



Previous Experience Matters: An in-Person Investigation of Expectations in Human–Robot Interaction

Julia Rosén¹ · Jessica Lindblom^{1,2} · Maurice Lamb¹ · Erik Billing¹

Accepted: 12 January 2024 / Published online: 29 February 2024
© The Author(s) 2024

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 min 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

✉ Julia Rosén
julia.rosen@his.se

Jessica Lindblom
jessica.lindblom@it.uu.se

Maurice Lamb
maurice.lamb@his.se

Erik Billing
erik.billing@his.se

¹ School of Informatics, University of Skövde, Höskolevägen 1, 541 28 Skövde, Sweden

² Department of Information Technology, Uppsala University, Lägerhyddsvägen 1, 751 05 Uppsala, Sweden

1 Introduction

The ways humans perceive situations and take action depends on their assumptions about the world and how various entities in the world, either objects or others, respond to one's actions [1–3]. These expectations and their emerging social interaction patterns are suggested to play major roles in shaping human-human interaction (HHI). It is proposed that humans rely on social expectations grounded in experiences from HHI when interacting with social robots, implying that humans tend to interact with these robots by using the same interaction patterns they developed to interact with other humans [4, 5]. However, it may also be the case that humans

have different social expectations about the behavior of social robots compared to humans [6]. In human–robot interaction (HRI) contexts, individuals begin interacting with a robot with some set of previous experience of what robots and artificial intelligence (AI), in general, can do.

Social robots, in particular, are becoming more sophisticated and stand out from other kinds of digital technologies as they occupy our physical and social space rather autonomously [7, 8]. Social robots are also designed to motivate humans to interact and communicate with these robots as they would with other people [9–11]. Thus, research conducted within the HRI field often assumes that HHI research is transferable to HRI, and many HRI studies are strongly inspired by the field of social psychology [12]. There is, however, a key difference between HHI and HRI, namely that social robots are not humans, despite considerable attempts to make it appear otherwise. Social robots are artifacts designed to be as human-like as technically feasible. As a result, the line between humans and artifacts is more arbitrary and blurred in social HRI than interactions with other kinds of technology [13]. The close connection between social robots and HHI means that individual response to and communication patterns with social robots are based on human expectations, both of technological artifacts and social agents. Variations in interactions with social robots can be based on a lack of understanding of the robot's behavioral, social, and cognitive capabilities, but also may arise from a mismatch between what we expect from robots compared to humans [4, 5, 14, 15]. As HRI researchers, it is challenging to control participant expectations because they may come from experiences other than first-hand interactions with robots, such as the exposition of robots in movies, books, or video games [4, 5, 14, 16]. Hence, it is important to examine individuals' expectations of social robots to further develop effective, smooth, and intuitive ways for humans to interact with robots so that robots can support the tasks as intended.

Expectations have previously been studied in HRI, and interest in the topic is steadily increasing [e.g., 4, 5, 10, 11, 14, 15, 17–25]. However, many studies of expectations in HRI are performed using images of robots, video clips of another individual interacting with a robot, or a human controlling the robot (i.e., the Wizard of Oz (WoZ) technique) as a surrogate. Hence, there is little insight into how expectations affect live interactions with a physical social robot in person.

Within social psychology, first-hand experiences are a distinctive source, which are often the basis for more accurate expectations than other experiences, such as watching videos or imagining an experience [1]. Notably, most individuals have either no first-hand experiences or very limited experiences interacting with social robots [15, 19, 26]. Edwards et al. [14] indicated that there is a significant knowledge gap regarding our understanding of the role of expectations in

shaping human–robot interactions. Similar ideas have been previously echoed in the HRI literature [e.g., 4, 5, 10, 11, 15, 25].

In this paper, we report results from a within-subject study of the relationship between expectations and experiences during an in-person social interaction with Pepper with the aim to investigate the role of expectations in HRI. The robot was equipped with dialogue system powered by the GPT-3 large language model, allowing open-ended dialogue with the robot. The primary purpose of the study was to investigate how the experience of interacting with a social robot affects individuals' expectations over time.

2 Background

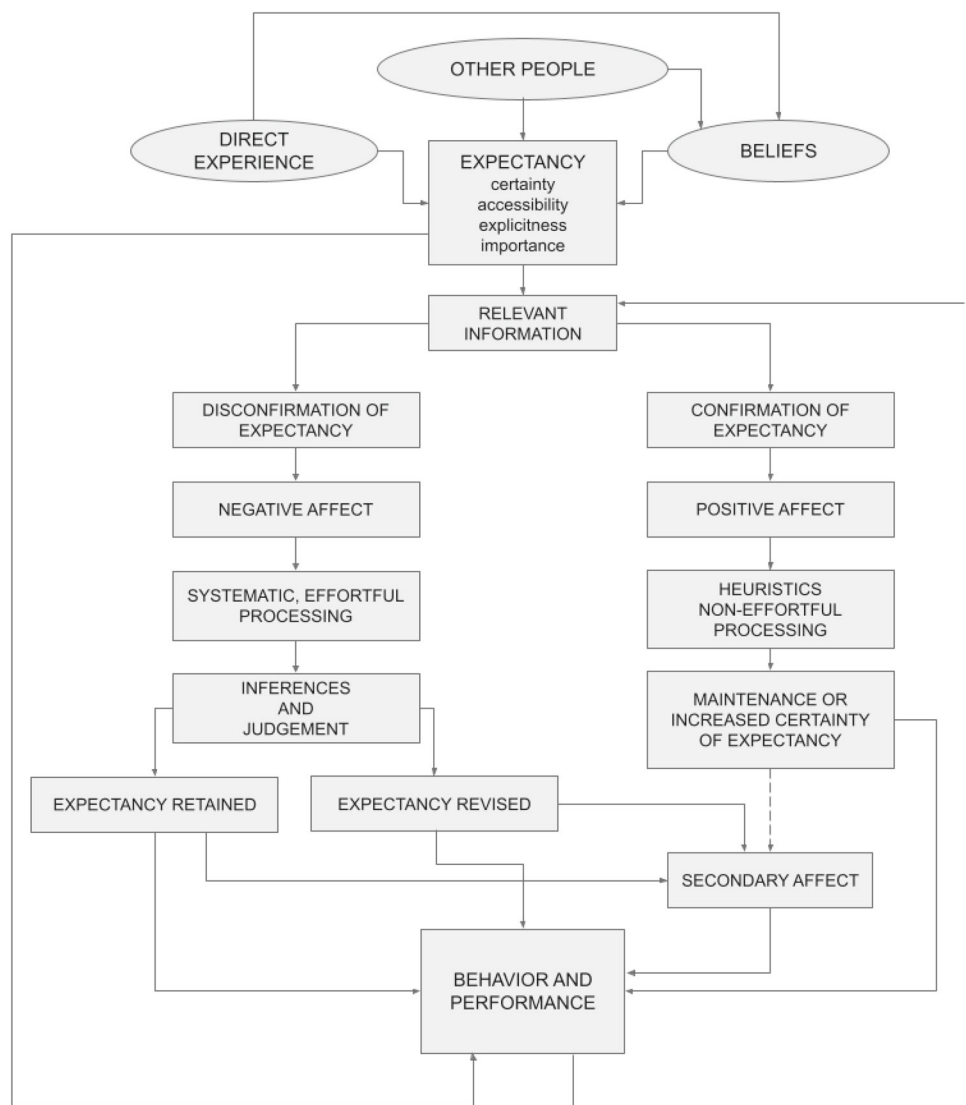
Expectations have been studied for decades in several fields, most noteworthy in social psychology [1, 3]. Expectations fundamentally affect action and play an important role within the human belief system. Expectations can be viewed as a vessel that can be filled with semantics, beliefs, and past experiences to guide us forward [2]. Since expectations also deal with predictions of future events, expectations can also be associated with wishful thinking and subsequent failure and disappointments (i.e., disconfirmed expectations) and can drastically alter an individual's emotions and behavior [13]. While social robots are becoming increasingly popular, many individuals still lack first-hand experience with them. As a result, the expectations people form about robots are likely grounded in some combination of interactions with media accounts or robots (factual and fictional), interactions with other interactive technologies, interactions with people or animals, or their own imaginations. Thus, understanding how expectations are formed and changed as well as how they affect experiences is important for the study of social HRI.

