a way so I should feel as comfortable as possible. However, I noticed that he lied to me since he said that he liked to eat food, which make me realize that maybe he lies about other stuff too . . . which was slightly uncomfortable."* The user continued: *"the robot's attempt to have no opinion and be impersonal makes him a little uncomfortable, it becomes difficult to categorize it [the robot] and we humans like to do that.. it feels strange that it doesn't have any personality."*

Several users struggled with engaging the robot in a dialogue, and when the robot did not respond swiftly, or not at all, they expressed confusion about how to behave. For instance, they repeated questions, raised their voice, and asked other questions. One user said: *"it was hard to know what to talk about with the robot, I don't know what level the robot is at."* One user felt embarrassed and guilty when the robots' reply was objective and not personal regarding what culture he liked: *"when I talked about cultures, he didn't answer which country [the robot liked] but that all cultures are exciting...[implying] that he has no preferences. . . it's very much like this [nervous laughing] unpredictable...No, it feels stiff..., but I felt like a bit crappy because I thought it was stiff, you felt a bit guilty...on behalf of the robot [laughing]."*

For another user, during the first interaction, when the dialogue had not gone well for a while, and the robot's responses were totally random with regard to the content of the raised questions, the dialogue ended. The same user expressed explicitly: *"I don't have any idea what to say, it [my mind] stands completely still."* Another user expressed similar thoughts: *"Instead of having a conversation, the robot responded with long answers like. . . pre-programmed... You can only ask one question and know it understood what I said... then I had to sit and think – what should I ask now?."* The above users raised rather polite or personal questions to the robot, in order to get to know the robot better and

many of these questions were rather personal, such as: *What is your name? How old are you? Where do you live? Do you have any friends? What do you like to eat?*

One of the users applied a different approach after a while and explicitly tried to test and challenge the robot's capability in more detail. He had asked the robot how old it was, and when the robot replied *"I'm three years old"*, the user followed up by asking: *"If you are three years old, what year were you born?."* The robot's response was delayed and then it answered *"2016."* The user then explained during the interview: *"I wanted to see how smart he was... [I'm] very impressed actually, but it didn't match the age when he tried to count... you still get the feeling that he is programmed in what he should say... He doesn't have his own... identity... I know he doesn't eat Sushi so you taught him that, because I can prove that robots don't eat Sushi."*

It was revealed that the users experienced less cognitive effort during the second interaction, indicating that they have acquired a certain way of interacting with the robot that was experienced more effortlessly. For example, the user that did not know what to say earlier now succeeded to engage in a dialogue and put a big smile on her face. Another user explained the difference between the first and second interaction: *"right at the beginning, I felt, when I didn't get any response and so, is it me who don't pronounce things properly... then you got a little pensive and a bit worried, but then it [the interaction] started and it felt better and then it was like when you are interacting with people, such as I ask a question and they have to come up with an answer... I have to come up with something new to ask. It was a bit more fun the second time when the robot used some body language... I was a bit amused."*

We also noticed during analysis that RAS's subscale 2: anxiety toward behavioral characteristics of robots could hint at cognitive processing. The subscale deals with how the robot may act during an interaction, which may affect cognitive processing as unexpected behavior causes extra cognitive processing to make sense of the behavior and how to react to it. The overall RAS S2 scores (range: 4–24) were 10 (SD = 4.75), 8 (SD = 4.24), and 7 (SD = 3.66) which suggested that the cognitive strain may have decreased after the interactions.

The identified UX problems were that the users have to construct questions that the robot could answer properly at its level of capability and that the robot itself seems to lack a kind of personality or identity that should add something extra while getting to know each other. Hence, there seem to be explicit user expectations that the Pepper robot is a machine, while they at the same time implicitly expect human-like aspects in the interaction.

UX Goal 3: The User Should Expect a Pleasant and Smooth Interaction. The findings show that on a general level, this UX goal was not fulfilled. For this particular UX goal, we focused mainly on the general flow between the users and the robot during the first and second interactions. For three of the users, it was rather hard to establish an interaction during the first interaction. The identified reasons for that were that the robot was unable to perceive the

users' voices, mainly because they spoke too quietly or that the robot could not recognize what they said. As one user reflected: *"it feels unnatural to talk so loudly when you are sitting as close as you do to talk to the robot."* On three occasions the robot needed to be restarted by the test leader. As a consequence, one of these users frequently turned her face towards the test leader, who was sitting behind a screen, in an attempt to get some support when the information flow was not fluent. Another consequence was that several users were rather hesitant and unsure of how to interact with the robot to experience a smooth interaction.

It was revealed that the quality of the interaction also depended on how the user's questions were raised and what kind of questions were asked. Many questions were more on declarative knowledge, like common facts and basic knowledge about Pepper. For these kinds of questions, the human-robot interaction went rather smoothly. But if the questions raised were about more procedural knowledge and skills, the robot's responses were not that highly appreciated. For example, some users asked if the robot could perform some movements, dance, and sing songs like 'Happy Birthday' or 'Baa Baa White Sheep'. These users seemed to examine whether, or to what extent, the robot was able to perform these tasks. They were rather disappointed; although the robot moved its arms, the robot neither danced nor sang. The robot's reply was that it was able to sing'Happy Birthday' (without singing the tune) and the response for the 'Baa Baaa White Sheep' sing request was to utter "bad, bad". As one user expressed it: *"I have to speak slowly, you can't speak too fast and [you have to] speak clearly too... as you might do with older people while the robot answers like a child... it becomes very shallow, I don't like to speak in this way."* Another user said: *"I don't know what to expect, there was a lot of stops [in the interaction], probably because it won't be the same conversations as with a person."* Other users argued that the interaction was a bit repetitive and stated that *"[if you make] short commands you got relevant answers, but otherwise it was not possible to have a good communication."*

However, three of the users experienced much more pleasant and smooth interactions and one user said: *"my first impression was that it was rather intelligent, but better than a chatbot, and he [the robot] thought about weather and could learn facial expressions... I had slightly pessimistic expectations before and I wondered if he [the robot] can read facial expressions, feels like he can do it. Surprisingly he did... I felt quite happy during the interaction, quite a unique experience!"* Two of the users' interactions with the robot went very well, one said: *"[the robot] was really cute, I thought it would be less advanced... and so it was super!"* and the other said *"it wasn't difficult or hard to talk with the robot"* both expressing very happy facial expressions and with fascination in their voices.

It was evident that the expectations of a more pleasant and smooth interaction were confirmed to a higher extent between the first and second interaction generally. One user that had a rather non-fluent interaction during the first session, said: *"it was one-sided, ... [like an] interview, it answered very gener-*

ally... kind of having google in front of you, but the second turn was better, it is fascinating compared to what people want." Another user raised similar thoughts, stating: *"I knew better how he [the robot] behaved... I knew a little better how it moved and didn't"*, and *"you know more what you have to play with...it's like that with all the people you meet at first, it's a bit tough...I knew a bit more the second time."* Although several users expressed that they have learned how they interacted with the robot between the interactions, there were some hesitations about the robot's actual interaction behavior and capabilities, as one user reflected: *"it was exciting that they [robots] have so many answers... it feels like they are looking up the answers to what you are talking about... and it is a bit spooky... you don't know what they are capable of."*

Another expressed that the experience of interacting with the robot was a bit unpleasant: *"I'm feeling curiosity and a little discomfort, not in such a way that you are in danger but more what will happen... so that if you talk about the same thing, he could answer something outside of what you talked about..."*

The identified UX problems were that the robot did not respond to some voices, the questions should be stated in a certain manner for a smooth interaction, and the robot was unable to respond by performing actions or behaviors asked for to a high extent.

UX Goal 4: The User Should Expect to Have Ease of Conversation. The findings show that on a general level, this UX goal was partly fulfilled. The target level for interruptions in the first interaction was met; however, there were interruptions in the second interaction. Four of the ten participants experienced an interruption by the robot; two of them experienced interruptions by the robot twice, and the other two experienced interruptions by the robot once.

For the qualitative data, we focused on the conversation quality between the users and the robot during the first and second interactions. As revealed in the third UX goal, the ease of conversation varied in the interactions between the users and the robot. One user had a non-fluent interaction with the robot from the start, because the robot did not recognize his voice. The user reacted to this by moving his chair closer to the robot and reaching for its hands. When the interaction was ongoing he then leaned backward. He later on explained, while gesturing vividly, that the expected the conversation to be more verbal and that the robot would be more engaged in the conversation. He experienced it rather surprising that the robot didn't respond, and that the robot sometimes made rather random moves that he considered a bit uncomfortable. He then explained that he felt a bit embarrassed when the robot didn't respond to his interaction attempts, verbally and non-verbally. He stated that he felt a little anxious, but at the same time curious about the robot during their uneasy conversations. The cumbersome conversations resulted in many hesitations and lack of interaction between them, and he concluded: *"I felt uncomfortable with the silence between us, as soon as he [the robot] answered, I wanted to ask another question... I didn't want it to be quiet and he would look at me... but this [characteristics of] human*

interaction is not there. . . .” Hence, the silence between them was experienced as uncomfortable from a human-human interaction perspective.

