



Multi-party Turn-Taking in Repeated Human–Robot Interactions: An Interdisciplinary Evaluation

Mateusz Żarkowski¹

Accepted: 11 October 2019 / Published online: 8 November 2019
© The Author(s) 2019

Abstract

As social robots become more popular, so arises the need for these social agents to operate in environments involving multiple users. The robot control systems that govern these multi-party interactions require to be evaluated both from the technical and social standpoints. This paper presents the methodology, setup and results for experiment involving the social robot EMYS participating in multi-party interaction where pairs of participants interacted with the robot in a trivia questions game lead by the robot. In total 32 people, 16 pairs, interacted with the robot twice, which resulted in 32 interactions and 64 filled questionnaires. The developed robot's multi-party interaction system was evaluated both in terms of performance and user assessment. The results show that the robot adhering to human turn-taking social norms reduced the number of occurring conversational errors, which improved the communicative performance from 51.5% to 80.5%, in addition, it made the robot perceived as more communicative, cooperative and fitting user expectations by up to 3 points on a 7 point scale. In addition, the study on repeated interactions revealed that user perception of the robot is affected by subsequent interactions, which can be of consequence in future experiments. This first impression caused lasting effect between 1 and 2 points on user assessment of several robot's aspects, even when contradicted by objective performance measurement of the robot's actual behavior.

Keywords Social robot · Human–robot interaction · Multi-party interaction · Robot control · User assessment

1 Introduction

In the last decade social robots have been receiving growing interest by the scientific community, partners in the industry and finally, the end-users of the robots. Social robots have successfully been placed in the roles of personal assistants and helpers [15,16], however, as of late their social environment is expanding and social robots are encountering more situations where there is a need to interact with multiple users simultaneously. Such examples include a robot participating in a discussion involving several users [28], mediation [19] and participating in a social game either as a player [39] or as a host [7].

One of the core requirements for a social robot is the ability to use natural communication in the interaction with the

user [8]. Within the human–robot interaction (HRI) community there tends to be a division between external and internal aspects of interaction [4]. The external aspects focus on user reaction and assessment of selected features of the robot, such as its morphology or personality, as well as the expression of emotions and intentions, both verbally and non-verbally [2,3,57,59,60]. The research on internal aspects focuses on sensory perception, modeling and control/decision making in the context of interpersonal interaction and its surroundings, natural language processing and machine learning. This research usually focuses on the implementation of a proof of concept scenario in a limited environment and evaluation of its performance [21,27]. The goal is to understand social norms regulating human–human interaction and communication through modeling, as well as to verify whether these principles have a comparable application in human–robot interaction. Examples of such studies are the topics of the robot's touch and embodiment [13,30,46,63], personal space and proxemics [36,55,61,62], and turn-taking [8,11,29,51], the last one being the subject of this paper.

In this paper we present a developed multi-party interaction system and its verification through an experimental

This research was supported by Grant No. 2012/07/N/ST7/03308 awarded by the National Science Centre of Poland.

✉ Mateusz Żarkowski
mateusz.zarkowski@pwr.edu.pl

¹ Wrocław University of Science and Technology, Wybrzeże
Stanisława Wyspiańskiego 27, 50-370 Wrocław, Poland



Fig. 1 The robotic expressive head EMYS [26]

scenario that involves a social robot EMYS (Fig. 1) hosting a game of trivia questions for its users. The system is verified from the perspective of performance evaluation and user opinion assessment. Furthermore, the impact of repeated interactions with the robot on user assessment was studied.

Robotic head EMYS (Emotive headY System) [24,26] was used as experimental platform. This robot was chosen due to its appearance, expressive abilities and well-documented software. EMYS's appearance was designed in a way that would allow him to bypass the problem of uncanny valley (stating that humanoids that appear almost, but not exactly, like real human beings elicit feelings of eeriness and revulsion in observers [35]), while still maintaining the ability to express human-like emotions [25]. The ability to move its head, eyes and eyelids allows EMYS to engage users through gaze, which is an important turn-taking cue for multi-party interaction. The robot had been successfully used in multiple experimental scenarios involving empathic robotic tutors [10,12], closing the mind-body loop [42], adapting the principles and practices of animation for robots [41] and long-term interactions [15].

2 Related Work

Not many robots can autonomously participate in turn-taking, even less can do so in a multi-party setting. This section describes the turn-taking phenomenon, the turn-taking cues are used by social robots, as well as examples of robots operating autonomously in multi-party interactions. It shows that robots operating in multi-party settings are rarely evaluated simultaneously both in terms of performance and

user assessment. Moreover, it was found that neither turn-taking nor multi-party interactions were evaluated through repeated interactions.

In human–human interaction turn-taking occurs naturally. Studies have revealed that general rules of turn-taking seem to be universal [31,47], with relatively small differences across cultures and languages [52]. The rules of turn-taking organize the conversation into turns, during which one of the participants has the right to speak while the others agree to listen. The number of speakers, as well as the length of turns, can vary. The speakers jointly regulate the flow of conversation in order to minimize both the gaps between turns and the overlap. Natural turn-taking is highly efficient and robust, as it works just as well without visual contact. Less than 5% of the conversation involves two or more simultaneous speakers (the modal overlap is less than 100 ms long), while the modal gap between turns is only around 200 ms [32]. Various turn-taking cues are used to signal the intents of the participants, depending on the available channels of communication [14,38]. Conversational analysis [50] recognizes gaze as one of the most important turn-taking cues for multi-party interaction and an indicator of listeners attention.

In case of social robots, most commonly used turn-taking used are pauses, prosody, gaze direction and body positioning. Mutlu [37] showed that a robot can impose a conversational role on participants by shifting attention via gaze during an interaction. Bruce et al. [9] showed that actively turning to human interaction partners significantly increases their willingness to interact. These studies used silence and gaze to facilitate turn-taking behavior, while the robot was teleoperated by the researchers.

Ideally the social robot should be autonomous, however, many HRI studies use the Wizard-of-Oz method [22] in exploratory experiments, in which the social situations are too complex or unpredictable for autonomous robot. In such cases the autonomous robot control system is replaced by a human operator, unbeknown to the participants of the experiment. Multi-party HRI studies that follow this approach involve the topics of building the speech corpus used in such interactions [53], modeling the role of the gaze as a turn-taking signal [1] or recognizing gaze patterns for different conversational roles (speaker, addressee, side-participant) [20]. It is vital that the knowledge gained in this fashion should be used to develop autonomous robots [44].

As of yet, there are not many robots outfitted to autonomously interact with human groups, despite that this type of interaction occurs rather frequently for humans. To provide a background we discuss below three examples of autonomous human–robot interaction in a multi-party context.