2.1 Model of the Expectation Process

A model of expectations has previously been proposed for HRI by Rosén et al. [25], presented in Fig. 1, modified from Social Psychology originally by Olson et al. [1].

As presented in this model, all expectations are derived from beliefs. Beliefs are statements we take to be true and expectations are the implications of these beliefs for the future [1, 25]. There are three sources of the beliefs that serve as the basis for expectations. *Direct experiences* generate beliefs based on first-hand information which are the basis for expectations that are typically more trustworthy and more confidently held. *Indirect experiences* generate beliefs based on the (direct and indirect) experiences of others. Expectations based on these beliefs are likely to be less trusted and have lower confidence than expectations

Fig. 1 The expectancy (expectation) process [1], with permission



grounded in beliefs formed by direct experiences. *Inferences* generate beliefs through reasoning about other beliefs and experiences. Beliefs and expectations can be changed and refined through various experiences, with all three types of experience often contributing to the beliefs and expectations individuals bring to interactions with robots in HRI studies.

Once expectations confront reality, the expectation is either confirmed or disconfirmed. When expectations are disconfirmed, inferences and judgments are made regarding the event which leads to either *retaining* or *revising* the expectation [1]. Retaining the expectation means that the person will keep their initial expectation despite evidence contradicting the expectation. Revising the expectation means that the expectation is updated to agree with the experience.

The potential effects of confirmed and disconfirmed expectations on an individual can be categorized as three factors affecting human experiences [24]. First, *cognitive processing* refers to how straining an expectation may be

on an individual's cognitive abilities; disconfirmed expectations typically require high levels of cognitive processing whereas confirmed expectations typically require low levels of cognitive processing. For example, when an expectation is disconfirmed, cognitive effort may go towards identifying and remembering the context of the disconfirmation as these details may be relevant to future unexpected events. Second, *behavior and performance* are the changes in an individual's deliberate actions based on confirmed and disconfirmed expectations. Expectations are the basis for behaviors, grounding our intentions and guiding our actions [1, 3, 25]. This can be easy to see in the extreme case of self-fulfilling prophecies, where an individual's expectations influence their behavior in a way that nearly guarantees that the expectation is confirmed. Lastly, *affect* refers to the emotional reactions, ranging from negative to positive, an individual may have after an expectation is confirmed or disconfirmed. There are many affective processes, though in

HRI it is common to focus on individual attitudes toward robots (e.g., Negative Attitude towards Robots Scale (NARS) [27]). When an expectation is confirmed, a person will typically judge the experience as pleasant, or at least neutral, but when an expectation is disconfirmed a person will typically judge the experience as uncomfortable or unpleasant [1]. For the present study, affect is the main factor we focus on.

2.2 Previous Research on Expectations in HRI

Lohse [26], explicitly addressed the role of expectations in HRI and provided a point of departure for introducing some assumptions about individuals' expectations, emphasizing the need to explore the influence of expectations in HRI research. Since Lohse [26], several authors have also identified a need to study the expected capabilities of robots versus their actual capabilities [e.g., 15, 20, 25]. Moreover, there is a growing pressure to study expectations with real robots in-person instead of surveys and observations of interactions. These real interactions are expected to provide a more accurate picture of participants' assessments of social robots and the quality of interactions with such robots [5]. In fact, a study by Wallkötter et al. [28] showed that by changing the context in a HRI study from online videos to real-world interaction conditions influenced the participants' perception of the robot's 'mind'. These results demonstrate how subjective measures may depend on the presentation context of social robots.

The physical appearance of robots may also affect users' expectations. Manzi et al. [23] demonstrated that the physical appearance and the behaviors performed by Nao and Pepper affected the interaction quality, independent of the particular robot. In addition, Edwards et al. [14] studied how initial expectations and impressions can be altered and confirmed through limited first-hand experience when communicating with Pepper. After a brief initial interaction with Pepper, many participants reported feelings of affinity and connectedness, whereas a nearly identical encounter with a human experimenter resulted in opposite outcomes.

Jokinen and Wilcock [19] investigated whether high expectations are associated with the users' experience in an interaction with Nao, and examined via a modified SASSI questionnaire if the users' before and after experiences with the robot have an impact on their self-assessment and quality aspects of the interaction. Their results confirm that expectations in general were rated higher than the actual experience. The results show that a majority of the participants perceived a positive interaction experience, and indicate that the participants perceived the interaction with Nao as more enjoyable and interesting. However, there were indications of a negative tendency related to their expectations of Nao's behavior

and to what extent the participants perceived that they were 'understood' by the robot. Interestingly, the authors observed that the most experienced participants seemed to be the most critical ones. The authors also emphasized that reducing the mismatch between individuals' expectations and their experiences during interaction is important for the development of trust between robots and their intended users over the long term.

Horstmann and Krämer [4] explicitly studied what kinds of expectations people have about social robots as well as the sources of their expectations (e.g., direct or indirect experience). The results indicate that previous experiences with social robots in movies and in social media lead to increased expectations regarding the ability of robots to be an active part of their personal lives and society. Moreover, individuals' awareness of negatively perceived fictional social robots increased negative expectations of robots as threats to humans, while having more knowledge about the capacities and limitations of robot technology showed reduced levels of anxiety towards social robots. They concluded that people most likely form expectations of social robots from various information sources, and more research is needed. Horstmann and Krämer [4] suggested that future work should examine what kinds of expectations and preconceptions people hold towards robots, and in what ways these influence their behavior when interacting first-hand with a social robot.

Finally, Paetzel et al. [29] showed that various perceptual dimensions stabilized within different time frames in an interaction. While perceived competence was judged quickly by the participants and remained stable after only two minutes of social interaction, game play with Furhat improved participants' impressions of the robot head's anthropomorphism and likability which continued to increase until the second session. However, the perceived threat and discomfort continued to fluctuate until the last session. Notably, this study highlights the importance of allowing participants time to interact with the robot before examining their perception of it.

The studies presented in this section demonstrate that expectations affect participants experience with a robot, based on many dimensions such as previous experience with robots, stimulus presentation, experimental context, the robot's behavior, and the design (appearance) of the robot. While these studies establish the relevance of investigating the impact of expectations in HRI research, they also highlight specific gaps in this research area. Notably, there is a lack of research that considers expectations before the interaction and how those expectations change over multiple interactions with a real robot. Moreover, there is a need for insights into how expectations affect individual experiences when interacting with a real social robot.