In contrast, a user with a very fluent and pleasant interaction said: *“It [the robot] was really cute, I thought it would be less advanced and so it was super. . . surprising how good it could answer things... could keep the thread and that it could ask a follow-up question and that it joined the conversation...you didn’t have to clarify... it was surprising.”* She then explained that she felt a lot of curiosity, little nervousness, and that it was a super interesting experience.

It was also revealed that the majority of users experienced qualitative differences in the ease of a conversation between the first and second interactions. As one of them said: *“in the first [interaction], I asked questions and in the second one, it felt more like an interaction ... a conversation ... in the second one it was a flow. . . because you ended up in the interaction more. . . I felt that the second time I started talking, the robot reacted to me. . . The first time the robot didn’t ask any questions back, it was stilted, but the other time, it flowed like that then it happened that I didn’t think about that... because the robot made suggestions that maybe we should do this or that. . . it wasn’t like the robot was leading the conversation but that I came up with other things [to say] and so on.”*

It became rather evident that the users, although aware that the robot was an artifact, still made comparisons to how humans act in a conversation. For example, one user explained: *“when I first met the robot and sat down I felt a little bit nervous, I did not know what to do and so, but then I thought that she [the robot] should have done something like ‘O hello, please sit down [while the user made a ‘have a seat’ gesture].”* He said that he was very clueless and nervous since he did not know what to do initially. He ended by arguing: *“she [the robot] should say the first things, she should start. . . ’Hi, welcome’ and stuff. . . because if you just sit. . . If you see how humans interact with each other, one always takes the first step, and when we talk with robots, we know that they are not as intelligent as humans, so they should say the first word or so just so we can feel relaxed.”*

The identified UX problems were that most of the users did not experience any ease of conversation, mostly because their expectations were that the conversation should mimic human conversation aspects although the robot was not a human but a machine. Hence the robot’s appearance and behavior resulted in mixed expectations.

4.2 Severity and Scope of the Identified UX Problems

In this section, we have arranged the identified UX problems into scope and severity. The scope was either global or local, where global UX problems entail the interaction with the robot as a whole, and local problems only entail certain moment(s) of the interaction. The severity of the UX problems provides insights into which kinds of re-design should be prioritized or what aspects need to be studied in more depth. The scope and severity dimensions provide some recommendations for how to reduce the disconfirmed expectations in the chosen

scenario as well as some insights into how and why the users altered or changed their expectations.

We identified two global UX problems with high severity. The first is a lot of times the users utterances are not recognized by Pepper causing a bad UX. The other one is that the dialogues are usually experienced too simple and superficial. A local problem of medium severity is that the robot was unable to respond to requests to perform certain actions and behaviors. The overall global UX problem with high severity and scope is the mixed messages perceived from the robot's appearance and behavior during human-robot interaction. The users seem to expect a human-like way of acting and interacting with the Pepper robot, although they grasp that it is a machine.

5 Discussion and Conclusion

Our findings show that UX goals 1 and 3 were not fulfilled, whereas UX goals 2 and 4 were partly fulfilled (Fig. 3). These obtained results are based on the majority of the users, however, it is worth noting that 3 out of the five had most UX goals fulfilled with successful interactions and overall positive UX. This shows that there is quite a drastic variation between users, with some having a more positive UX while other's having a more negative UX.

A major insight derived from our analysis is that a lot of changes, relating to the three factors of expectations, actually emerged during the first-hand encounter of interacting with a robot. To interact per se with the robot in real life seems to have a big impact on the users, whether it resulted in a positive or negative UX.

Our findings also showed that the interactions, generally, improved between the first and second interactions. Several users had better interactions the second time, and no users had a more positive UX in the first interaction compared to the second interaction. These results show the importance of studying expectations temporally, as it changes over time in human-robot-interaction. These results are in line with previous research on how expectations can change in HRI [7,21,27], although this research is still in its infancy.

Another interesting insight is that users seem to implicitly expect human-like behavior from the robot and subsequently experience disappointment when these expectations are not confirmed, relating to Olson et al.'s [20] dimension of explicitness of expectation. During the post-test interview, many users stated that they were not impressed and made it clear that it was a robot, but at the same time compared the robot's behavior to a human from an anthropomorphic perspective. As robots are indeed not actually human, this sets up for disconfirmed expectations and bad UX as robots cannot live up to these expectations. This anthropomorphic expectation did not appear to become stable during the two interactions, similar to the study by Paetzel [21] who saw that anthropomorphism became stable at the end of the second interaction. It is possible that this expectation dimension would be adjusted if the users had the chance to interact for longer time periods with the robot. Future research includes increasing the

interaction periods in order to uncover how expectations work on a deeper level, develop other kinds of scenarios, and use other robots than Pepper.

During our analysis, we found that the RAS questionnaire and its subscales were appropriate for several UX goals, covering more expectation factors than affect. For example, subscale 2's theme relates to the behavioral characteristics of the robots, which could also be used to measure UX goal 2 as unexpected behavior from the robot may put a strain on the cognitive processing by the users. Subscale 3's theme relates to discourse with robots, which could also be used to measure UX goal 4 as unexpected discourse from the robot may lead to less ease of conversation. This shows how interrelated the expectation factors are, and strengthens the argument to use data triangulation when analyzing users' expectations.

We also noticed during the post-test interview that several users admitted to have had *some* experience with Pepper before this study, usually non-interactive but having been presented with the robot in various contexts, despite they reported that they have no previous experience with robots.

We want to note that there is some discrepancy between our framework (presented in [24]) and the present study, as we added two more measures (three items for a closeness questionnaire [30], and one item asking for the perceived capability of the robot). These measures are not present in this work and will be presented, along with the other questionnaires, in future publications. With this work, we have started to disentangle how expectations work and affect users' experiences during human-robot interactions. More work needs to be done in order to validate the social robot expectations gap evaluation framework, including considering more of the metrics (e.g., length of conversations) from the original framework.

Our findings indicate that more aspects of Olsen's et al. [20] model should be incorporated into the future development of our framework. We are inclined to further investigate how the dimensions of expectations; certainty, accessibility, explicitness, and importance are aligned to the user experience [20]. In particular, we want to investigate and analyze accessibility, explicitness, and their relatedness in more detail. Accessibility denotes how easy it is to activate and use a certain expectation, which we suggest is partly involved in the initial user experience when users are interacting with a social robot first-hand. Explicitness denotes to what degree expectations are consciously generated. The perceived mixed messages between knowing that a robot is a machine, but still comparing the conversations with the robot with human-like interactions imply that there are some hidden expectations that effects these mixed messages due to users' non-existing or limited first-hand experiences of interacting with social robots.

To conclude, we would like to point out that studying the relationship between expectations and user experience is of major concern for future social HRI research since this kind of social interactive technology allows humans to become more socially situated in the world of artificial systems [29]. As we hopefully have highlighted in this paper, investigating and analyzing how humans' expectations in interacting with social robots affect user experience may provide

additional significant insights concerning the fundamentals of human-human interaction. It is in relationship to something more familiar, as social robots, that the unknown becomes visible. Thus, by studying human-like robots, albeit machines, we learn more about ourselves as humans.

Acknowledgements. We would like to give a big thanks to the users that participated in this study, including Erik Lagerstedt who agreed to be depicted in Fig. 2.

This study was submitted for ethical review to the Swedish Ethical Review Authority (#2022-02582-01, Linköping) and was found to not require ethical review under Swedish legislation (2003:615). There were no physical or mental health risks to the users, and they were informed of their tasks prior to receiving an informed consent form. All data have been de-identified during collection. No sensitive personal information was collected. Video recordings are stored locally on a computer that is password protected. These recordings are only available for the researchers that analyzed the data and will be deleted after publication.