Kondo et al. [28] study the impact of robot's gestures on the user interaction assessment and the duration of interaction in multi-party conditions. The experiments involved an autonomous android robot in a large number (1662 people)

of short interactions (under a minute). It was shown that the use of non-verbal communication through gesticulation increases the length of interaction up to twofold and make the users assess the the robot more favorably. Noteworthy in these experiment is the open nature of the interaction with the robot. In contrast, there is no defined task for the robot to accomplish and there is no interaction between users due to the short duration of these meetings, therefore, while the environment is considered as multi-party, the interactions are almost exclusively two-sided.

The work of Pereira et al. [39] explores building social presence for artificial opponents. An autonomous social robot is placed into the role of a board-game opponent for three people during a game of risk. The study focuses on user assessment and the perception of the robot. As a result it presents guidelines for designing socially present opponents. According to the authors, such agent should be embodied, use verbal and non-verbal communication, show emotions, have social memory (history of previous games), and simulate social roles (like motivator, rival or helper) during the game. While usually the multi-party human–robot interaction involves a cooperative scenario, this work is exceptional in that it presents multi-party turn-taking in a situation of conflicting goals of conversation participants.

Bohus and Horvitz developed a multi-party interaction system for a virtual agent [5,6]. The research focuses on facilitating multi-party dialog through gaze, gestures and speech with two and three users. The proposed verbal and non-verbal cues affect the success rate of the action of releasing the floor to a selected user over different dialog contexts (question, confirmation, etc.) In some specific cases noticeable improvements were reported, for example for verbal confirmations within interactions involving two participants and the system, the participant to whom the system had released the floor was the first to speak in 86.2% of the cases. In an other paper [7] a subjective user assessment of the system rated it within 4.5–5.0 on a 7-point Likert scale, however no comparison to any control condition was presented.

Other noteworthy research include: the effects of robot moderation in a team collaborative game, in which the robot influenced the trade-off between social cohesion of the group and the task performance [49], robot controlling the level of engagement between main and side-participants in a four-party setting [34], facilitating inter-group trust through exhibition of vulnerable behavior [54], building relationships and facilitating with children through praise, competition encouragement, sympathy, stimulation [48]. Recent work [18] indicates how human–robot interaction can be affected by factors such as group presence, cohesiveness and group social norms. The above studies show that social robots are taking up more active roles in group interaction and start to influence the behaviors of the individuals, as well as the group altogether.

In conclusion, multi-party HRI studies involving an autonomous robot rarely provide simultaneous performance evaluation and user assessment, as was the case in [7]. The reports either focus solely on user assessment [28,39,58] or on performance evaluation of some multi-party interaction component, e.g. gaze and lip movement detection [43], user engagement classification [17], addressee detection and selection [33]. In contrast, the study presented in this paper utilizes both behavioral and survey measures to provide an interdisciplinary evaluation of multi-party interaction system.

Regarding repeated human–robot interaction, Jones and Schmidlin [23] explain the importance of understanding HRI beyond participants first impressions. Złotowski et al. [64] show that repeated interactions with a robot can reduce the uncanny valley effect in the perception of the robot. The repeated exposure to the robot can improve robot’s likeability and reduce its eeriness. Robins et al. [45] present the effect of long-term repeated human–robot interaction on the children with autism. Over time children got accustomed to the robot reporting more emotional significance and meaning to the experiences with the robot. However, no studies regarding the effects of repeated interactions has been found, neither in the case of human–robot turn-taking, nor in the case of multi-party HRI.

To summarize, the goals of the study presented in this paper were to:

- use an autonomous social robot,
- develop an interaction system for EMYS that acts in accordance with human multi-party turn-taking norms,
- evaluate this interaction system both in terms of performance and user assessment,
- compare the results with the basic EMYS interaction system,
- quantify the effect of first impression in the context of robot multi-party turn-taking behavior.

3 Methodology

The main purpose of this study was to verify the interaction system developed for a social robot to participate in multi-party interaction. Three research questions are considered in this paper:

RQ1 How does the robot’s adhering to human turn-taking norms impact its performance, measured in percentage of correct turn-exchanges, in terms of multi-party communication?

RQ2 How does the robot’s adhering to human turn-taking norms impact the user assessment of the robot in terms of multi-party communication?

RQ3 How does repeated interactions with the robot influence its performance and user assessment?

In this section we describe the proposed multi-party interaction system and present the design for the experiment.

3.1 Multi-party Interaction System

A multi-party interaction system was developed to extend the abilities of social robot EMYS. The basic capabilities of EMYS use a spoken dialog system for a conversational agent that relies on mostly on speech, which has proven to be enough for interactions with a single user, however it is prone to conversational errors in multi-party interaction. The purpose of the experiments was to test the proposed multi-party interaction system (M), which implemented human turn-taking behavior and multi-party capabilities, in relation to this basic interaction system (B).

The control system of social robot EMYS is comprised of a three-layer architecture: lowest, middle and highest layer [26]. The lowest layer provides an access point for actuators, sensors and external software. The middle layer implements robot's competencies, i.e. tasks that the robot is able to perform. These competencies are based on lowest layer modules or extend other competencies to carry out more complex tasks. The highest layer is where the competencies are utilized for the robot to function in a specific scenario or application, while the implementation can vary from remote control to fully autonomous control system.

The basic interaction system in EMYS utilizes speech recognition engine [26] that detects speech events and relies on pauses to segment users utterances. It does not take into account any visual cues, nor does it process them to track the turn-taking in the conversation, which, in multi-party setting, results in a number of conversational errors and a reduced smoothness of interaction. We argue that this can be significantly improved upon by using a system that detects and expresses turn-taking cues, especially through combination of gaze and speech cues.

The multi-party interaction system expands existing robot control system. In the lowest layer the extension included support for multiple microphones, an interface for tracking multiple people through the Kinect sensor, as well as user gaze detection and estimation. There were multiple competencies added to the middle layer. The robot's perception was enhanced by detection and tracking of turn-taking cues, while for the robot's expression the gaze and the speech were combined into tasks: *speak to user(s)*, *listen to user(s)*. These abilities are utilized in *Turn-taking Manager* to oversee the conversation flow by ensuring that the robot has the floor before speaking and that the proper attention is given to the users when they are talking. Finally, the *Dialog Manager* encapsulates all of the above multi-party turn-taking competencies and allows language generation for robot's

Table 1 Differences in behavior between multi-party interaction system and basic interaction system

M-multi-party interaction system	B-basic interaction system
Robot's gaze tracks the current speaker or the attention is shared equally among the conversation participants	Robot's gaze tracks the most active person according to Kinect sensor
Robot allows the users to interact with each other after each utterance	Robot waits for silence before speaking
Robot reacts if being looked at by last speaker or after prolonged silence (2 s)	Robot reacts after medium-length silence (1 s)

utterances along with interpretation of the user's responses for the robot's programming logic.