2.3 Research Question and Hypotheses

With previous research on expectations in HRI in mind, we designed an experiment with the aim to better understand the relationship between human expectations of social robots and multiple first-hand interactions with a real robot. Specifically, we address the following research question: *how does experience of interacting with a social robot affect users' expectations over time?* Considering previous research we've formulated three hypotheses related to this question:

Hypothesis 1 The variability between participants' expectations towards the robot will decrease over time

A key component of the expectancy process by Olson et al. [1] is that expectations are formed on basis of prior interaction. If this assumption is correct, individuals that meet the same interaction partner, in this case a social robot, should over time move towards similar views of the robot, even if they began with very different expectations. As a consequence, the variability in participants' expectations, reflected in perceived capability of and affect towards the robot, should decrease over time.

Hypothesis 2 Previous experience affects expectations of robots

If Olson et al. [1] is correct in that expectations are formed on basis of previous interaction, participants' previous experiences interacting with robots should be reflected in their expectations. Thus, participants' expectations, reflected in perceived capability of and affect towards the robot collected before they interact with the robot in this study, should differ significantly between participants with and without previous first-hand experience of interacting with robots.

Hypothesis 3 Expectations will change based on experience with the robot

As stated in the expectation process by Olson et al. [1], expectations change continuously during an interaction, especially in new situations. Given the novelty of the GPT-powered dialogue system that is used in this experiment and the fact that none of the participants have interacted with this specific robot setup before, there should be a change in expectations over time, reflected in the mean scores of perceived capability of and affect towards the robot.

3 Method

With the research question and the three hypotheses in mind, we designed a within-subject experiment to measure how expectations may affect a forthcoming interaction and how they may change over time throughout an interaction. In other words, we investigated how time (i.e., experience with

the robot) affected participants' expectations in a human–robot interaction. The current study's experimental design were guided by the Social Robot Expectation Gap Evaluation Framework proposed by Rosén et al. [25]. The framework outlines a methodological approach for investigating and analyzing individuals' expectations before, during, and after interaction with a social robot from a human-centered perspective. Moreover, the framework focuses on measures for three factors of expectations—cognitive processing, behavior & performance, and affect. Here, we focus primarily on the affect-component of expectations using three measures commonly used within research on HRI: NARS [27], RAS [27], and Closeness, inspired by the IOS scale [30]. We also created a single question asking participants about the perceived capability of the robot.

3.1 Participants

The participants' ($N = 31$) ages ranged between 20–54 years old ($M = 29$), with 45% identifying as male and 55% identifying as female (no-one self-described or chose non-binary). The interaction with the robot was conducted in English; 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. We asked these questions in order to investigate if accent could be a confounding variable in the study. Participants were recruited by flyers on campus as well as emails to faculty and students.

Previous experience with robots was assessed through a single question: *What's your previous experience with robots?* This question was answered on a scale between 1 (*I have no previous experience*) and 5 (*I have a lot of experience*). Out of all the participants, 48% had no previous experience while 52% had some previous experience (29% chose 2, 16% chose 3, 7% chose 4, and no-one chose 5 in the scale). Interests in robots was assessed through a single question: *How interested are you in robots?* This question was answered on a scale between 1 (*No interest*) and 5 (*Very interested*). Out of all the participants, no-one chose 1 or 2, 42% chose 3, 16% chose 4, and 42% chose 5.

3.2 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 was performed in accordance with the Declaration of Helsinki. There were no physical or mental health risks to the participants of this study. Participants were informed of their tasks and right to withdraw prior to providing consent through signing an

informed consent form. All data has been de-identified during collection when possible. No sensitive personal information were 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 were deleted after the publication phase.

3.3 Procedure

The participants were instructed to interact with Pepper with the freedom to explore what conversations are possible for two and a half minutes, in two interactions total. Participants were told prior to the first interaction that we were investigating how individuals interact with robots that are intended to be used in the home and that they could ask the robot anything. Once the participants entered the lab room, they were asked to read and sign the consent form an informed of their right to withdraw consent at any time without penalty. Then they filled in questionnaires, followed by the first interaction, then another round of questionnaires, followed by the second interaction, then the final set of questionnaires, and lastly an open-ended verbal interview about their experience of interacting with Pepper. If the participant's speech was not recognized by the robot, the test leader told the participant to try to speak more loudly. There were instances where participants asked the test leader why the robot was not responding, in which the test leader said the same thing. A movie ticket was given for their participation in the study. During debriefing, participants were informed of the study's aims and were provided with an opportunity to ask further questions. We also disclosed how the robot and its speech function worked.

3.4 Materials and Technological Setup

For the present study, Aldebaran's Pepper with a customized dialogue system was used [31]. The dialogue system utilized the OpenAI GPT-3 language model for producing responses to participants' verbal input [32]. The dialogue system was implemented as a text completion, using the *text-davinci-002* language model; i.e., GPT was asked to generate a probable continuation of the presented prompt. Before interaction with the first participant, the language model was initiated with a short prompt: *You are talking to Pepper. We are currently at the Interaction Lab in a town called Skövde. We are in the country Sweden.* No other adaptation of the GPT language model was made.

The dialogue was always initiated by the participant. The participant's speech was transformed into text using Google's speech-to-text service, and the initial prompt was combined with the participant's verbal utterance. The GPT system was then responding with a probable answer given both the initial prompt and the verbal input. On the following requests, all previous dialogues were included in the prompt, appended

with the most recent verbal utterance by the participant. This way, the robot's responses were not only based on each individual utterance, but on the entire dialogue with that participant.

Produced text completions were transformed into spoken utterances by the robot using the NaoQi ALAnimatedSpeech service, resulting in both synthetic robot speech as well as arm and head gestures. Additionally, the built in *autonomous life* functionality was used, providing simulated breathing and basic attention (head turns) towards the participant. The robot was configured not to locomote during the study. Technical details of the dialogue system can be found in [33] and the source code is available <https://github.com/ilabsweden/pepperchat>. An example of a conversation with the robot, illustrated by the first author, is available at <https://youtu.be/zip90jyv1i4>.

The experiment was performed in the Interaction Lab at the University of Skövde, Sweden. The lab is a 60 m² room of which about half is the open space dedicated to the interaction. The remaining part of the room was arranged with a desk for the experimental leader, computers, and other equipment used in the lab. The participants were asked to sit on a chair approximately one meter in front of Pepper (Fig. 2). There were two cameras recording the interactions and the post-test interviews, one from the side and one right behind Pepper (for a clear view of the participants' facial expressions and bodily movements).

3.5 Measures

The dependent variable in this experiment is expectancy, measured via negative attitudes, anxiety, closeness, and perceived capability. The independent variable was time, i.e., experience from the interactions with the robot. Data collection occurred throughout the interaction, with questionnaires before the first interaction, after the first interaction, and after the second interaction. In addition, previous experience with robots (see Sect. 3.1) was used as a between subjects factor in analysis.

3.5.1 Negative Attitudes Towards Robots

The negative attitudes towards robots scale (NARS) is a 14-item questionnaire which seeks to further understand humans behavior and negative attitudes toward robots [27]. NARS consists of three subscales. The first sub-scale, *S1: Negative attitude toward situations of interaction with robots* has a summary assessment range of 6–30. The second sub-scale, *S2: Negative attitude toward social influence of robots* has a summary assessment range of 5–25. The third sub-scale, *S3: Negative attitude toward emotions in interaction with robots* has a summary assessment range of 3–15. Participants were



Fig. 2 Experimental set-up viewed through the two cameras used to record interaction between participant and robot. Example scene taken from one of the participants (with permission)

asked to assess each question on a scale of 1–5 Likert scale, with 1 being *I strongly disagree*, and 5 being *I strongly agree*.