References

1. OpenAI. https://openai.com/
2. Alač, M.: Social robots: things or agents? AI Soc. **31**(4), 519–535 (2016)
3. Alenljung, B., Lindblom, J., Andreasson, R., Ziemke, T.: User experience in social human-robot interaction. In: Rapid Automation: Concepts, Methodologies, Tools, and Applications, pp. 1468–1490. IGI Global (2019)
4. Billing, E., Rosén, J., Lamb, M.: Language models for human-robot interaction. In: Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI '23 Companion), 13–16 March 2023, Stockholm, Sweden. ACM, New York, NY, USA (2023). https://doi.org/10.1145/3568294.3580040
5. Breazeal, C., Dautenhahn, K., Kanda, T.: Social robotics. In: Siciliano, B., Khatib, O. (eds.) Springer Handbook of Robotics, pp. 1935–1972. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-32552-1_72
6. Dautenhahn, K.: Some brief thoughts on the past and future of human-robot interaction. ACM Trans. Hum.-Robot Interact. (THRI) **7**(1), 4 (2018). https://doi.org/10.1145/3209769, https://dl.acm.org/citation.cfm?id=3209769
7. Edwards, A., Edwards, C., Westerman, D., Spence, P.: Initial expectations, interactions, and beyond with social robots. Comput. Hum. Behav. **90**, 308–314 (2019)
8. Hartson, H., Pyla, P.: The UX Book. Morgan Kaufmann (2018)
9. Hassenzahl, M., Tractinsky, N.: User experience - a research agenda. Behav. Inf. Technol. **25**(2), 91–97 (2006)
10. Horstmann, A., Krämer, N.: Great expectations? Relation of previous experiences with social robots in real life or in the media and expectancies based on qualitative and quantitative assessment. Front. Psychol. **10**, 939 (2019)
11. Horstmann, A., Krämer, N.: When a robot violates expectations. In: Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, pp. 254–256 (2020)
12. Jokinen, K., Wilcock, G.: Expectations and first experience with a social robot. In: Proceedings of the 5th International Conference on Human Agent Interaction, pp. 511–515 (2017)
13. Kaasinen, E., Kymäläinen, T., Niemelä, M., Olsson, T., Kanerva, M., Ikonen, V.: A user-centric view of intelligent environments. Computers **2**(1), 1–33 (2013)

14. Kwon, M., Jung, M., Knepper, R.: Human expectations of social robots. In: 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 463–464. IEEE (2016)
15. Lindblom, J., Alenljung, B., Billing, E.: Evaluating the user experience of human-robot interaction. In: Lindblom, J., Alenljung, B., Billing, E. (eds.) Human-Robot Interaction, pp. 231–256, vol. 12. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-42307-0_9
16. Lohse, M.: The role of expectations in HRI. New Front. Hum.-Robot Interact. 35–56 (2009)
17. Mahdi, H., Akgun, S.A., Saleh, S., Dautenhahn, K.: A survey on the design and evolution of social robots-past, present and future. Robot. Auton. Syst. 104193 (2022)
18. Nomura, T., Kanda, T., Suzuki, T., Kato, K.: Psychology in human-robot communication. In: RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No. 04TH8759), pp. 35–40. IEEE (2004)
19. Nomura, T., Suzuki, T., Kanda, T., Kato, K.: Measurement of anxiety toward robots. In: ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication, pp. 372–377. IEEE (2006)
20. Olson, J., Roese, N., Zanna, M.: Expectancies, pp. 211–238. Guilford Press (1996)
21. Paetzel, M., Perugia, G., Castellano, G.: The persistence of first impressions: the effect of repeated interactions on the perception of a social robot. In: Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, pp. 73–82 (2020)
22. Patton, M.Q.: Qualitative Research & Evaluation Methods: Integrating Theory and Practice. Sage Publications, Thousand Oaks (2014)
23. Rosén, J.: Expectations in human-robot interaction. In: Ayaz, H., Asgher, U., Paletta, L. (eds.) AHFE 2021. LNNS, vol. 259, pp. 98–105. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-80285-1_12
24. Rosén, J., Lindblom, J., Billing, E.: The social robot expectation gap evaluation framework. In: Kurosu, M. (eds.) HCII 2022. LNCS, vol. 13303, pp. 590–610. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-05409-9_43
25. Roto, V., Law, E., Vermeeren, A., Hoonhout, J.: User experience white paper - bringing clarity to the concept of user experience. In Dagstuhl Seminar on Demarcating User Experience (2011)
26. Sandoval, E.B., Mubin, O., Obaid, M.: Human robot interaction and fiction: a contradiction. In: Beetz, M., Johnston, B., Williams, M.-A. (eds.) ICSR 2014. LNCS (LNAI), vol. 8755, pp. 54–63. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11973-1_6
27. Serholt, S., Barendregt, W.: Robots tutoring children: longitudinal evaluation of social engagement in child-robot interaction. In: Proceedings of the 9th Nordic Conference on Human-Computer Interaction, pp. 1–10 (2016)
28. SoftBank Robotics: https://www.softbankrobotics.com (2018). Accessed 05 Jan 2018
29. Stephanidis, C., et al.: Seven HCI grand challenges. Int. J. Hum.-Comput. Interact. **7318** (2019)
30. Woosnam, K.M., et al.: The inclusion of other in the self (iOS) scale. Ann. Tour. Res. **37**(3), 857–860 (2010)

PAPER V

INVESTIGATING NARS

Reprinted from Julia Rosén, Erik Lagerstedt, and Maurice Lamb (2023). “Investigating NARS: Inconsistent Practice of Application and Reporting”. In: *The 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2023), Busan, South Korea, 2023*. IEEE, pp. 922–927 with permission from IEEE.

2023 32nd IEEE International Conference on Robot and
Human Interactive Communication (RO-MAN)
August 28-31, 2023. Paradise Hotel, Busan, Korea

2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) | 979-8-3503-3670-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/RO-MAN57019.2023.10309650

Investigating NARS: Inconsistent practice of application and reporting

Julia Rosén[1,*], Erik Lagerstedt[1,*] and Maurice Lamb[1]

***Abstract*— The Negative Attitude toward Robots Scale (NARS) is one of the most common questionnaires used in the studies of human-robot interaction (HRI). It was established in 2004, and has since then been used in several domains to measure attitudes, both as main results and as a potential confounding factor. To better understand this important tool of HRI research, we reviewed the HRI literature with a specific focus on practice and reporting related to NARS. We found that the use of NARS is being increasingly reported, and that there is a large variation in how NARS is applied. The reporting is, however, often not done in sufficient detail, meaning that NARS results are often difficult to interpret, and comparing between studies or performing meta-analyses are even more difficult. After providing an overview of the current state of NARS in HRI, we conclude with reflections and recommendations on the practices and reporting of NARS.**

I. INTRODUCTION

Questionnaires of various kinds are commonly used in human-robot interaction (HRI) research to study various phenomena [1], and the Godspeed Questionnaire Series (GQS), the NASA Task Load Index (NASA-TLX), and the Negative Attitude toward Robots Scale (NARS) are the three most common named questionnaires in the field [2]. The NASA-TLX [3] was developed in the late 1980s and has been used in different domains where cognitive load is a relevant factor to measure. In contrast, both GQS [4] and NARS [5] are developed specifically for the field of HRI. More specifically, GQS is used to evaluate a specific robot or situation, whereas NARS is used to measure general attitudes toward robots.

To be able to appropriately interpret measurements, it is necessary to understand the tools used to measure, and this is also the case when the tools in question are questionnaires. The practices surrounding the GQS has been examined for that purpose [6]. At the time, GQS was cited roughly 160 times according to Google Scholar [6], but has now more than ten times that amount. NARS has also been investigated in a similar way [7], however, that was at an initial stage when the use of NARS had only been reported tens of times. NARS is now cited hundreds of times, which is large enough to provide a good sample, but small enough to allow for adjustments to budding conventions before they become too widespread.

We initially set out to perform a meta-analysis based on the reported NARS data in conjunction with the examination of practices regarding NARS. Such secondary analysis of already published data can be an important step to identify general trends or phenomena not considered in the individual or initial studies [8]. However, we found that NARS is not applied consistently, nor is it reported in sufficient detail to allow for such an analysis (a problem identified also for other scales used in HRI [6], [9]). For that reason we decided on investigating the following:

1) Where are articles using NARS published?
2) How are methods involving NARS reported?
3) How is NARS data analyzed?
4) How are results related to NARS reported?

II. THE NEGATIVE ATTITUDES TOWARDS ROBOTS SCALE — NARS

NARS is a kind of Likert scale [10] and was introduced specifically to investigate both the short- and long-term affects of attitudes (such as anxiety) toward computers in education and therapy, though the authors also proposed that scale could be applied more generally [5]. Since its introduction, NARS has been used in a variety of contexts extending beyond its original use. The original scale was established in Japanese [5], but an English translation has been validated as well [11], [7].