In summary, the comparison is between a spoken dialog system that uses pauses in speech for turn exchanges (basic interaction system), and a multi-modal system that combines gaze and speech to track and express turn-taking behavior (multi-party interaction system). However, rather using various turn-taking cues and other factors, the focus is placed on comparison between minimal system requirements for smooth interaction.

The experimental scenario was implemented so that in both cases the interaction could be finished successfully to the best of each system's capabilities. The differences, presented in Table 1, were in the robot's gaze patterns, patience in taking the conversational floor and recognizing when being spoken to.

3.2 Experimental Scenario

The goal of the experimental scenario was to facilitate turn-exchanges in multi-party interaction setting. Two participants interacted with the robot while playing a trivia questions game in which the robot served as a host. After each question, the participants were to consult the answer and provide it to the robot upon agreement. The setup for the experiment is presented in Fig. 2.

Below is an excerpt from the transcript of a dialog between the robot and the users.

ROBOT:

The Roman god of love was called?

- A. Amor
- B. Jupiter
- C. Eros

USER A:

I'm sorry, can you repeat the question?

ROBOT: [repeats]



Fig. 2 The experimental setup

USER B: [after consulting each other]

We decided upon Amor.

ROBOT:

Are You sure that You want to answer A – Amor?

USER A:

Yes, we are.

ROBOT:

Correct! / Sorry, the correct answer is ...

Did You know that Amor was the son of Venus and Mars? His Greek equivalent was Eros.

[waits for the users before continuing]

Next question: ...

The experimental scenario was prototyped through preliminary experiments in order to create the vocabulary and grammar necessary for user speech recognition, as well as to design robot utterances. Moreover, the preliminary experiments allowed to fine-tune the system: select latency threshold for responding, take into account the limitations of the robot sensors, such as positioning in the camera range or noise reduction and resolve unpredicted situations that could affect the study.

3.3 Experimental Design

The main goal of the experiment was to evaluate and compare both interaction systems in terms of performance (*RQ1*) and user assessment (*RQ2*). This comparison should be made on unbiased interactions with robot, i.e. first-time interactions.

The secondary goal was to measure the effect of repeated interactions (*RQ3*). The first meeting with the robot establishes some preconceptions regarding its abilities and results in participants modifying their behavior and expectations. These expectations can influence the perception of the robot in further interactions. We studied how does user assessment differ in the second (biased) interaction with the robot after the first interaction has set some expectations, in the case of both improvement (i.e. first basic, then multi-party interac-

tion system), as well as deterioration (i.e. first multi-party, then basic).

The participants were divided randomly into two groups that both interacted with robot twice but in different order. One group interacted with basic interaction system first and then with multi-party interaction system, resulting in experimental conditions *basic-first* (B1) and *multi-party-second* (M2). For the other group the order was reversed, resulting in conditions *multi-party-first* (M1) and *basic-second* (B2). Note that interactions B1 and M1 are unbiased, while interactions B2 and M2 will be biased by previous interaction.

The experimental design focuses on following aspects:

Evaluate the developed multi-party interaction system

The evaluation and comparison both interaction systems in term of performance (*RQ1*) and user assessment (*RQ2*) was done on unbiased interactions, i.e. comparing conditions B1 with M1.

Measure the effect of user expectations in repeated interactions

The effect of repeated interactions (*RQ3*) was measured by comparing biased interactions to their unbiased counterparts, i.e. B2 with B1 and M2 with M1. For example, we know for condition B2 (*basic-second*) that the user have previously interacted with multi-party interaction system (M1), which may have set high preconceptions and expectations, by comparing B2 with B1 the effect of this factors can be measured.

Do not reveal the goal of the study to the participants

The information about the true goal of the experiment can cause participants to consciously and unconsciously influence the results of the study, therefore this information should not be revealed. This effect was reduced by presenting the interaction in the form of a game.

Moreover, the robot has been awarding points for correct answers, which may have led the participants to believe that the real goal was to test their knowledge rather than the robot's communicative abilities.

Do not reveal the experimental group to the participants

The participants were not aware what the difference is between the two interactions neither with which version of the robot they are currently speaking. This a further consequence of not revealing the goal of the study.

Reduce the influence of the researchers on the results

The behavior of the researchers can also indirectly influence the results of the study, therefore, their contact with the participants should be reduced to the necessary minimum. During the recruitment process the participants have only been informed that they will participate in a game with the robot and about the estimated time it will take (up to 40 min). Any further questions that they may have had, have been answered after the completion of both interactions and filling of the surveys during the

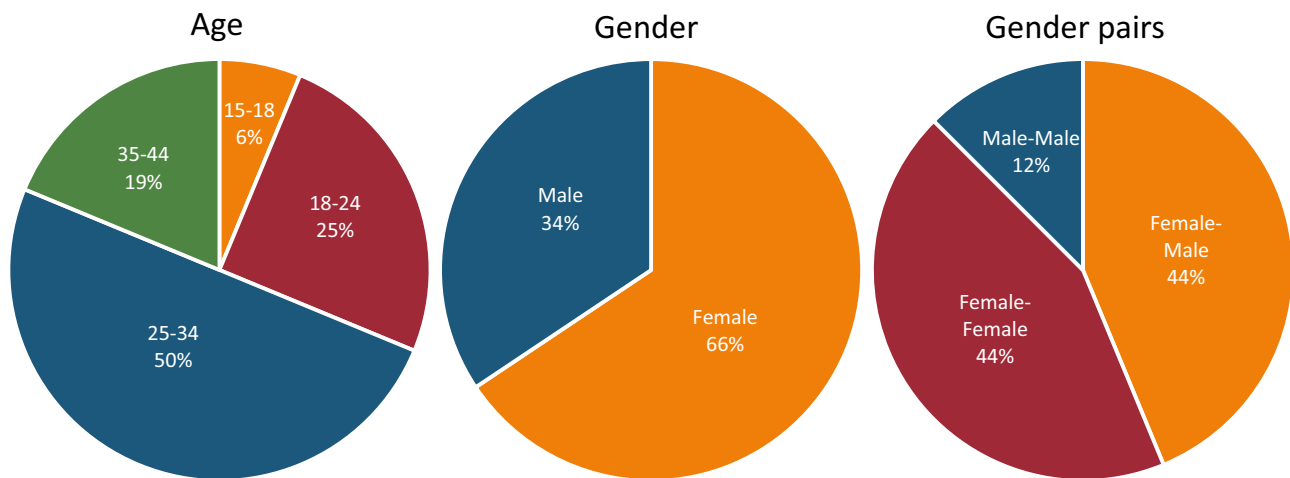


Fig. 3 Age and gender of the participants

in-depth interviews. After guiding the participant to the room and their seats, the researcher left the room and the robot took the role of the host, who explained the rules of the game and the next course of events.