3.5.2 The Robot Anxiety Scale

The robot anxiety scale (RAS) is a 11-item questionnaire that measures the altered behavior participants may have towards robots based on their anxiety towards robots [27, 34]. The RAS consists of three subscales. The first sub-scale, *S1: Anxiety toward communication capability of robots* has a summary assessment range of 3–18. The second sub-scale, *S2: Anxiety toward behavioral characteristics of robots* has a summary assessment range of 4–24. The third sub-scale, *S3: Anxiety toward discourse with robots* has a summary assessment range of 4–24. Participants were asked to assess each question on a scale of 1–6 Likert scale, with 1 being *I do not feel anxiety at all*, and 6 being *I feel anxiety very strongly*.

3.5.3 Closeness

We based our Closeness questions on the Inclusion of the Other in the Self Scale (IOS) [30] which is a questionnaire that measures how close the participants felt to the robot in the experiment. As this scale is not originally intended for HRI, we decided to include three questions that have been used as a part of scale validation [35]. The first question is *Q1: Please, select the appropriate number below to indicate*

to what extent you would use the term “WE” to characterize you and the robot, the second question is *Q2 Relative to all your other relationships (both same and opposite sex) how would you characterize your relationship with the robot?*, and the third question is *Q3: Relative to what you know about other people’s close relationships, how would you characterize your relationship with the robot?*. The participants were asked to rate each question on a 1–7 scale, with 1 being *Not at all* for Q1 and *Not close at all* for Q2 and Q3, and 7 being *Very much so* for Q1, and *Very close* for Q2 and Q3.

3.5.4 Perceived Capabilities

Perceived Capabilities is a question that was created for this experiment which asked the participants: *How capable do you think the robot in this study is?* on a scale of 1 to 9, with 1 being *Not capable at all* and 9 being *Extremely capable*. The exact type of capability (e.g., cognitive or social) was not specified, but rather participants were asked to rate a more general idea of capability.

4 Results

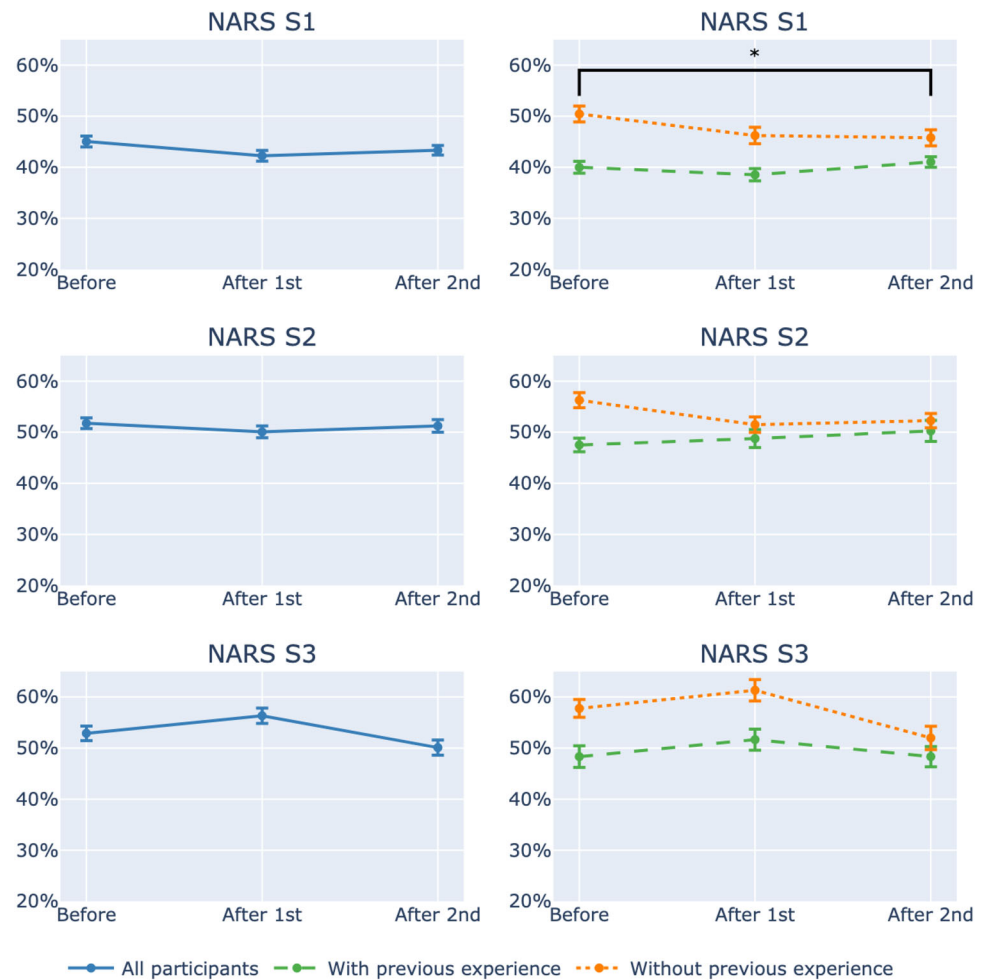
In the present study, we investigated how the collected measures (NARS, RAS, Closeness, and Perceived Capability) changed over time, specifically *before the first interaction*, *after the first interaction*, and *after the second interaction*. We also investigated the effect of previous experience with robots, collected as part of the pre-questionnaire (Sect. 3.1). Group 1 includes only participants that answered 1 (no previous experience of robots) and group 2 includes all participants responding 2 or above. Hypothesis 1 was evaluated using F-tests performed in R-studio, while Hypotheses 2 and 3 were evaluated using ANOVA in Jasp. Bonferroni correction was used to compensate for repeated tests. Results for each measure are presented below.

4.1 NARS

The NARS questionnaire responses were analyzed as the sum of each participant’s responses to each subscale (c.f., Sect. 3.5.1). Mean results are presented in Fig. 3. The overall score for S1 (range: 6–30) were 13.52 ($SD = 3.56$), 12.68 ($SD = 3.49$), 13.00 ($SD = 3.13$). The overall score for S2 (range: 5–25) were 12.94 ($SD = 2.92$), 12.52 ($SD = 3.21$), 12.81 ($SD = 3.47$). The overall score for S3 (range 3–15) were 7.94 ($SD = 2.38$), 8.45 ($SD = 2.51$), 7.52 ($SD = 2.49$).

To test Hypothesis 1, separate two-sided F-tests was used to test the difference in variance between the data collected before the first interaction and after the last interaction. No statistically significant effects on variability were found.

Fig. 3 Mean scores for the three NARS components for all participants (left) and separated based on previous experience with robots (right), as a percentage of the maximum score for each component. Error bars indicate standard error of the mean



To test Hypothesis 2 and 3, a repeated measures ANOVA was performed on each subscale in relation to time as a within subjects factor and previous experience with robots as a between subjects factor. No statistically significant effects of time were found for any of the subscales. However, there was a statistically significant main effect of previous experience with robots for S1 ($F(1, 29) = 4.76, p < 0.05$). A post-hoc pairwise comparison revealed that group 1, without previous experience with robots, provided significantly higher responses to S1 than group 2. Similar trends, with more negative attitudes from group 1, were observed also for S2 and S3. However these differences were not significant.