The original formulation of NARS consists of fourteen statements (or Likert items) divided into three sub-scales, each related to a different theme of attitudes. Attitudes is, according to the authors, a psychological construct that "is defined as mental states prepared before behaviors" [5, p.35]. The first sub-scale, S1, is named "Negative Attitude toward Situations of Interaction with Robots" and the summary assessment range for this sub-scale is from 6–30. The second sub-scale, S2 or "Negative Attitude toward Social Influence of Robots", ranges from 5–25. The third sub-scale, S3 or "Negative Attitude toward Emotions in Interaction with Robots" ranges from 3–15. For each statement, the respondent indicates how strongly they agree using one out of five different response options, from "I strongly disagree" (coded numerically as "1") to "I strongly agree" (coded numerically as "5"). Three of the fourteen items are negatively formulated, and all of those items belong to the third sub-scale. A participant's result consists of three values which are calculated as the sum of the participant's responses to each of the sub-scales. Calculating a single NARS number by combining the three sub-scales is not meaningful since the sub-scales are designed to measure different kinds of attitudes.

The appropriate way of analyzing Likert scales is an ongoing discussion in scientific discourse in general [12],

[1]Interaction Lab, School of Informatics, University of Skövde, 54955 Skövde, Sweden [julia.rosen, erik.lagerstedt, maurice.lamb] @his.se

*Both authors contributed equally to this research.

[13], but also in HRI in particular [14], [9]. Typically, the recommendation is to not analyze individual Likert items (that is, the responses to the individual statements), but instead consider the aggregated result for each sub-scale. By aggregating the responses to the individual Likert items (which are a kind of ordinal data), the resulting numbers can arguably be analyzed as parametric data (for instance allowing for ANOVAs and t-tests). In its initial presentation [5], NARS was administered to 238 Japanese university students who were also asked about prior experience with real robots. A two-way ANOVA indicated that both gender and prior experience with robots affected the responses to sub-scale S1.

There are some aspects of NARS to be aware of when using or interpreting the scale. For instance, the statements in the scale are simply about "robots", which is a very broad term that can refer to a diverse range of machines [15]. The responses to the questionnaire may therefore depend on the participants' assumptions and expectations in relation to what kind of machine is referred to. Such assumptions might be biased by geographical or cultural variations [15]. In addition, there are indications that the interpretation of the statements can be sensitive to priming effects by, for instance, presenting participants with unrelated pictures of robots while filling out the questionnaire [16].

Another aspect to be aware of is that NARS measures the degree of *negative* attitudes toward robots respondents harbor. The lack of negative attitudes does not necessarily mean that there are positive attitude, but could simply be indifference. Also a participant with strong negative attitudes could overall have an ambiguous attitude toward robots if they simultaneously have strong positive attitudes. It is not a problem that NARS is focused specifically on negative attitudes, but it is important to be aware of to avoid conflating the concepts. The inability to distinguish between neutral (no strong feelings) and ambiguous (strong mixed feelings) attitudes is, however, far from a unique property of NARS [17].

To get a better understanding of the actual practice in relation to NARS, how it is used and reported, we conducted a literature review. Apart from providing insights regarding methodological practices in HRI in general, it can more specifically inform the development of new conventions regarding NARS.

III. METHOD

To find the publications in which NARS results are reported, we searched for the presence of the term "NARS" or (inclusive) "Negative Attitudes toward Robots" anywhere in the publication using the databases IEEE Xplore and ACM Digital Library. We only considered papers in scientific conferences, journals, and book chapters, excluding publication types like magazines, books, and standards. Since the original paper that introduced NARS was published in 2004, only results from that year and up until (and including) 2021 were analyzed. The initial search returned 380 papers, and after removing 28 duplicate papers (found in both databases), 352 papers remained. Further, 192 other publications unrelated to robotics were removed after manual inspection, resulting in a final corpus of 160 documents. We did not specifically rely on articles citing the original paper (although many papers in our corpus do), to include publications using NARS without reference or using some other paper as the source (e.g., [18]).

The review of the corpus was done in two rounds. In the first round, each document was classified depending on the reason for mentioning NARS; (1) because NARS was used in some way in an empirical study, (2) NARS was mentioned as background or context, or (3) NARS was used as a reference or starting point when creating a new scale. In addition, it was in this round noted whether the data in the respective publication was presented in a way that would allow a meta-analysis and if the scale had been modified. During the second round of reviews, more detailed information regarding practice and reporting was extracted. This included how data was presented, what statistical tests were used, what other methods or questionnaires were used in conjunction with NARS, the number of response options participants could choose from, and how the raw data was processed to get the NARS numbers.

Two researchers (first authors of this article) performed all the reviews. Each researcher reviewed half of the corpus for the first round, and swapped papers for the second round. Each paper was therefore reviewed by both researchers. Before each round, 10 papers were randomly selected and reviewed by both researchers to find consensus regarding interpretations of the classifications.

IV. RESULTS

We distinguish between actually using NARS as a measure, mentioning NARS as background or for context, and using NARS as a reference point when creating a novel survey. The most common use of NARS, in 115 of the 160 reviewed papers (72%), was as a measure of attitudes in the respective studies (see Figure 1). In 29 of the 160 reviewed papers (18%), NARS was mentioned as an example or discussed as background but not actually used. In the remaining 16 of the 160 papers (10%), NARS was instead used as a benchmark, inspiration, or basis for the creation of a new scale. The rest of this results section is organized into two parts. In the first part, patterns regarding when and where NARS shows up in the literature; all 160 publications are used for this. The second part of the results contains more specific details regarding how NARS was presented, used, and analyzed, based on the subset of 115 publications where NARS was actually used as a measure.

A. NARS in the literature

NARS was introduced in 2004 and initially the scale was rarely mentioned or used (see Figure 2). Within a few years, NARS started to gain more attention in the literature, and the number of mentions are still generally increasing. Assuming a linear increase over the last 10 years (2012–2021) results in a Pearson correlation coefficient of $r = 0.832$, and the

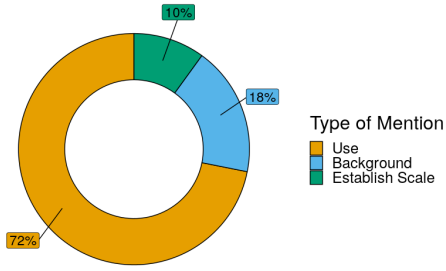

Fig. 1. The proportions of different purposes of mentioning NARS in the publications.

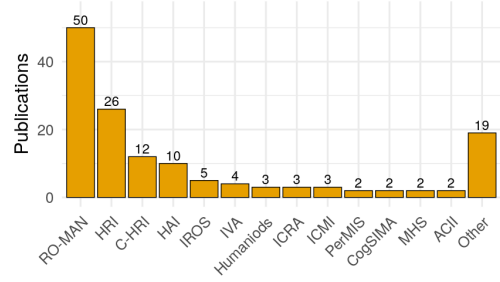

Fig. 3. Number of publications mentioning NARS in the most common conferences since its introduction in 2004. C-HRI is the companion of the HRI conference, and mainly contains late breaking reports. All 19 conferences in the "Other"-category had one instance each.

least square approximation for the same interval results in a regression slope of 2.3 publications per year.

Of the 160 reviewed articles, 143 articles (90%) were published in conference proceedings, and the rest appeared in scientific journals (15 articles, 9%) or book chapters (2 articles, 1%). For the conference papers, the most common conferences were IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN) with 50 papers, followed by ACM/IEEE International Conference on Human-Robot Interaction (HRI) with 38 papers of which 12 appeared in the compendium proceedings (see Figure 3). The remaining 55 conference papers were published across 29 other conferences, mainly focused on topics such as robotics, computing, or interaction but some with a slightly broader scope of technology.

In 108 of the 160 articles mentioning NARS (68%), other questionnaires were also mentioned by name. In addition to the named questionnaires, it was common to also ask about demographic information, or adding additional questions related to specific research questions or contexts. The most commonly mentioned additional questionnaires included Godspeed (25 papers, 12.5%), Big 5 (11 papers, 5.5%), RAS (10 papers, 5%), and RoSAS (9 papers, 4.5%) where the reported percentages are relative to the 108 papers identifying questionnaire(s) in addition to NARS (see Figure 4).