Keep the duration of the experiment representative, albeit short

The length of the interaction should provide enough time for the participants to get accustomed to the situation and develop their opinion, however at the same time be relatively short to not bore or fatigue the participants. We decided upon 10–15 min for the interaction and 3–5 min for filling out the questionnaire, basing our decision on so-called ‘tv-series’ attention span. The participants interacted with the robot two times in total, filling the questionnaire after each interaction.

Recruit participants with ease

Larger groups of people are more difficult to recruit (and organize) than smaller ones. The possible technical limitations of the robot’s sensors were also taken into consideration. On this basis, we decided upon 3-party interaction between two people and a robot. The location selected for the experiment and recruitment methods, described below, also helped with this aspect.

3.4 Experimental Procedure

The experiments were conducted near the city center, outside the main university campus, which resulted in increased number of participants, as well as provided diversity among them. The rooms where the experiment took place were prepared to ensure the neutral appearance by removal of elements that were suggestive or distracting.

The participants were recruited by two methods: online internet registration form sent through social media (snow-

ball method) and by inviting pedestrians to take part in the experiment. In total, 32 people took part in the experiment, 21 were female and 11 were male. The age of participants was between 15 and 44 with an average of 29 years old. The distribution of gender and age of the participants is presented in Fig. 3.

The participants interacted with the robot over the course of playing a trivia questions game. The goal of the participants was to select the correct answer. There were 12 randomly selected questions asked in total (4 easy, 4 medium and 4 hard) and the robot presented three possible answers. The duration of the game was in the range of 10 to 15 min. After finishing the game, the participants were moved to stations, where they were asked to fill out the questionnaire, which took between 3 and 5 min. After the first experiment, the participants were asked to interact with the robot again. Throughout the whole experiment the robot was working autonomously.

User assessment of the robot was done through questionnaires after each interaction with the robot followed by in-depth interviews with the participants. The questionnaires designed for this study consisted of 15 questions graded on a 7-point Likert scale, based on approaches presented in [7,56]. The aim of the question was to inquire about the perceived communication skills of the robot, its intuitiveness, politeness and expressiveness, as well as user expectations and reception of the robot. For full list of questions please consult Table 3.

The questionnaire filling stations were placed in a separate room from the robot. The role of the researchers has been minimized to welcoming participants into the room, seating them in their respective places and showing them to the questionnaire filling station after the interaction. Neither during the interaction with the robot, nor during the filling of questionnaires was the researcher present in the room. The

Table 2 Performance of the robot interaction systems as a percentage of correct turn-exchanges during the experimental scenario in unbiased conditions (B1 and M1) with breakdown of conversational errors

Interaction system	Correct behavior	Repetition needed	Response wrongly understood	Prolonged silence	Unnecessary question repeat	Reacted without prompt	Performance
Basic-first (B1)	9.86 (± 0.34)	4.43 (± 0.89)	3.14 (± 0.62)	0.14 (± 0.34)	1.43 (± 0.50)	0.14 (± 0.34)	51.5% ($\pm 2.0\%$)
Multi-party-first (M1)	13.00 (± 2.63)	1.38 (± 1.93)	0.38 (± 0.72)	0.00 (± 0.00)	0.50 (± 0.52)	0.88 (± 0.96)	80.5% ($\pm 9.4\%$)

in-depth survey was conducted only after both interactions took place. The survey consisted of open discussion between both participants and the experimenter.

The total size of the data set was: 16 pairs that interacted with the robot twice, which resulted in 32 interactions and 64 filled questionnaires.

4 Results

4.1 Performance

The performance of the interaction systems was calculated as a percentage of correct responses during the whole interaction. The following conversational errors were considered:

- User had to repeat himself
- User response wrongly interpreted
- Prolonged ‘awkward’ silence
- Robot repeated the question without request
- Robot spoke out of turn

Two judges were asked to review the video recordings of the experiments and count occurrences of correct turn-exchanges or the above error types. The judges were otherwise not involved in the study, they were not aware of the research hypotheses or experimental conditions (i.e. basic/multi-party), nor could they differentiate between these conditions. Each interaction, 8 for condition B1 and 8 for condition M1, was from 10 to 15 min long and contained from 15 to 22 places that needed classification, totaling almost 300 observations. Independently, the judges annotated a subset of 30 randomly selected segments (10% of total observations), to verify inter-rater reliability, which resulted in Cohen’s Kappa $K = 0.76$ indicating acceptable agreement.

The robot equipped with the multi-party interaction system is expected to show improved performance in comparison with basic interaction system (RQ1). Indeed, as shown in Table 2, in an unbiased conditions (B1 and M1) the performance of multi-party interaction system ($M = 80.5\%$, $SD = 9.4\%$) was much higher than the performance of basic interaction system ($M = 51.5\%$, $SD = 2.0\%$). The

analysis of variance has confirmed that this difference was significant [$F(1, 30) = 85.56$, $p < 0.001$].

In addition, the length of the interaction (measured in total places to classify by the judges) was also shorter in multi-party condition ($M = 16.12$, $SD = 2.71$) than in basic condition ($M = 19.12$, $SD = 0.81$), which was significant [$F(1, 30) = 18.07$, $p < 0.001$]. This coincides with lower number of repetitions that the robot was asked to perform.

4.2 User Assessment

User assessment serves as a validation that the multi-party interaction system improved how the robot is seen by its users in comparison with the basic interaction system (RQ2). The assessment of the robot was done through 15 questions regarding the robot’s perceived abilities and user experience, which were rated on a 7-point Likert scale. The verification of multi-party interaction system focused on unbiased interactions: conditions basic-first (B1) and multi-party-first (M1), in which the participants talked to the robot for the first time and should not have any previous opinions and preconceptions about the robot. The user responses are presented in Table 3, as well as visually by plots in Figs 4 and 5, see conditions B1, M1 and (M1–B1).