4.2 RAS

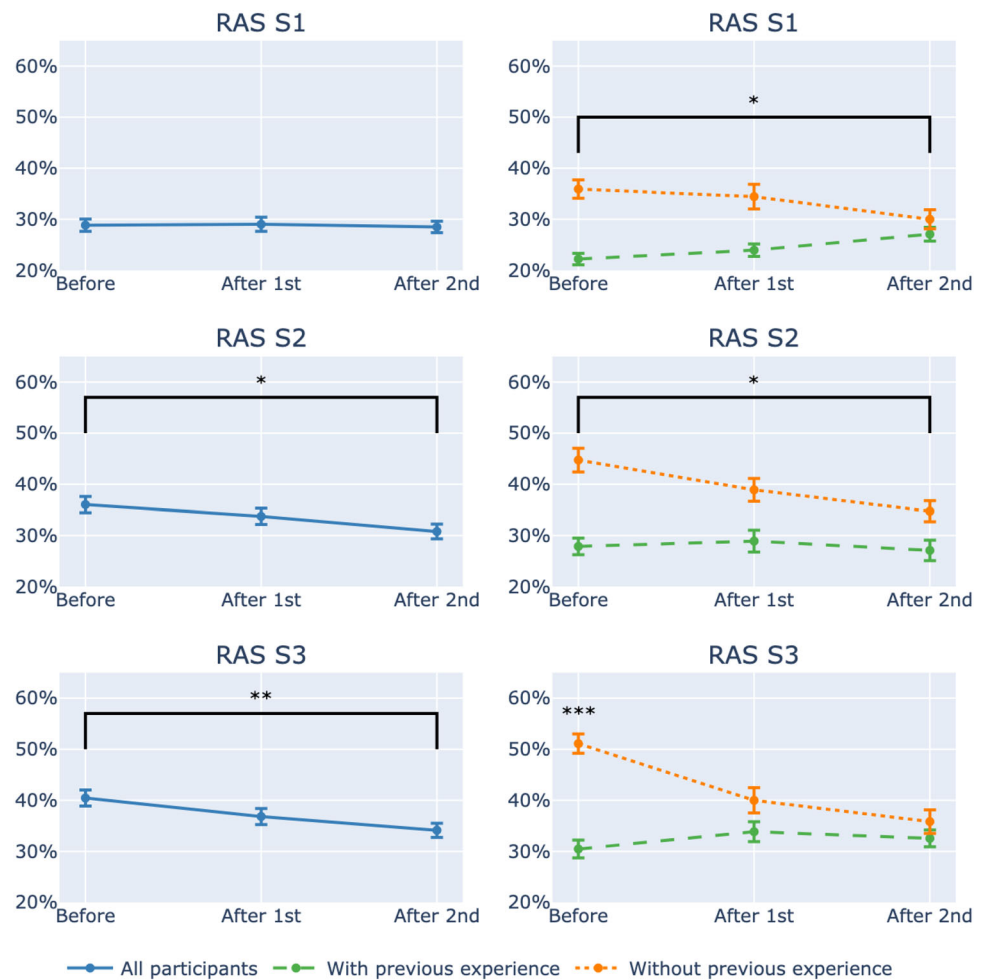
The RAS questionnaire responses were analyzed as the sum of each participant's responses to each subscale (c.f., Sect. 3.5.2). Mean results are presented in Fig. 4. The overall scores for S1 (range: 3–18) were 5.19 ($SD = 2.41$), 5.23 ($SD = 2.79$), 5.13 ($SD = 2.26$). The overall scores for S2 (range: 4–24) were 8.65 ($SD = 4.22$), 8.10 ($SD = 4.23$), 7.39

($SD = 3.87$). The overall scores for S3 (range 4–24) were 9.71 ($SD = 4.20$), 8.84 ($SD = 4.17$), 8.19 ($SD = 3.70$).

To test Hypothesis 1, separate two-sided F-tests was used to test the difference in variance between the data collected before the first interaction and after the last interaction. No significant effects on variability were found.

To test Hypothesis 2 and 3, a repeated measures ANOVA was performed on each subscale in relation to time as a within subjects factor and previous experience with robots as a between subjects factor. There were statistically significant main effects of time for S2 ($F(2, 58) = 4.14, p < 0.05$) and S3 ($F(2, 58) = 5.19, p < 0.01$). A post-hoc pairwise comparison revealed that both subscales had statistically significant before interaction, compared to after the second interaction. Significant main effects of previous experience were found for S1 ($F(1, 29) = 4.53, p < 0.05$) and S2 ($F(1, 29) = 4.49, p < 0.05$). Significant interaction effects of time and previous experience were found on all three subscales, S1 ($F(2, 58) = 4.86, p < 0.05$), S2 ($F(2, 58) = 3.26, p < 0.05$), and S3 ($F(2, 58) = 10.24, p < 0.001$). Group 1 showed a negative trend, reduced anxiety, on all three subscales, while group 2 showed positive (S1) or flat (S2, S3) trends.

Fig. 4 Mean scores for the three RAS components for all participants (left) and separated based on previous experience with robots (right), as a percentage of the maximum score for each component. Error bars indicate standard error of the mean



4.3 Closeness

The responses to questions on Closeness were analyzed individually for each of the three questions (c.f., Sect. 3.5.3). Mean results are presented in Fig. 5. The mean score for Q1 were 2.93 ($SD = 1.55$), 2.97 ($SD = 1.62$), 2.97 ($SD = 1.83$). The mean score for Q2 were 1.77 ($SD = 1.09$), 2.35 ($SD = 1.47$), 2.29 ($SD = 1.44$). The mean score for Q3 were 1.84 ($SD = 1.27$), 2.42 ($SD = 1.48$), 2.26 ($SD = 1.53$).

To test Hypothesis 1, separate two-sided F-tests were used to test the difference in variance between the data collected before the first interaction and after the last interaction. No significant effects on variability were found.

To test Hypothesis 2 and 3, a repeated measures ANOVA was performed on each question in relation to time as a within subjects factor and previous experience with robots as a between subjects factor. There were statistically significant main effects of time on Q2 ($F(2, 58) = 6.39$, $p < 0.01$) and Q3 ($F(2, 58) = 6.86$, $p < 0.01$). A post-hoc pairwise comparison revealed that participants' responses to these questions increased significantly between measures taken before the interaction and after the first interaction. Additionally, both

measures changed significantly between before interaction and after the second interaction. There was no statistically significant main effect of previous experience on any of the three questions related to Closeness.

Significant interaction effects of time and previous experience were found on Q1 ($F(2, 58) = 3.22$, $p < 0.05$). A post-hoc pairwise comparison revealed that responses from group 2 were significantly higher than group 1 before interaction, a difference that disappeared after the first interaction.

4.4 Perceived Capability

The result for Perceived Capability is the average for all participants, scale from 1–9. The mean scores for the respective measurement times were 5.26 ($SD = 1.39$), 4.71 ($SD = 2.07$), and 4.64 ($SD = 2.30$). Mean results are presented in Fig. 6.