B. NARS when used

Looking specifically at the 115 articles reporting actual use of NARS, the work of identifying trends and practices was made difficult by the inconsistent and sparse reporting of relevant information. In the first round of reviews we identified two particular aspects of NARS that seemed inconsistent in how it was applied; (1) the *number of response options* participants had for each Likert item, and (2) what *item evaluation method* was used by the researchers when preparing the data for analysis. We found that most papers reported either both (39 papers, 34%) or neither (44 papers, 38%). Twenty papers (17%) reported only the *item evaluation method*, and 12 papers (10%) reported only the *number of response options* (see Figure 5). Thirteen of the

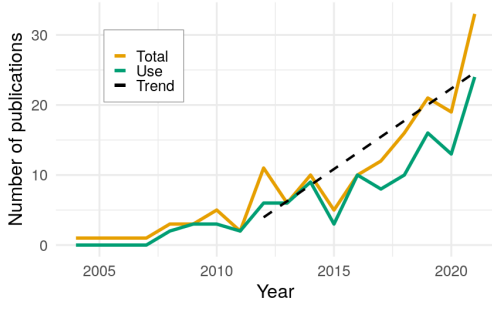

Fig. 2. Number of publications per year mentioning NARS since its introduction in 2004. "Total" is the total number of publications mentioning NARS each year, and "Use" is the subset of the publications that actually used NARS. The dashed line is the linear regression over the last 10 years.

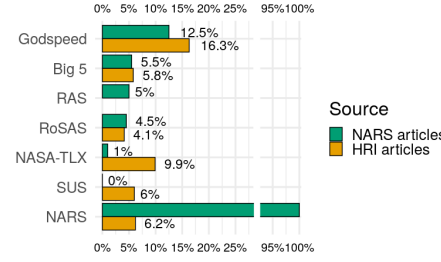

Fig. 4. Prevalence of most common scales used together with NARS, presented together with the most common scales found in 2022 by Zimmerman et al. [2] in their survey of the HRI field. Note the broken axis, and that NARS is found in 100% of the papers in our corpus due to the selection criteria of our survey.

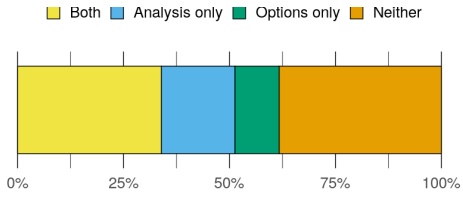

Fig. 5. The distribution of the 115 papers using NARS that report both *number of response options* and *item evaluation method*, neither, or only one of the categories.

39 papers reporting both the *number of response options* and *item evaluation method* (33%, 11% of the 115 papers using NARS) followed the original conventions [5] both with regards to the number of response options (5 options) and item evaluation method (the sum of responses to the Likert items for each sub-scale). In 12 of the 39 papers (31%), the conventional five options were used but together with an unconventional method of evaluation, and in 14 papers (36%) neither convention was observed. No paper reported the conventional evaluation method with a different number of response options.

When it comes to practices of reporting the NARS results, of the 115 papers using NARS, 30 papers (26%) did not present any numbers related to NARS, and 35 papers (30%) only presented the numerical results of NARS through some kind of statistical test. In the 50 remaining papers (43%) some reportings of NARS results were done, but only 14 of these papers (12% of the papers reported using NARS) included mean, standard deviation, and number of respondents, which would be necessary for meta-analysis [19].

The original formulation of NARS specified 5 response options for each statement in the scale [5]. In the papers using NARS ranges were reported anywhere from binary to 8 possible responses. In 40 papers (35%) it was possible to identify that five options were used, whereas it was not possible (from the provided text) in 56 papers (49%) to identify how many response options existed for each item. In 14 papers (12%) it was possible to identify that seven options were used, and the remaining 5 papers (4%) used some other range of options, all of which had an even number of options (see Figure 6).

To calculate the score of each sub-scale for a respondent, the sum of the responses to the corresponding Likert items are calculated. Sixty-four papers (56%) did not report how the results for the sub-scales are calculated, nor is it possible to make any conclusive inferences based on reported numbers (see Figure 7). Of the 51 papers (44%) where it is possible to identify the method, only 15 papers (29%, 13% of all papers using NARS) used the sum (as originally intended [5]), whereas 35 papers (69%, 30% of all papers using NARS) instead reported calculating the mean item response for the respective sub-scales. One paper (2%, 1% of all papers using NARS) reported analyzing the response data at an individual item level.

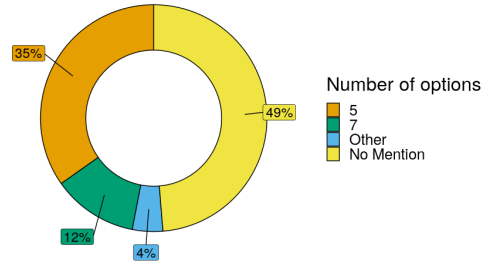

Fig. 6. The number of possible response options when using NARS (validated with 5 options [5]). The "other" category captures the instances when 2, 4, 6 or 8 options are used.

Papers that used NARS often reported statistical tests. In 107 papers (93%), some kind of statistical test was reported, however, only 76 papers (66%) report statistical tests in relation to NARS. In 4 of those papers, the results of the statistical tests were presented without any reference to what kind of test was made. The most popular statistical tests including NARS data were ANOVAs (and to a lesser degree ANCOVA and MANOVA), various kinds of regression analysis, and Student's t-tests.

As a final note on the use of statistics, there were 16 papers among the reviewed 160 (10%) that created new scales. Among those 16 papers, 8 papers (50%) reported a measure of internal consistency (in relation to some other measure of the respective phenomena of interest), primarily through Cronbach's alpha. Two of the 16 papers (13%) reported that tests of internal consistency had been used, but no further details were reported. The remaining 6 papers (38%) did not report any internal consistency or reliability of the items. Among the 115 papers that used NARS, 9 papers (8%) reported Cronbach's alpha to confirm the internal consistency of the scale.

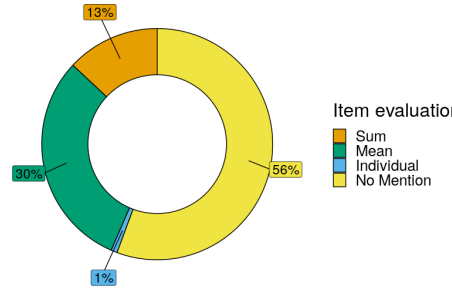

Fig. 7. The analysis practice when calculating the score for the respective sub-scale for each participant (validated using summation [5]).

V. DISCUSSION

Despite not being able to perform the initially intended meta-analysis, we did identify three general patterns; (1) the reported use of NARS is increasing, (2) the specifics of how it is used varies, and (3) so does the level of detail in the reporting. After a brief discussion on each of these respective findings, we will conclude with a general discussion and recommendations.

A. NARS reporting is increasing

We found that the number of new publications per year mentioning or using NARS is increasing. It is, however, likely that the linear regression is on the overly-conservative side, and the growth might be faster than linear. GQS had, for instance, roughly 160 citations six years after its publication, but has now (another eight years later) more than ten times the citations. The rate of growth is slower for NARS compared to GQS, evident by NARS being slightly older while still being cited fewer times, but the number of annual publications using NARS is growing nonetheless. We had expected a larger total number of publications using NARS, given how well known that questionnaire is in the field and that general attitudes toward robots could arguably be a relevant confound in most HRI and social robotics. The relatively low number of citations could potentially be explained by null results often remaining unreported, however, such a hypothesis is difficult to examine. In the reviewed papers, it was not unusual to mention that NARS was used, but not report any results related to NARS, which could be seen as weak support for such a hypothesis. There are, however, other reasons for such reporting, such as studies with several results might spread the reporting of the result over several publications. The absence of evidence is after all not evidence of the absent.

B. NARS is not applied consistently

NARS was used in many different kinds of studies, often complementing other methods and techniques. This is one of the strengths of NARS; to get a quick and comparable measure of the participants' attitudes toward robots. Such measure can be used as a basis for a result in its own right, or simply to examine a potential confound. However, NARS was established and validated using a set of 14 different Likert items, phrased in a particular way. The scale was also established and validated using five response options for each item, and each of the three sub-scales are to be analyzed by summarizing the responses to the related items. Despite that, we saw a large variation in how NARS was used, such as removing, adding, and rephrasing Likert items of the scale, changing the number of response options, and changing way to analyze the response options. This was often done without validating the new version of the scale. Without a new validation, it is difficult to tell to what extent the results of the questionnaire relates to the actual phenomenon of interest [1]. Changing too many things will also make it difficult to determine how the NARS results of one study compares to the NARS results of another.