Analysis of variance with Bonferroni correction for $\alpha = 0.05/15 = 0.0033$ has shown significant differences in 10 cases between multi-party-first and basic-first conditions (M1–B1):

- Q1: EMYS met my expectations (2.38, $p < 0.001$)
- Q3: EMYS made a good impression on me (1.25, $p = 0.002$)
- Q7: EMYS correctly reacts to the environment (1.50, $p < 0.001$)
- Q8: EMYS is cooperative (3.38, $p < 0.001$)
- Q9: EMYS can communicate (3.13, $p < 0.001$)
- Q10: EMYS can assess a situation (2.00, $p < 0.001$)
- Q11: EMYS understands when being spoken to (2.63, $p < 0.001$)
- Q13: EMYS completed his task (2.38, $p < 0.001$)
- Q15: EMYS is intuitive (1.88, $p < 0.001$)

Table 3 User assessment questionnaire responses on 7-point Likert scale (−3, 3) and supporting indicators with corresponding significance tests

	M1	B1	B2	M2	F(3,60)	p	M1-B1	B2-B1	M2-M1	M2-B2
Q1: Emys has met my expectations	1.75 (±0.68)	-0.63 (±1.45)	1.00 (±0.73)	0.25 (±0.45)	19.97	<0.001	2.38 p<0.001	1.63 p<0.001	-1.50 p<0.001	-0.75 p=0.103
Q2: I like Emys	1.88 (±0.96)	1.00 (±0.52)	1.63 (±0.89)	1.38 (±0.50)	4.02	0.011	0.88 p=0.008	0.63 p=0.093	-0.50 p=0.239	-0.25 p=0.778
Q3: Emys made a good impression on me	2.38 (±0.72)	1.13 (±0.96)	1.38 (±1.15)	1.75 (±0.86)	5.43	0.002	1.25 p=0.002	0.25 p=0.873	-0.63 p=0.241	0.38 p=0.669
Q4: Emys has large knowledge	1.00 (±1.46)	0.00 (±0.52)	0.88 (±1.59)	0.13 (±0.34)	3.50	0.021	1.00 p=0.057	0.88 p=0.117	-0.88 p=0.117	-0.75 p=0.221
Q5: Emys is interesting to me	2.00 (±0.73)	1.13 (±1.20)	1.88 (±0.62)	1.50 (±0.52)	3.80	0.015	0.88 p=0.017	0.75 p=0.054	-0.50 p=0.311	-0.38 p=0.562
Q6: Emys behaves like a human	0.50 (±1.15)	-0.88 (±1.59)	-0.13 (±1.09)	-0.75 (±1.91)	2.95	0.040	1.38 p=0.050	0.75 p=0.481	-1.25 p=0.089	-0.63 p=0.630
Q7: Emys reacts correctly to the environment	1.50 (±0.89)	0.00 (±0.52)	0.75 (±1.77)	0.50 (±0.52)	5.60	0.002	1.50 p<0.001	0.75 p=0.197	-1.00 p=0.046	-0.25 p=0.908
Q8: Emys is cooperative	2.13 (±0.81)	-1.25 (±1.91)	0.50 (±1.55)	0.88 (±0.96)	16.29	<0.001	3.38 p<0.001	1.75 p=0.004	-1.25 p=0.041	0.38 p=0.869
Q9: Emys can communicate	1.75 (±0.68)	-1.38 (±1.26)	0.75 (±1.13)	0.88 (±0.81)	28.40	<0.001	3.13 p<0.001	2.13 p<0.001	-0.88 p=0.073	0.13 p=0.985
Q10: Emys can assess a situation	0.75 (±0.86)	-1.25 (±1.24)	-0.38 (±1.78)	-0.38 (±0.72)	7.21	<0.001	2.00 p<0.001	0.88 p=0.190	-1.13 p=0.049	0.00 p=1.000
Q11: Emys understands when being spoken to	1.38 (±0.72)	-1.25 (±1.84)	0.00 (±1.46)	0.38 (±0.89)	11.02	<0.001	2.63 p<0.001	1.25 p=0.043	-1.00 p=0.145	0.38 p=0.849
Q12: Emys is well-mannered	2.50 (±0.52)	2.00 (±1.26)	1.88 (±1.09)	1.75 (±1.00)	1.71	0.175	0.50 p=0.501	-0.13 p=0.985	-0.75 p=0.162	-0.13 p=0.985
Q13: Emys completed his task	2.13 (±1.09)	-0.25 (±1.44)	0.88 (±1.31)	1.75 (±0.45)	13.81	<0.001	2.38 p<0.001	1.13 p=0.034	-0.38 p=0.787	0.88 p=0.141
Q14: Emys understands complex situations	0.50 (±0.73)	-0.50 (±1.03)	-0.13 (±0.96)	-0.38 (±0.72)	4.18	0.009	1.00 p=0.010	0.38 p=0.618	-0.88 p=0.030	-0.25 p=0.849
Q15: Emys is intuitive	0.50 (±0.89)	-1.38 (±1.26)	0.00 (±0.52)	-0.63 (±0.89)	12.23	<0.001	1.88 p<0.001	1.38 p<0.001	-1.13 p=0.006	-0.63 p=0.236

Statistical significance marked for $p < \alpha = 0.05/15 = 0.0033$

Note the magnitude of the changes, e.g. EMYS cooperativeness (Q8) has shifted more than 3 points, from 'rather poor' ($M = -1.25$, $SD = 1.91$) to 'well' ($M = 2.13$, $SD = 0.81$). The results confirm that the robot had better communicative and interactive abilities, was better perceived by the users and fit their expectations better when it was equipped with the multi-party interaction system rather than the basic interaction system.

In addition, it is worth mentioning the questions that didn't reach statistical significance due to alpha correction, but still had $p < 0.05$, which makes them strong candidates for further study.

- Q2: I like EMYS (0.88, $p = 0.008$)
- Q5: EMYS is interesting to me (0.88, $p = 0.017$)

- Q6: EMYS behaves like a human (1.38, $p = 0.050$)
- Q14: EMYS understands complex situations (1.00, $p = 0.010$)

4.3 Repeated Interactions

The effect of repeated interactions was measured both in performance and user assessment (RQ3). The participants interacted with the robot two times, each in time with a different interaction system, however the order of the interactions was randomized. Specifically, half of participants interacted with condition *basic-first* (B1) followed by *multi-party-second* (M2), while the other half started with condition *multi-party-first* (M1) followed by *basic-second* (B2). If the order of interaction was not of importance the ratings should