To test Hypothesis 1, a two-sided F-test was used to test the difference in variance between the data collected before the first interaction and after the last interaction. Although no statistical significance was found ($F(1, 29) = 2.7$, $p = 0.072$),

Fig. 5 Mean scores for the three Closeness questions for all participants (left) and separated based on previous experience with robots (right), as a percentage of the maximum score for each component. Error bars indicate standard error of the mean

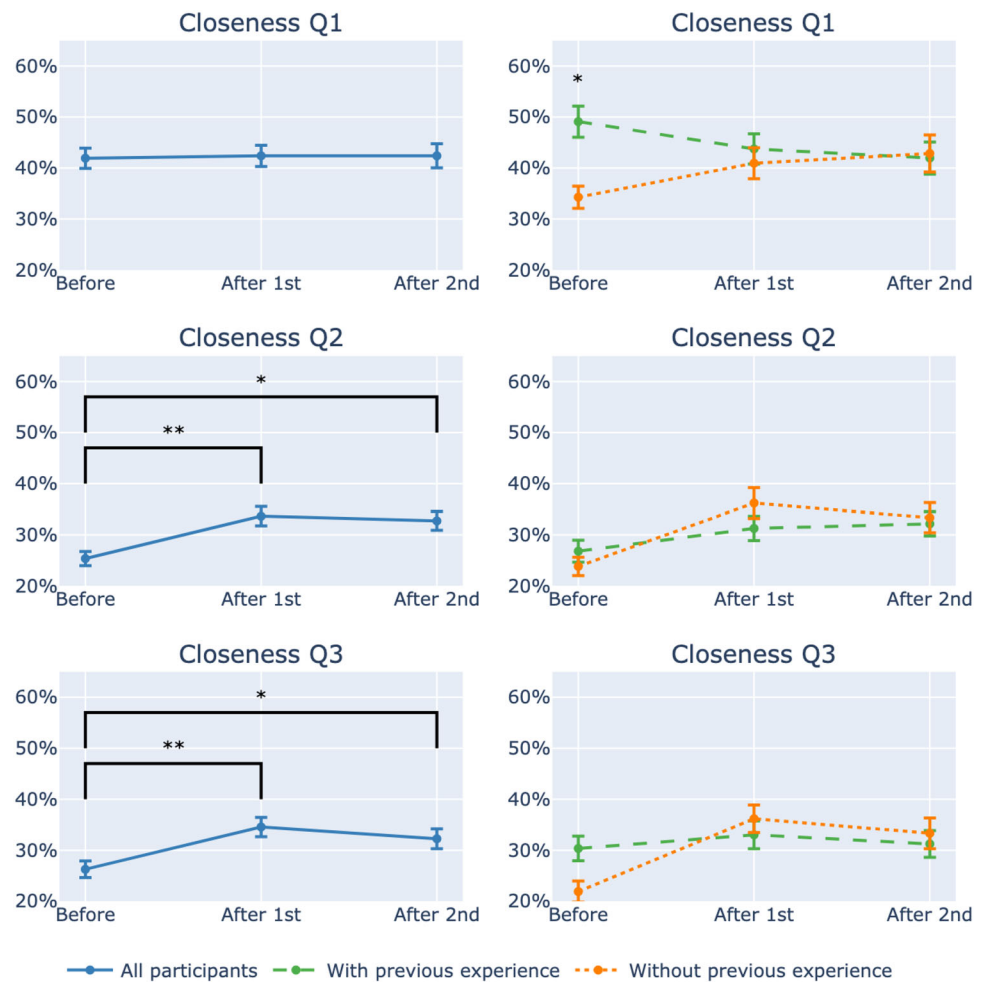
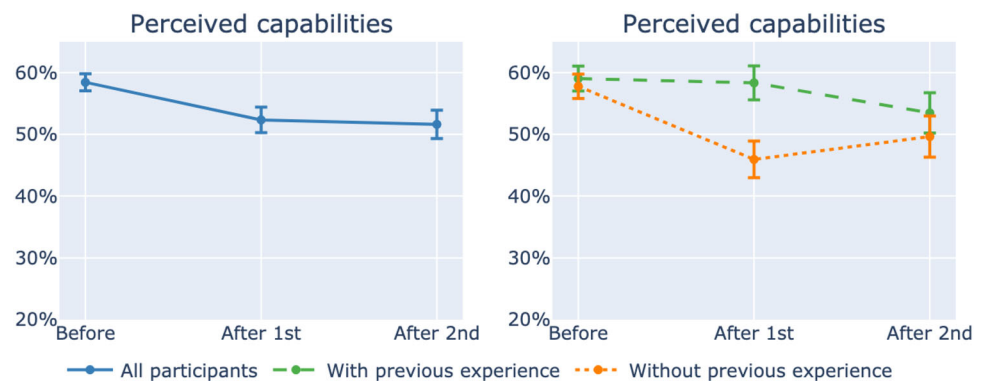


Fig. 6 Mean scores for the Perceived Capabilities question for all participants (left) and separated based on previous experience with robots (right), as a percentage of the maximum score for each component. Error bars indicate standard error of the mean



results reveal a strong tendency of increasing variability with time.

To test Hypothesis 2 and 3, a repeated measures ANOVA was performed on Perceived Capability in relation to time as a within subjects factor and previous experience with robots as a between subjects factor. No statistically significant relationships were found.

5 Discussion

In this study, we investigated dimensions of human expectations of robots in an open-ended in-person interaction between participants and a social robot. Participants were asked to have two short interactions with Pepper and to fill in

questionnaires related to expectations before the interaction, after the first interaction, and after the second interaction. We were interested in how the experience of interacting with a social robot affects expectations over time. We hypothesized that variability between participants would decrease over time, previous experience would affect the expectations, and that expectations would change over time.

Results show that participants' responses did not move towards agreement and that participants tended to stick with their initial expectations based, in part, on their previous experience with robots. Therefore, Hypothesis 1, related to a decrease in variability was rejected, whereas Hypothesis 2, related to the effects of previous experience with robots on expectations, was supported. A mixed picture appeared in relation to Hypothesis 3, concerning change in subjective measures over time. Overall, participants' responses changed less over the course of the interaction than what we expected. It appears that participants' initial expectations of robots were sufficiently robust that they were only moderately affected by the two interactions with the robot. In fact, the results indicate that participants' responses on several measures were influenced more by their previous experience with robots than the human–robot interaction they had just experienced.

In the following discussion, we consider possible explanations of our results in relation to each of the three hypotheses.

Hypothesis 1 The variability in participants' expectations towards the robot will decrease over time

We hypothesized that direct experience interacting with the robot would cause participants to adjust their expectations and reduce the gap between an individual's expectations and the actual capabilities of the robot, leading to reduced variability in subjective measures. However, no significant decrease in variability over the two interactions was observed. In fact, variability in reports of perceived robot capability appeared larger, though not significantly, after interacting with the robot, compared to measures taken before interaction.

Given that participants likely have very different initial expectations of the robot due to different previous experiences, we expected that experience with the same robot would cause their expectations to converge as their expectations would revise to a more accurate picture of the robot, in line with the Expectancy Process by Olson et al. [1]. As we did not see this, it is possible that these sources of the expectation had such a strong effect that the variability did not have time to decrease, thus participants (at least initially) retained their expectations [1].

Another possible explanation is that, despite interacting with the same robot, the content and flow of participant interactions with the robot were too distinct for each participant that the experiences were too different to induce a decrease in variability. This interpretation is supported by

the strong tendency for increased variability over time in the Perceived Capability measure. For this measure, a large proportion of participants did revise their expectations, but apparently in different directions. The GPT-based dialogue system used in the present study may be a reason for this, which allowed for open and uncontrolled dialogues with the robot. Although challenging from a methodological point of view, some variability between participants is likely inherent in real in-person interactions between humans and robots where dialogue may be affected by numerous personal and environmental factors. Moreover a more controlled setting, for example, achieved through a limited dialogue system and scripted robot behavior, would also limit participants' ability to actually *interact*, effectively transforming the robot into a stimulus–response system rather than an interaction partner. A WoZ style design may have provided the desired flexibility with more controlled variability, but ultimately directly introduces the expectations of the human actor and the experimenters into dialogue. Moreover, because the GPT-3 model is a technical solution, it is much nearer to potential interactions with future robot interactions than WoZ style human generated text, which may only change expectations because it is actually a dialogue with a human pretending to be a robot.