C. NARS is not always reported in sufficient detail

There were also many unreported aspects related to the questionnaire. It was often difficult to tell how and when the questionnaires were provided to the participants, how the data was processed and analyzed, and what the analysis resulted in (including how it was interpreted). For some of those aspects, such as number of response options and phrasing of Likert items, there are conventions established with the validation [5], and it is reasonable to assume that those conventions are followed unless explicitly stated otherwise. We did, however, find several papers where the reported NARS numbers were not possible given the conventional methods, so there is always the risk that such assumptions are wrong. Again, the absence of evidence is not evidence of the absent, so not reporting certain practices is not conclusive evidence that they were not followed.

Other aspects are not dictated by the validation of the scale. Some aspects were, for instance, related to the particularities of the specific study, such as what statistical tests to perform or how to integrate the questionnaire with other tasks and measures. Providing such details in each respective case is important to make it possible to understand the context in which the scale is used, which is particularly important given indications that the measure might be sensitive to some priming [16], [15].

D. General discussion and recommendations

Since NARS is such a popular and well known tool for researchers in HRI while still having a manageable amount of published work (similar to GQS when examined by [6]), this is an opportune time to learn from the practices regarding NARS so far, and identify and adopt appropriate conventions moving forward. Although some conventions are already established in relation to NARS, there might be reasons to change some of them. Changes might be due to particularities of a specific study, or as some general improvement of the tool. An example of the former is when the phrasing of the Likert items needs to change to be understandable for the participants. It might also be necessary to translate it to a different language (e.g., Turkish [20]) or make minor adjustments to make the statements comprehensible to specific groups (e.g, children [21]). When doing so, it is necessary to validate the questionnaire in the new format to be sure that the same things are measured. Being clear about precisely what has changed as well as how (e.g., when adding Likert items [7]), is also necessary to allow the results to be appropriately interpreted. The necessity of clarity and motivation is also the case for the other established conventions (such as number of response options and summing the responses to items before analyzing the scales). Alternative conventions might be introduced, however such cases may require additional validation, and should be explicitly stated to avoid inappropriate comparisons to previously published results. When conventions are followed, the source of the conventions should be cited (e.g., [5]).

In terms of less established methodological aspects, factors such as when (and under what circumstances) in the study

NARS is distributed would often be helpful when interpreting the results. Documenting the results and experiences regarding this decisions would also be an important resource when making decisions in future studies, and in establishing and improving methodological conventions for the field of HRI (proposed by e.g., [7]).

A surprising amount of papers did not report any results at all regarding NARS, and a reason for that might be that NARS was not the only, nor the main, measure of the respective studies. Although it is understandable to omit results in such cases, it is still preferable with reported results for NARS as well, even if it is just a simple statement that no significant results were found, or explaining why those results are omitted. When actually reporting results, it is important that it is clear what condition or sub-group the results refer to as well as the number of participants in the respective groups (which is not always clear in the reviewed literature). For each sub-scale, the mean and standard deviation should be reported to facilitate interpretation and comparisons. If statistical tests are conducted, it should be clear which tests are used, and relevant statistics should be reported.

VI. CONCLUSIONS

We have investigated the current practices related to the questionnaire NARS through a literature review. The reported use of the scale is increasing, and it is evident that there is considerable variability in the use and reporting of NARS in HRI research. Moreover, the lack of standard practices only amplifies the difficulties related to comparing findings of different studies. There are several reasons for making changes to the scale and how it is applied, however, such revisions need to be validated to ensure that the improved version is still measuring what is intended. Some of the changes might only be relevant in specific contexts, but there might also be a need for a general revision of NARS and its related conventions. Clear, unambiguous reporting of the current practices would provide a valuable empirical resource for such work. In terms of using NARS as a measure, relying on current conventions is important to facilitate interpretation of the results in respective studies, as well as making it possible to compare results between studies. Reporting in a way that would also allow for meta-analyses would make it possible to build on the collective work of the field to examine more general phenomena related to attitudes and robots. Clear and unambiguous reporting is an important step for that reason as well.

REFERENCES

[1] M. Rueben, S. A. Elprama, D. Chrysostomou, and A. Jacobs, "Introduction to (re)using questionnaires in human-robot interaction research," in *Human-Robot Interaction: Evaluation Methods and Their Standardization*. Springer, 2020, pp. 125–144.

[2] M. Zimmerman, S. Bagchi, J. Marvel, and V. Nguyen, "An analysis of metrics and methods in research from human-robot interaction conferences, 2015-2021," in *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '22. IEEE Press, 2022, p. 644–648.

[3] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," in *Advances in psychology*. Elsevier, 1988, vol. 52, pp. 139–183.

[4] 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*, vol. 1, pp. 71–81, 2009.

[5] T. Nomura, T. Kanda, T. Suzuki, and K. Kato, "Psychology in human-robot communication: An attempt through investigation of negative attitudes and anxiety toward robots," in *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No. 04TH8759)*. IEEE, 2004, pp. 35–40.

[6] A. Weiss and C. Bartneck, "Meta analysis of the usage of the godspeed questionnaire series," in *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2015, pp. 381–388.

[7] K. M. Tsui, M. Desai, H. A. Yanco, H. Cramer, and N. Kemper, "Using the "Negative Attitude Toward Robots Scale" with telepresence robots," in *Proceedings of the 10th performance metrics for intelligent systems workshop*, 2010, pp. 243–250.

[8] G. V. Glass, "Primary, secondary, and meta-analysis of research," *Educational researcher*, vol. 5, no. 10, pp. 3–8, 1976.

[9] M. L. Schrum, M. Johnson, M. Ghuy, and M. C. Gombolay, "Four years in review: Statistical practices of likert scales in human-robot interaction studies," in *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 2020, pp. 43–52.

[10] R. Likert, "A technique for the measurement of attitudes." *Archives of psychology*, vol. 22, no. 140, pp. 5–55, 1932.

[11] T. Nomura, T. Suzuki, T. Kanda, and K. Kato, "Measurement of negative attitudes toward robots," *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, vol. 7, no. 3, pp. 437–454, 2006.

[12] S. E. Harpe, "How to analyze likert and other rating scale data," *Currents in pharmacy teaching and learning*, vol. 7, no. 6, pp. 836–850, 2015.

[13] J. Murray, "Likert data: what to use, parametric or non-parametric?" *International Journal of Business and Social Science*, vol. 4, no. 11, 2013.

[14] M. Gombolay and A. Shah, "Appraisal of statistical practices in HRI vis-á-vis the t-test for likert items/scales," in *2016 AAAI Fall Symposium Series*, 2016, pp. 46–54.

[15] T. Nomura, T. Kanda, T. Suzuki, and K. Kato, "People's assumptions about robots: Investigation of their relationships with attitudes and emotions toward robots," in *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.* IEEE, 2005, pp. 125–130.

[16] S. Thellman and T. Ziemke, "Social attitudes toward robots are easily manipulated," in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 2017, pp. 299–300.

[17] J. G. Stapels and F. Eyssel, "Let's not be indifferent about robots: Neutral ratings on bipolar measures mask ambivalence in attitudes towards robots," *PloS one*, vol. 16, no. 1, p. e0244697, 2021.

[18] T. Nomura, T. Kanda, and T. Suzuki, "Experimental investigation into influence of negative attitudes toward robots on human–robot interaction," *Ai & Society*, vol. 20, pp. 138–150, 2006.

[19] T. Li, J. Higgins, J. Deeks, J.J. Deeks, J. Higgins, and D. Altman, "Collecting data," in *Cochrane Handbook for Systematic Reviews of Interventions*, J. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. Page, and V. Welch, Eds. Cochrane, 2022, ch. 5, version 6.3 (updated February 2022), Available from www.training.cochrane.org/handbook.

[20] J. Kanero, I. Franko, C. Oranç, O. Uluşahin, S. Koşkulu, Z. Adıgüzel, A. C. Küntay, and T. Göksun, "Who can benefit from robots? effects of individual differences in robot-assisted language learning," in *2018 Joint IEEE 8th International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*. IEEE, 2018, pp. 212–217.

[21] D. Robert and V. van den Bergh, "Children's openness to interacting with a robot scale (COIRS)," in *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 2014, pp. 930–935.

PAPER VI

PREVIOUS EXPERIENCE MATTERS

Julia Rosén et al. (Under review). “Previous Experience Matters: An In-Person Investigation of Expectations in Human-Robot Interaction”. In: *Under review for scientific journal*, pp. 1–19

Previous experience matters: an in-person investigation of expectations in human-robot interaction

Julia Rosén[1*], Jessica Lindblom[1,2], Maurice Lamb[1] and Erik Billing[1]

[1*]School of Informatics, University of Skövde, Högskolevägen 1, Skövde, 541 28, Sweden.
[2]Department of Information Technology, Uppsala University, Lägerhyddsvägen 1, Uppsala, 751 05, Sweden.