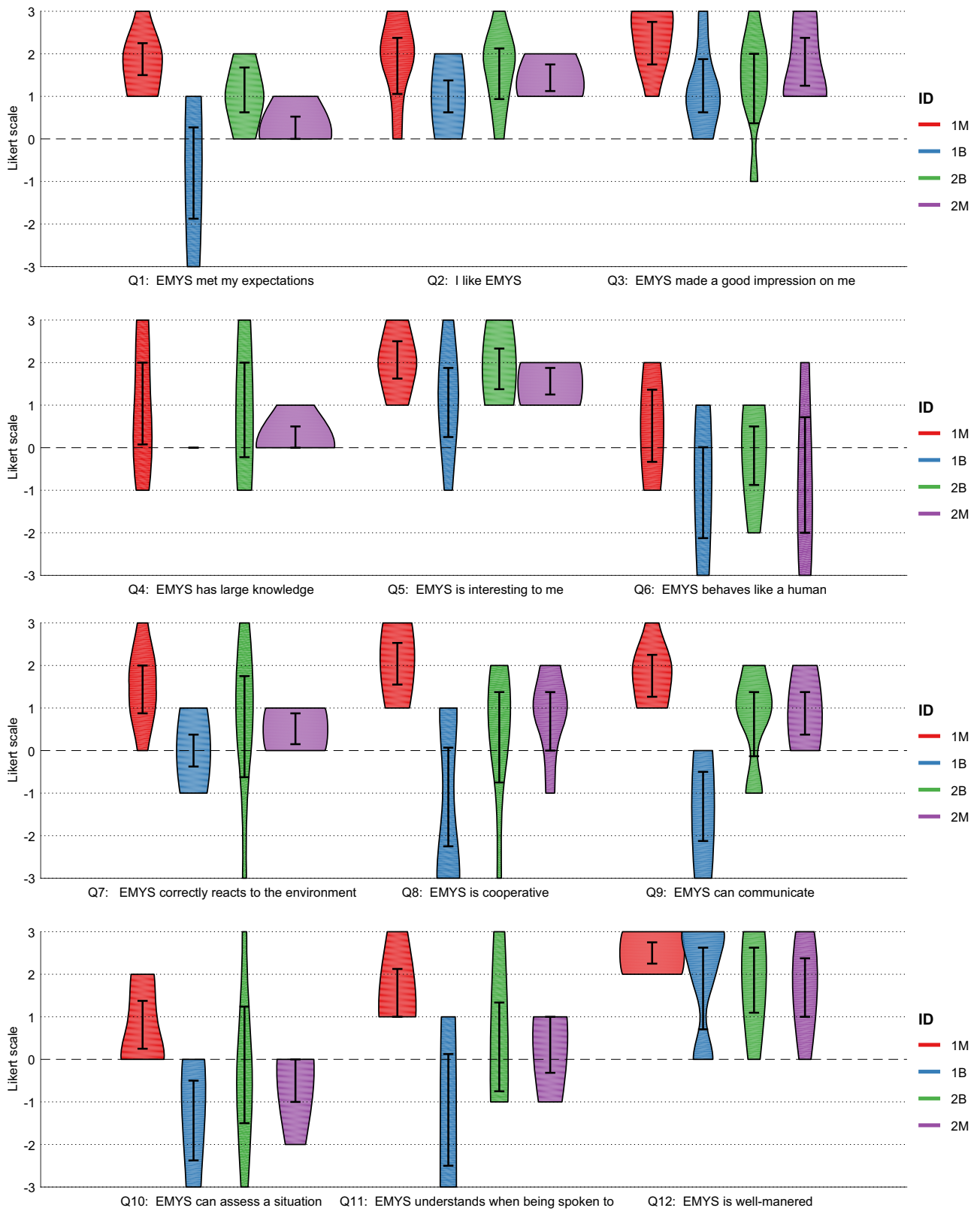


Fig. 4 User assessment questionnaire responses (questions Q1–Q12). Confidence intervals plotted for $\alpha = 0.05$

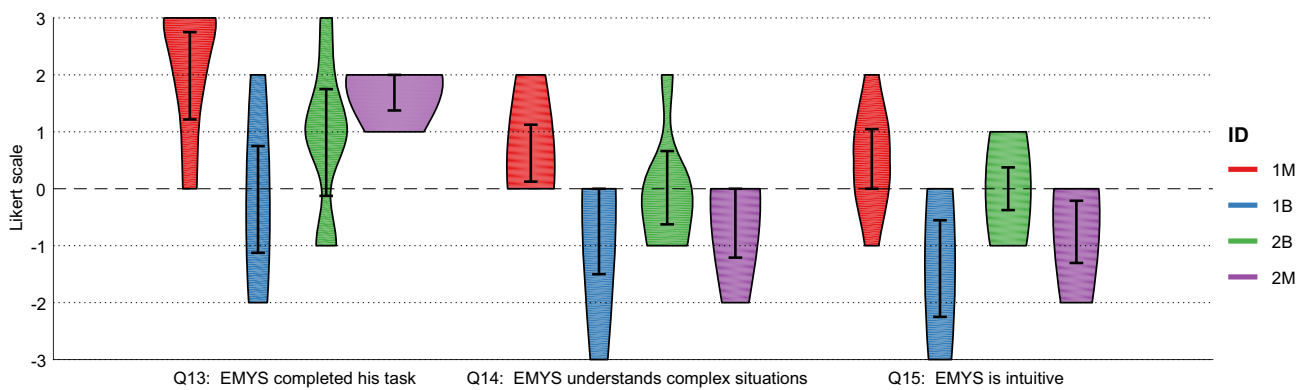


Fig. 5 User assessment questionnaire responses (questions Q13–Q15). Confidence intervals plotted for $\alpha = 0.05$

be equal, i.e. $M1 = M2$, $B1 = B2$. However, this does not hold—some aspects of interaction have been assessed differently. How do these impressions and convictions of users manifest themselves? For further reference, we have named the observed effects as *benefit of the doubt* and *caution*.

4.3.1 Benefit of the Doubt (B2–B1)

The difference ($B2 - B1$) will determine the effect of the users' first impressions after interaction multi-party-first (M1) that preceded condition basic-second (B2). If multi-party interaction system M has been evaluated positively, a kind of *benefit of the doubt* can be expected, i.e. since the robot previously had some features (in condition M1), the participants may be convinced that the robot still has these features, even if it does not show them (condition B2).

The performance in the basic-second condition (B2) ($M = 42.5\%$, $SD = 15.6\%$) was worse than the basic-first condition (B1) ($M = 51.5\%$, $SD = 2.0\%$), which was significant [$F(1, 30) = 5.38$, $p < 0.027$].

The user assessment, as shown in Table 3 (column B2–B1), yielded significant differences ($p < \alpha = 0.05/15 = 0.0033$) for following questions:

- Q1: EMYS met my expectations (1.63, $p < 0.001$)
- Q9: EMYS can communicate (2.13, $p < 0.001$)
- Q15: EMYS is intuitive (1.38, $p < 0.001$)

The Cronbach's $\alpha = 0.810$ for measuring the effect of benefit of the doubt confirms the internal consistency for the above questions.