Ultimately, while any design choice introduces challenges we believe that the open dialogue system used allowed participants to move away from any dialogue they may have previously had with similar digital agents and increased the likelihood that their expectations would change. Based on our results, it seems likely that more detailed investigations are needed to understand the impact of different approaches to testing and imagining dialogue generation in HRI contexts.

While there was no statistically significant decrease in variance for participants overall, we did see significant interaction effects of time and previous experience, revealing a pattern of reduced group differences between participants with and without previous experience of robots. This effect was significant for RAS S3, and for Closeness Q1. This may be seen as partial support for Hypothesis 1, but only on the group level. Notably, where group 1 responses changed significantly, they tended to move towards the response patterns of the more experienced group 2 participants.

Hypothesis 2 Previous experience affects the expectation an individual has of the robot

We hypothesized that participants' previous experience with robots will affect the measures collected before interacting with the robot, which was supported in our results for NARS S1, all RAS subscales, and for Closeness Q1. Surprisingly, the difference between responses from participants with and without previous experience with robots persisted also after interacting with the robot, a pattern that was most prominent for the RAS questionnaire. All effects of previous experience pointed in the same direction, that is, more positive responses

from participants with previous experience of robots. Opposite results have been shown in the study by Jokinen and Wilcock [19], who, in their case, saw a difference in results based on previous experience of robots with participants having more previous experience with robots being more critical. Since there are several differences between the present work and the study by Jokinen and Wilcock [19] it is difficult to say where these differences come from. One possible explanation may be the more open dialogue system used in the present work responded well to more complex input, and was able to impress the more experienced users that came with higher demands on the interaction.

This found effect on previous experience with robots strengthens the argument that the sources of expectations are a strong factor for the experience in an human–robot interaction, in accordance with the Expectancy Process by Olson et al. [1]. The expectations were not only different before the interaction with the robot, but also remain different after the interaction; in terms of the Expectancy Process, this would (similar to Hypothesis 1) support that the sources of the expectations causes participants to retain their expectations in an interaction, at the minimum for the two shorter interactions. As explained earlier, there are three sources of expectations: direct experience, indirect experience, and inferences [1]. Expectations that are built on indirect experience and inferences are not held with the same level of certainty [1]. The main source of expectations of social robots for the public is from media exposure, which has been noted many times in the HRI literature [16, 36, 37]. This is also highlighted in the work by Horstmann and Krämer [4] where the authors found that movies and social media lead to increased expectations of robots' capabilities, and people who have more accurate expectations of robots have less anxiety towards robots. Thus, the observed effect for previous experience and no previous experience is in line with how different sources could result in different expectations.

Hypothesis 3 Expectations will change based on experience with the robot over time

Our third hypothesis was that participants will change their expectations of the robot after interacting with the robot. Results showed that this was true for RAS S2 and S3, and for Closeness Q2 and Q3, but not for any of the NARS scales nor Perceived Capability. Overall, for the measures that did change, participants became increasingly positive towards interacting with the robot.

While we expected a bigger change in collected measures, there is at least one study indicating that previous experience of robots can be difficult to displace [29]. Paetzel et al. [29] found that first impressions have persistent effects on upcoming interactions, with different dimensions stabilizing at different time frames, demonstrating how robust humans' perception of the robot can be. Their study could

explain why we only saw partial significant results as different dimensions solidified at different time frames. In our study, we saw a significant change for anxiety and closeness toward the robot but not for attitudes and Perceived Capability of the robot. It is possible that anxiety and closeness, being more emotional than the other two, are more unstable in a human–robot interaction.

These results also highlight the potential issue with subjective measures in HRI as expectations seem to be a strong confounding variable in a human–robot interaction. As such, positive or negative attitudes, feelings of anxiety, or even reported relationship with the robot, may stem from many factors other than the precise robot interaction under investigation.

5.1 Limitations and Future Work

One of the main limitations of this experiment is the potential for priming participants by having them fill out the questionnaires before the interaction. Having individuals think about different dimensions of expectations may make implicit expectations explicit. However, this is likely a challenge for any study on expectations that wants to compare expectations previous to an interaction to those after the interaction. It is even more challenging to study implicit expectations without making them explicit since implicit expectations are often made explicit through the process of thinking about or interacting with the object that expectations are directed towards [1].

Another limitation lies in the duration of the experiment. As discussed, it is possible that the variance would decrease if interaction times were longer and there were more interactions (possibly over several weeks or months). However, even if the variance does not decrease over time, it seems that expectations would be affected by more experience and thus more long-term studies may be important in HRI.

For future work, we propose a qualitative analysis, in which we will investigate and analyze the interaction quality during the interactive sessions as well as conduct a reflexive thematic analysis [38]. A qualitative analysis of the collected data from the video recordings of the HRI sessions and the post-test interviews could potentially provide some additional insights to the obtained quantitative results, and would address other aspects of the Social Robot Expectation Gap Evaluation Framework by Rosén et al. [25]. The purpose of such a qualitative analysis would be to shed light on how humans, as the users of robots, experience this kind of interaction, deepening our understanding of the aspects that influence their experiences and the ways that implicit and explicit expectations are manifested.

Moreover, because we found that previous experience was a major factor of how participants experienced the robot, we propose to investigate what kind of previous experience par-

ticipants had in more in-depth for future work. Both indirect (e.g., social media, news, science fictions) and direct experience (work situation, public space, hospital setting) can be quite diverse, and there are many future directions that can map this out further to understand how expectations may vary. What kind of robots (e.g., social, industrial) and even what exact robot (e.g., Pepper, Baxter) may also provide further insight into participants' expectations in HRI.

6 Conclusion

With this work, we present the results of an empirical study investigating how the experience of interacting with a social robot affects users' expectations over time. Participants were instructed to converse with Pepper and explore the interaction, in two sessions lasting 2.5 min each. Questionnaires were used to measure affective responses to the interaction and participants were asked to rate the capability of the robot as they saw it. We found that previous experience with robots has an effect on collected measures, not only before interacting with the robot, but to large extent also after interaction.

Our results highlight the importance of tracking sources of expectations (i.e., previous experience with robots) and considering potential effects on forthcoming interactions. This tracking could include what kind of previous experience the users actually have (e.g., indirect experience with media such as science fiction or direct experience with actual human–robot interaction such as hospitality robots at hotels). Further, our results have been analyzed through the lens of the Expectancy Process by Olson et al. [1] from psychology and thus applied to an HRI context. We also demonstrate how aspects of the framework by Rosén et al. [25] works in an empirical application.

We demonstrate how expectations may be measured throughout an interaction, considering how expectations change (or not) over the course of interactions. Many HRI experiments involve a single interaction, with questionnaires filled out *after* the interaction. If participants come into the study with strong expectations, the results may reflect such preconceived notions of social robots to a larger degree than the human–robot interaction setting being studied. This phenomenon may be increasingly problematic in cases where user expectations affect the interaction as such, as discussed in relation to Hypothesis 1 above (Sect. 5). In most experimental HRI research with human participants we assume the target population (users) to be a homogeneous group that react in a similar way to a number of conditions (independent variables). However, in real interaction, these conditions are rarely met. Participants are not necessary a homogeneous group and their different feelings and expectations will shape the interaction, making different users to experience different things.