*Corresponding author(s). E-mail(s): julia.rosen@his.se;
Contributing authors: jessica.lindblom@his.se; maurice.lamb@his.se; erik.billing@his.se;

Abstract

The human-robot interaction (HRI) field goes beyond the mere technical aspects of developing robots, often investigating how humans perceive robots. Human perceptions and behavior are determined, in part, by expectations. Given the impact of expectations on behavior, it is important to understand what expectations individuals bring into HRI settings and how those expectations may affect their interactions with the robot over time. For many people, social robots are not a common part of their experiences, thus any expectations they have of social robots are likely shaped by other sources. As a result, individual expectations coming into HRI settings may be highly variable. Although there has been some recent interest in expectations within the field, there is an overall lack of empirical investigation into its impacts on HRI, especially in-person robot interactions. To this end, a within-subject in-person study (N=31) was performed where participants were instructed to engage in open conversation with the social robot Pepper during two 2.5 minute sessions. The robot was equipped with a custom dialogue system based on the GPT-3 large language model, allowing autonomous responses to verbal input. Participants' affective changes towards the robot were assessed using three questionnaires, NARS, RAS, commonly used in HRI studies, and Closeness, based on the IOS scale. In addition to the three standard questionnaires, a custom question was administered to capture participants' views on robot capabilities. All measures were collected three times, before the interaction with the robot, after the first interaction with the robot, and after the second interaction with the robot. Results revealed that participants to large degrees stayed with the expectations they had coming into the study, and in contrast to our hypothesis, none of the measured scales moved towards a common mean. Moreover, previous experience with robots was revealed to be a major factor of how participants experienced the robot in the study. These results could be interpreted as implying that expectations of robots are to large degrees decided before interactions with the robot, and that these expectations do not necessarily change as a result of the interaction. Results reveal a strong connection to how expectations are studied in social psychology and human-human interaction, underpinning its relevance for HRI research.

Keywords: Expectations, previous experience, social robot, human-robot interaction, experiment, expectation gap, pepper, GPT, large language models

PAPER VII

DISENTANGLING PEOPLE'S EXPERIENCES AND EXPECTATIONS WHEN INTERACTING WITH THE SOCIAL ROBOT PEPPER

Jessica Lindblom et al. (Manuscript). "Disentangling People's Experiences and Expectations when Interacting with the Social Robot Pepper: A Qualitative Analysis". In: *Manuscript for scientific journal*, pp. 1–41

Disentangling People's Experiences and Expectations when Interacting with the Social Robot Pepper: A Qualitative Analysis

Jessica Lindblom[1,2*], Julia Rosén[1], Maurice Lamb[1] and Erik Billing[1]

[1*]School of Informatics, University of Skövde, Högskolevägen 1, Skövde, 541 28, Sweden.
[2*]Department of Information Technology, Uppsala University, Lägerhyddsvägen 1, Uppsala, 751 05, Sweden.

*Corresponding author(s). E-mail(s): jessica.lindblom@his.se, jessica.lindblom@it.uu.se;
Contributing authors: julia.rosen@his.se; maurice.lamb@his.se; erik.billing@his.se;

Abstract

The use of social robots in many sectors of society is supposed to progressively increase, although ordinary people are still not that familiar with interacting first-hand with these robots. Humans are experts at interacting socially, being capable of interacting with a diverse array of social actions and interactions. Within social Human-Robot Interaction (sHRI), social robots are purposely designed to have the appearance of agency that encourages users to interact and communicate with them in socially appropriate ways that resemble human social interaction and cognition. This technological development as well as the societal push towards applying social robots in various contexts, both domestic as well as professional ones, increases the expectations of such robots. Two conditions for this foreseen future are that social robots also need to achieve their intended benefits for human users, while at the same time being positively experienced by them. The study of users' expectations towards social robots is still an emerging area within the interdisciplinary sHRI field, and there is a need to narrow the knowledge gap of how expectations play a role in users' experiences when interacting with these robots over time. The overarching purpose of the present work is to disentangle how humans, as the users of robots, experience this specific kind of in-person interaction with social robots, but also to deepen the understanding regarding what aspects that influence their expectations over time. Qualitative data from interactions with the social robot Pepper, equipped with the OpenAI GPT-3 language model, were collected from an experiment. A reflexive thematic analysis (RTA) was applied to facilitate the analysis and to identify central themes. The major findings consist of various levels of interaction quality, different types of interaction strategies, a more nuanced picture of the social robot expectation gap, core elements that influenced the users' expectations and experiences, and positive and negative user experiences that vary along four dimensions. For a majority of the participants, the initial in-person encounter left a positive impression, which indicates that the robot surpassed their initial expectations. However, there were also negative user experiences, mainly due to a lack of proper verbal dialogue, and feelings of oddness and awkwardness of the current situation. These findings underscore the intricate nature of user expectations. Recognizing the importance of implicit expectations is critical in understanding users' experiences, particularly as the participants appeared to encounter challenges in expressing their explicit expectations. Moreover, it became obvious that the participants adapted their interactions with the robot based on their perceived capability of the robot, which shaped their experiences, revealing that positive user experience is not solely determined by the interaction quality. To conclude, there is an interplay among many aspects when interacting with a social robot, which makes it challenging to study certain aspects in isolation because the users' experiences of the sHRI are influenced by prior experiences as well as expectations, which depend on those experiences.

Keywords: Expectations, social robot, human-robot interaction, user experience, social robot expectation gap, The social robot Pepper, OpenAI GPT-3 language model

PUBLICATIONS IN THE DISSERTATION SERIES

PUBLICATIONS IN THE DISSERTATION SERIES

1. Berg Marklund, Björn (2013). Games in Formal Educational Settings: Obstacles for the development and use of learning games. Licentiate Dissertation, ISBN 978-91-981474-0-7.
2. Aslam, Tehseen (2013). Analysis of manufacturing supply chains using system dynamics and multi-objective optimization. Doctoral Dissertation, ISBN 978-91981474-1-4.
3. Laxhammar, Rikard (2014). Conformal anomaly detection: Detecting abnormal trajectories in surveillance applications. Doctoral Dissertation, ISBN 978-91-981474-2-1.
4. Alklind Taylor, Anna-Sofia (2014). Facilitation matters: A framework for instructor-led serious gaming. Doctoral Dissertation, ISBN 978-91-981474-4-5.
5. Holgersson, Jesper (2014). User Participation In Public e-service Development: Guidelines for including external users. Doctoral Dissertation, ISBN 978-91-981474-5-2.
6. Kaidalova, Julia (2015). Towards a definition of the role of enterprise modeling in the context of business and IT alignment. Licentiate Dissertation, ISBN 978-91-981474-6-9.
7. Rexhepi, Hanife (2015). Improving healthcare information systems: A key to evidence based medicine. Licentiate Thesis, ISBN 978-91-981474-7-6.
8. Berg Marklund, Björn (2015). Unpacking Digital Game-Based Learning: The complexities of developing and using educational games. Doctoral Dissertation, ISBN 978-91-981474-8-3.
9. Fornlöf, Veronica (2016). Improved remaining useful life estimations for on-condition parts in aircraft engines. Licentiate Thesis, ISBN 978-91-981474-9-0.