The questions below didn't reach statistical significance, but are worth mentioning for further research due to $p < 0.05$:

- Q8: EMYS is cooperative (1.75, $p = 0.004$)
- Q11: EMYS understands when being spoken to (1.25, $p = 0.043$)

- Q13: EMYS completed his task (1.13, $p = 0.034$)

The additional inclusion of these questions raises the Cronbach's $\alpha = 0.906$

This confirms that repeated interactions affect the performance of the interaction system, as well as the user assessment of the robot. Curiously enough, the objective performance in the biased basic-second condition was worse by 9% than in basic-first condition, but user assessment of the robot was better in the range 1.13–2.13 point on a 7-point scale.

4.3.2 Caution (M2–M1)

In contrast, the difference ($M2 - M1$) will describe the impact of the first impression after interaction basic-first (B1). We expect increased conservativeness and *caution* when assessing the robot in multi-party-second condition (M2), which should manifest in lower scores than the unbiased interaction (M1).

In this case the performance in the multi-party-second condition (M2) ($M = 69.8\%$, $SD = 11.8\%$) was also worse than the multi-party-first condition (M1) ($M = 80.5\%$, $SD = 9.4\%$), the difference was significant [$F(1, 30) = 4.45$, $p < 0.043$].

In terms of user assessment as presented in Table 3 (column M2–M1) only 1 question reached significance for $p < \alpha = 0.05/15 = 0.0033$:

- Q1: EMYS met my expectations (-1.50 , $p < 0.001$)

Other questions worth considering for further study due to $p < 0.05$ are:

- Q7: EMYS correctly reacts to the environment (-1.00 , $p = 0.046$)
- Q8: EMYS is cooperative (-1.25 , $p = 0.041$)
- Q10: EMYS can assess a situation (-1.13 , $p = 0.049$)

- Q14: EMYS understands complex situations (-0.88 , $p = 0.030$)
- Q15: EMYS is intuitive (-1.13 , $p = 0.006$)

This set of questions resulted in Cronbach's $\alpha = 0.832$ for the effect of caution.

The performance deteriorated 10.7%. It can be observed that the overall set of questions mostly overlaps with the previous case with the differences at level of 0.88–1.50. This time, however, the worse performance coincides with worse user assessment.

5 Discussion

The aim of the study was to evaluate the proposed multi-party interaction system in terms of performance (RQ1) and user assessment (RQ2). The secondary objective was to study how multiple interactions with the robot affect the user assessment of the robot (RQ3).

5.1 Evaluation of Multi-party Interaction System

The multi-party interaction system has shown a significant improvement in comparison to basic interaction system. The basic interaction system relies on speech cues for turn-exchanges while the multi-party system provided turn-taking mechanisms by combining speech and gaze cues, which resulted in performance improvement from 51.5% to 80.5%. This manifests in greatly reduced number of errors and shorter length of the conversation, which accounts for more fluent interaction (RQ1).

The analysis of the questionnaires revealed that, on the basis of user assessment, the developed multi-party interaction system was perceived significantly better than the basic interaction system in 10 different aspects (RQ2). The changes in assessment were in the range of 1 to 3 points in 7-point Likert scale, mostly showing that the multi-party interaction system exhibited some trait that the basic interaction system lacked, for example the perception of EMYS ability to communicate has shifted from 'rather poor' to 'well'. The participants rated the robot equipped with multi-party interaction system as, among others, more communicative, perceptive and willing to cooperate. These aspects can be described as indigenous elements of multi-party communication. Apart from communication skills, it was also stated that EMYS equipped with the multi-party interaction system has made an overall better impression, satisfied the expectations of the users better (in the context of user expectations of social robots) and also performed his task better (the role of the game host). It is evident that the ability to communicate naturally has a significant impact on the perception of social robots and that these skills are perceived as crucial for role

of a game host. In addition, during the in-depth interviews conducted after the experiments, the participants described the robot with the multi-party interaction system as listening actively and showing attention, pointing towards the robot's gaze as a factor that created such an impression.

A larger difference was expected in the assessment of the robot manners, but in both cases they were rated about equally high, in the range of 2.00–2.50 which places them between 'well' (2) and 'exceptional' (3). This is probably due to the nature of the interaction with the robot, as well as the patient (respectful) way of taking the floor. Placing the robot in a conflict scenario could lead to more conversational errors (i.e. interruptions) and, in the context of turn-taking, good manners reflect in the way of tactfully resolving such situations. In a similar way, the latency in taking the floor was set relatively long and its reduction would increase the number of interruptions and misunderstandings during the conversation [40]. It is possible that this would have a significant impact on the perception of the other aspects of the robot, especially on the robot's likeability.

5.2 Repeated Interactions

The other part of the analysis concerns the effect of repeated interactions with the robot on the robot's assessment. It shows that even the first interaction with the robot can leave an impression on the user that will affect the user in later interactions with robot and influence his/hers assessment (RQ3). Positive and negative effect of magnitude in the range of 0.88–2.13 on the 7-point Likert scale were observed, which were defined as a benefit of the doubt and caution.

Benefit of the doubt occurred after initial interaction with the multi-party interaction system (condition M1) in the user second interaction that used the basic interaction system (B2). The participants rated the basic interaction system better in this biased condition B2 than in the unbiased condition B1, in which the user interacted with the basic interaction system for the first time. This effect is even more interesting if one takes into account that the measured performance of the basic interaction system in the condition B2 was actually lower than in the condition B1 (42.5% vs 51.5%).

Caution, a situation opposite to the above, occurs after initial interaction with the basic interaction system (condition B1) when assessing multi-party interaction system in the second interaction (M2). It was observed that the participants are more conservative in assessing the multi-party interaction system in biased condition M2 than if it was the first encounter with the robot as it was in condition M1. However, in this case the worse performance between conditions multi-party-second M2 and multi-party-first M1 coincides with worse user assessment (69.8% vs 80.5%).

Both situations are consistent and mostly symmetric. There are no cases in which the multi-party interaction

system would cause caution or the basic version would develop benefit of the doubt in the users. In terms of size, these two effects are mostly equal (consider comparison $|B2 - B1| - |M2 - M1|$) with the exception of question ‘Q15: EMYS can communicate’, in which the benefit of the doubt left a seemingly stronger impression of 1.25 points.