While we believe that controlling for previous experience is one way to approach this problem, it is worth noting that little or no previous experience with social robots is no guarantee that participants come in with the same expectations. Participants with less actual experience form interacting with robots are likely to base their expectations on other sources (e.g, indirect experiences and inferences). First-hand experiences with robots may quickly accumulate enough to cause expectations to be more accurate and as such, having participants interact, over an extended time period with a social robot, could reduce the risk that previous experience act as a confound in HRI research. With this work, we contribute to the small, but growing, body of work investigating expectations in HRI.

Acknowledgements This work was partially supported through the Knowledge Foundation as a part of both the Recruitment and Strategic Knowledge Reinforcement initiative.

Funding Open access funding provided by University of Skövde.

Data Availability The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors have no competing interests to disclose. The contents of the manuscript have been unanimously agreed by all co-authors. We certify that the submission is original work and not submitted for consideration elsewhere.

Ethical approval This study involved human participants. Written consent was taken from all participants. The study was conducted in accordance with the Declaration of Helsinki and 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).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Olson JM, Roese NJ, Zanna MP (1996) Expectancies. In: Higgins ET, Kruglanski AW (eds) Social psychology: handbook of basic processes. Guilford Press, New York, pp 211–238

2. Borg MB, Porter L (2010) Following the life-course of an expectation: examining the exchange of expectations in a homeless shelter in New York City. In: Le-n P, Tamez N (eds) *The psychology of expectations*. Nova Science Publishers Hapague, New York, pp 1–47
3. Roese NJ, Sherman JW (2007) Expectancy. In: Kruglanski AW, Higgins ET (eds) *Social psychology: handbook of basic principles*. The Guilford Press, New York
4. Horstmann AC, Krämer NC (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. *Front Psychol* 10:939
5. Horstmann AC, Krämer NC (2020) When a robot violates expectations. In: *Companion of the 2020 ACM/IEEE international conference on human–robot interaction*, pp 254–256
6. Kahn Jr PH, Reichert AL, Gary HE, Kanda T, Ishiguro H, Shen S et al (2011) The new ontological category hypothesis in human–robot interaction. In: *Proceedings of the 6th international conference on human–robot interaction*, pp 159–160
7. Lindblom J, Alenljung B, Billing E (2020) Evaluating the user experience of human–robot interaction. *Human–robot interaction*. Springer, Berlin, pp 231–256
8. Lindblom J, Alenljung B (2020) The ANEMONE: theoretical foundations for UX evaluation of action and intention recognition in human–robot interaction. *Sensors* 20(15):4284
9. Fong T, Nourbakhsh I, Dautenhahn K (2003) A survey of socially interactive robots. *Robotics Auton Syst* 42(3–4):143–166
10. Schramm LT, Dufault D, Young JE (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
11. Berzuk JM, Young JE (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
12. Dautenhahn K (2018) Some brief thoughts on the past and future of human–robot interaction. *ACM Trans Hum Robot Interact (THRI)* 7(1):4
13. Alač M (2016) Social robots: things or agents? *AI & Society* 31(4):519–535
14. Edwards A, Edwards C, Westerman D, Spence PR (2019) Initial expectations, interactions, and beyond with social robots. *Comput Hum Behav* 90:308–314
15. Kwon M, Jung MF, Knepper RA (2016) Human expectations of social robots. In: *2016 11th ACM/IEEE international conference on human–robot interaction (HRI)*. IEEE, pp 463–464
16. Sandoval EB, Mubin O, Obaid M (2014) Human robot interaction and fiction. In: *International conference on social robotics*. Springer, Berlin, pp 54–63
17. Lohse M (2011) Bridging the gap between users' expectations and system evaluations. In: *2011 RO-MAN*. IEEE, pp 485–490
18. Meister M (2014) When is a robot really social? An outline of the robot sociologicus. *Sci Technol Innov Stud* 10(1):107–134
19. Jokinen K, Wilcock G (2017) Expectations and first experience with a social robot. In: *Proceedings of the 5th international conference on human agent interaction (HAI 2017)*, Bielefeld, Germany, October 17–20, pp 511–515
20. Moore RK (2017) Is spoken language all-or-nothing? Implications for future speech-based human-machine interaction. In: *Dialogues with social robots*. Springer, Berlin, pp 281–291
21. Moore RK (2019) Talking with robots. *arXiv preprint arXiv:1912.00369*
22. Horstmann AC, Krämer NC (2020) Expectations vs. actual behavior of a social robot. *PloS One* 15(8):e0238133
23. Manzi F, Massaro D, Di Lernia D, Maggioni MA, Riva G, Marchetti A (2021) Robots are not all the same: young adults' expectations, attitudes, and mental attribution to two humanoid social robots. *Cyberpsychology Behav Soc Netw* 24(5):307–314
24. Rosén J (2021) Expectations in Human-Robot Interaction. In: *Neuroergonomics and cognitive engineering: proceedings of the international conference on applied human factors and ergonomics*. Springer, Berlin, pp 98–105
25. Rosén J, Lindblom J, Billing E (2022) The social robot expectation gap evaluation framework. In: *International conference on human–computer interaction*. Springer, Berlin, pp 590–610
26. Lohse M (2009) The role of expectations in HRI. *New Front Hum Robot Interact* 35–56
27. Nomura T, Kanda T, Suzuki T, Kato K (2004) Psychology in human-robot communication. In: *RO-MAN 2004. 13th IEEE international workshop on robot and human interactive communication (IEEE Catalog No. 04TH8759)*. IEEE, pp 35–40
28. Wallkötter S, Stower R, Kappas A, Castellano G (2020) A robot by any other frame: framing and behaviour influence mind perception in virtual but not real-world environments. In: *Proceedings of the 2020 ACM/IEEE international conference on human–robot interaction. HRI'20*. Association for Computing Machinery, New York, pp 609–618
29. Paetzel M, Perugia G, Castellano G (2020) The persistence of first impressions: the effect of repeated interactions on the perception of a social robot. In: *Proceedings of the 2020 ACM/IEEE international conference on human–robot interaction*, pp 73–82
30. Aron A, Aron EN, Smollan D (1992) Inclusion of other in the self scale and the structure of interpersonal closeness. *J Personal Soc Psychol* 63(4):596
31. Aldebaran. <http://www.aldebaran.com>
32. OpenAI. OpenAI. Available from: <https://openai.com/>
33. Billing E, Rosén J, Lamb M (2023) Language models for human–robot interaction. In: *Companion of the 2023 ACM/IEEE international conference on human–robot interaction (HRI'23 companion)*, March 13–16, 2023, Stockholm, Sweden. ACM, New York
34. Nomura T, Suzuki T, Kanda T, Kato K (2006) Measurement of anxiety toward robots. In: *ROMAN 2006—The 15th IEEE international symposium on robot and human interactive communication*. IEEE pp 372–377
35. Gächter S, Starmer C, Tufano F (2015) Measuring the closeness of relationships: a comprehensive evaluation of the inclusion of the other in the self scale. *PloS One* 10(6):e0129478
36. Yogeewaran K, Złotowski J, Livingstone M, Bartneck C, Sumioka H, Ishiguro H (2016) The interactive effects of robot anthropomorphism and robot ability on perceived threat and support for robotics research. *J Hum Robot Interact* 5(2):29–47
37. Allan DD, Vonasch AJ, Bartneck C (2022) The doors of social robot perception: the influence of implicit self-theories. *Int J Soc Robotics* 14(1):127–140
38. Braun V, Clarke V (2021) One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qual Res Psychol* 18(3):328–352

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.