10. Ohlander, Ulrika (2016). Towards Enhanced Tactical Support Systems. Licentiate Thesis, ISBN 978-91-982690-0-0.
11. Siegmund, Florian (2016). Dynamic Resampling for Preference-based Evolutionary Multi-objective Optimization of Stochastic Systems: Improving the efficiency of time-constrained optimization. Doctoral Dissertation, ISBN 978-91-982690-1-7.
12. Kolbeinsson, Ari (2016). Managing Interruptions in Manufacturing: Towards a Theoretical Framework for Interruptions in Manufacturing Assembly. Licentiate Thesis, ISBN 978-91-982690-2-4.
13. Sigholm, Johan (2016). Secure Tactical Communications for Inter-Organizational Collaboration: The Role of Emerging Information and Communications Technology, Privacy Issues, and Cyber Threats on the Digital Battlefield. Licentiate Thesis, ISBN 978-91-982690-3-1.
14. Brolin, Anna (2016). An investigation of cognitive aspects affecting human performance in manual assembly. Doctoral Dissertation, ISBN 978-91-982690-4-8.
15. Brodin, Martin (2016). Mobile Device Strategy: A management framework for securing company information assets on mobile devices. Licentiate Thesis, ISBN 978-91-982690-5-5.
16. Ericson, Stefan (2017). Vision-Based Perception for Localization of Autonomous Agricultural Robots. Doctoral Dissertation, ISBN 978-91-982690-7-9.
17. Holm, Magnus (2017). Adaptive Decision Support for Shop-floor Operators using Function Blocks. Doctoral Dissertation, ISBN 978-91-982690-8-6.
18. Larsson, Carina (2017). Communicating performance measures: Supporting continuous improvement in manufacturing companies. Licentiate Thesis, ISBN 978-91-982690-9-3.
19. Rexhepi, Hanife (2018). Briding the Information Gap: Supporting Evidence-Based Medicine and Shared Decision-Making through Information Systems. Doctoral Dissertation, ISBN 978-91-984187-1-2.
20. Schmidt, Bernard (2018). Toward Predictive Maintenance in a Cloud Manufacturing Environment: A population-wide approach. Doctoral Dissertation, ISBN 978-91-984187-2-9.
21. Linnéusson, Gary (2018). Towards strategic development of maintenance and its effects on production performance: A hybrid simulation-based optimization framework. Doctoral Dissertation, ISBN 978-91-984187-3-6.
22. Amouzgar, Kaveh (2018). Metamodel Based Multi-Objective Optimization with Finite-Element Applications. Doctoral Dissertation, ISBN 978-91-984187-4-3.

23. Bernedixen, Jacob (2018). Automated Bottleneck Analysis of Production Systems: Increasing the applicability of simulation-based multi-objective optimization for bottleneck analysis within industry. Doctoral Dissertation, ISBN 978-91-984187-6-7.

24. Karlsson, Ingemar (2018). An Interactive Decision Support System Using Simulation-Based Optimization and Knowledge Extraction. Doctoral Dissertation, ISBN 978-91-984187-5-0.

25. Andersson, Martin (2018). A Bilevel Approach To Parameter Tuning of Optimization Algorithms Using Evolutionary Computing. Doctoral Dissertation, ISBN 978-91-984187-7-4.

26. Tavara, Shirin (2018). High-Performance Computing For Support Vector Machines. Licentiate Dissertation, ISBN 978-91-984187-8-1.

27. Bevilacqua, Fernando (2018). Game-calibrated and user-tailored remote detection of emotion: A non-intrusive, multifactorial camera-based approach for detecting stress and boredom of players in games. Doctoral Dissertation, ISBN 978-91-984187-9-8.

28. Kolbeinsson, Ari (2019). Situating interruptions in manufacturing assembly. Doctoral Dissertation, ISBN 978-91-984918-0-7.

29. Goienetxea Uriarte, Ainhoa (2019). Bringing Together Lean, Simulation and Optimization: Defining a framework to support decision-making in system design and improvement. Doctoral Dissertation, ISBN 978-91-984918-1-4.

30. Gudfinnsson, Kristens (2019). Towards facilitating BI adoption in small and medium sized manufacturing companies. Doctoral Dissertation, ISBN 978-91-984918-2-1.

31. Kaidalova, Julia (2019). Integration of Product-IT into Enterprise Architecture: a metod for participatory Business and IT Alignment. Doctoral Dissertation, ISBN 978-91-984918-3-8.

32. Brodin, Martin (2020). Managing information security for mobile devices in small and medium-sized enterprises: Information management, Information security management, mobile device. Doctoral Dissertation, ISBN 978-91-984918-4-5.

33. Bergström, Erik (2020). Supporting Information Security Management: Developing a Method for Information Classification. Doctoral Dissertation, ISBN 978-91-984918-5-2.

34. Gustavsson, Patrik (2020). Virtual Reality Platform for Design and Evaluation of the Interaction in Human-Robot Collaborative Tasks in Assembly Manufacturing. Doctoral Dissertation, ISBN 978-91-984918-6-9.

35. Ruiz Zúñiga, Enrique (2020). Facility layout design with simulation-based optimization: A holistic methodology including process, flow, and logistics requirements in manufacturing. Doctoral Dissertation, ISBN 978-91-984918-9-0.

36. Ventocilla, Elio (2021). Visualizing Cluster Patterns at Scale: A Model and a Library. Doctoral Dissertation, ISBN 978-91-984919-0-6.

37. Danielsson, Oscar (2020). Augmented reality smart glasses as assembly operator support: Towards a framework for enabling industrial integration. Licentiate Dissertation, ISBN 978-91-984919-1-3.

38. Morshedzadeh, Iman (2021). Managing virtual factory artifacts in extended product lifecycle management systems. Doctoral Dissertation, ISBN 978-91-984919-2-0.

39. Ståhl, Niclas (2021). Integrating domain knowledge into deep learning: Increasing model performance through human expertise. Doctoral Dissertation, ISBN , ISBN 978-91-984919-3-7.

40. Lidberg, Simon (2021). Evaluating fast and efficient modeling methods for simulation-based optimization. Licentiate Dissertation, ISBN 978-91-984919-4-4.

41. Senavirathne, Navoda (2021). Towards privacy preserving micro-data analysis – A machine learning based perspective under prevailing privacy regulations. Doctoral Dissertation, ISBN 978-91-984919-5-1.

42. Lennerholt, Christian (2022). Facilitating the implementation and use of self service business intelligence. Doctoral Dissertation, ISBN 978-91-984919-6-8.

43. Liu, Yu (2022). Integrating life cycle assessment into simulation-based decision support. Licentiate Dissertation, ISBN 978-91-984919-7-5.

44. Tavara, Shirin (2022). Distributed and federated learning of support vector machines and applications. Doctoral Dissertation, ISBN 978-91-984919-8-2.

45. Kävrestad, Joakim (2022). Context-Based Micro-Training – enhancing cybersecurity training for end-users. Doctoral Dissertation, ISBN 978-91-984919-9-9.

46. Jiang, Yuning (2022). Vulnerability Analysis for Critical Infrastructures. Doctoral Dissertation, ISBN 978-91-987906-0-3.

47. Toftedahl, Marcus (2022). Being Local in a Global Industry: Game Localization from an Indie Game Development Perspective. Doctoral Dissertation, ISBN 978-91-987906-1-0.

48. Danielsson, Oscar (2022). Augmented reality smart glasses as assembly operator support – A framework for enabling industrial integration. Doctoral Dissertation, ISBN 978-91-987906-2-7.

49. Su, Yanhui (2022). Bringing game analytics to indie game publishing – Method and tool support for indie mobile game publishing. Doctoral Dissertation, ISBN 978-91-987906-3-4.

50. Svensson, Torbjörn (2023). From Games to News – Creating an Engagement Model for Digital Local News. Doctoral Dissertation, ISBN 978-91-987906-4-1.

51. Barrera Diaz, Carlos Alberto (2023). Simulation-Based Multi-Objective Optimization for Reconfigurable Manufacturing Systems. Doctoral Dissertation, ISBN 978-91-987906-5-8.

52. Smedberg, Henrik (2023). Knowledge Discovery for Interactive Decision Support and Knowledge-Driven Optimization. Doctoral Dissertation, ISBN 978-91-987906-6-5.

53. Sweidan, Dirar (2023). Data-Driven Decision Support in Digital Retailing. Doctoral Dissertation, ISBN 978-91-987906-7-2.

54. Sahlin, Johannes (2023). Designing Advertising Systems with Human-Centered Artificial Intelligence. Doctoral Dissertation, ISBN 978-91-987906-8-9.

JULIA ROSÉN

Julia Rosén holds a BA in Psychology with an English Minor from Long Island University, USA, and an MA in Cognitive Science from Lund University, Sweden. Her academic interests include social robots, user-centered research, UX, ethics, inclusivity, and methodological practices.

In this thesis, Julia presents her work on users' expectations in social Human-Robot Interaction (sHRI). The result of her work can be divided into two major findings. The first finding is the Social Robot Expectation Gap Evaluation Framework. Focusing on the temporal aspect of expectations, the framework offers a strategy to systematically study expectations in sHRI. The second finding identifies the role and relevance of users' expectations, emphasizing that expectations will differ based on direct and indirect experiences with social robots. These results underscore the potential for expectations to act as a confounding variable in sHRI research. Moreover, these results demonstrate the need to consider individual expectations, how they change temporally, and their impact on user needs and preferences. By considering these findings, we can better manage expectations and, in doing so, reduce the social robot expectation gap.

ISBN 978-91-987906-9-6
Dissertation Series, No. 55 (2024)