This means that the initial opinion about the robot is difficult to change once it is established. The bias is so strong that once the user witnesses some trait exhibited by a robot, he/she will still continue attributing this trait to the robot, even if the objective measurement of performance proves otherwise. This seems to be a result of attribution bias applied to social robots.

In addition, this drop in performance may indicate that the users try adapt to the way the robot communicates, even if this unconscious. It shows that if the way the robot communicates changes, the users may need additional time to get used to it.

As a consequence, for experiments using social robots, it is recommended to take into account these factors during the experimental design process, paying special attention to multiple experiments with a social robot involving the same research groups, especially across different studies. At the same time, from the perspective of social robot as a commercial product, this study ascertains that it is difficult to influence the established opinion of the users, so if achieving a particular impression is needed, the effects of *benefit of doubt* and *caution* should be considered when introducing subsequent versions of the robot.

5.3 Observations

During the analysis of video recordings from experiments, it was noticed that people gradually examine the perceptual and cognitive abilities of the robot. Basing on these observations, the users build their model of the robot’s capabilities. For example, some people tried to elicit a reaction from the robot by joking, to see if the robot would react to humor, when this attempt was unsuccessful they adapted to the state of the robot’s abilities and no longer used jokes in the messages directed towards the robot, but were still using them in communication with the other person. This supports that people tend to instinctively verify the communication capabilities of the other party, social robots included, and then adjust their way of communication. Consequently, people construct their own model of robot competence and it is possible that the re-evaluation of this model in the case of these competencies changing is difficult and may take time, which is an important factor to consider when adding new functions to existing robots.

The in-depth interviews that followed the experiment has shown that the most noticed aspect of the interaction was the gaze of the robot tracking the current speaker. This means that the ability to actively listen, and thus provide feedback

to the speaker is an important part of communication. Lack of this behavior may cause the robot to be ignored during the interaction, and thus not treated as its full participant. Such situation took place in the research described in [19]. Moreover, the backchannel feedback is also presented verbally, which symbolizes an understanding of the current statement (‘aha’, ‘mhm’). These intrusions do not signify the intention to take the floor, on the contrary, they encourage the speaker to continue; in a way, this is the action of giving the floor in advance. In our opinion, the issue of expressing such feedback signals and their impact on the speaker is a promising direction for future research

In both experimental cases the robot reacted emotionally to the responses give by the participants, i.e. acted happy when the answer was correct and sad when the answer was incorrect, which served as a way of presenting empathy. As a result the robot could have been perceived more positively in terms of likeability.

5.4 Limitations

The following factors were not controlled for in the experiments and can be a subject for further research.

Alternative measures of performance

In the domain of dialog systems two common measures of performance are accuracy, used in this study, and latency. Latency has been shown to affect conversation in the following ways: awaiting too long to respond can prolong the conversation and impact its fluency, while responding too quickly can cause interruptions and misunderstanding [40]. In case of evaluating and comparing different variants of multi-party interaction system the researchers should consider using latency.

Differences in communication between friends, acquaintances, strangers and enemies

Because the participants were recruited in pairs, it should be assumed that they knew each other before the study and were on friendly terms. Considering various social robot working environments, these kinds of situations are more common than multi-party conversations including two strangers or two people in direct conflict. The interpersonal relationship could influence the way they communicate, as well as their perception of the robot as a result of a positive association. An opposite situation would be a scenario in which the robot acts as a judge or an arbiter between two opposite parties. However, this scenario does not encourage mutual communication between parties, but rather places the robot as an intermediary.

Participants influencing each other

The participants were not separated during the task of

filling the questionnaires. This could have caused strong correlations in the questionnaire scores in each pair of participants that interacted with the robot.

Individual differences between participants

The analysis did not take into account participants gender, age, personality, understanding of the technology, hobbies etc. It is difficult to say to what extent these factors can influence the results of the study, but the strongest candidate for a more in-depth assessment would be the personality types. Research indicates that people with different personalities prefer different traits in their companions [15], therefore these differences can reflect in the way they communicate. The factors to consider are the patterns of open/closed, social/asocial people in relation to multi-party interaction.

Group dynamics and interpersonal relationships

In recent research Fraune et al. [18] observed 2714 people interacting with the social robot in a naturalistic setting and reported how group presence, group cohesiveness and group social norms can influence the human–robot interaction. We argue that interpersonal relationships, such as: family, friendship, work, as well as age and sex differences, can be a strong factor in turn-taking, especially in the case of our experimental scenario, which required reaching a consensus in selecting the answer. For a social robot to operate in (and among) such relationships is an emerging and promising research direction.

Physiological measurements

The gathered survey data were not verified through any physiological measurements.

6 Conclusion

In this paper we present the experimental verification of the developed social robot multi-party interaction system, from the perspective of both performance evaluation and users assessment of the interaction with this system. In the context of social robotics this kind of simultaneous two-sided evaluation is rarely performed due its difficulties. The multi-party interaction system improved the performance, expressed as a percentage of correct turn-exchanges, of the basic interaction system of robot EMYS from 51.5% to 80.5%, which resulted in more fluent interaction due to reduced number of errors and shorter length of the conversation.

User feedback assessment based on the analysis of surveys has shown that the multi-party interaction system makes the robot perceived as more communicative, cooperative, intuitive, fitting the user expectations and making an overall better impression.

The other problem studied was the effect of repeated human–robot interaction on the user assessment of the robot. It was shown that the interaction with the robot may leave a

lasting impression on the user, which impacts the perception of the robot in future interactions. This effect can be both positive, i.e. benefit of the doubt, or negative, i.e. caution, to the assessment of the robot. We advise to take this effect into consideration either during the social experiments with the robot involving the same participants or in the process of updating the existing social robots. If the goal is not specifically to measure an individual user opinion, but to obtain an objective/unbiased assessment, the experimental design should refrain from using participants that had any previous contact with the robot, even across different experiments. Moreover, in case of social robot development, it may be that a set of small updates to the robot can have a diminished cumulative effect on users than a larger combined update.

As open directions for further research, we point towards including more people in the conversation, making it more dynamic and changing the role of the robot in the interaction. In particular, attention should be paid to examining a greater range of different social situations in which the robot can operate. The relationship between the robot and its users may be symmetrical or asymmetrical. The goals of the robot may vary, which can create scenarios of cooperation or conflict. The hierarchy between the interlocutors may differ, as well as the use of formal and informal language. Finally, the available means of expression may be limited (e.g. one of the users is available only by voice, but not visually, which may take place during teleconferences) or extended (e.g. telepresence on a tv-screen or a mobile phone). The overall aim of the HRI research should gradually shift from modeling specific use cases into describing general social situations towards a coherent model of (multi-party) interaction.

Funding This research was supported by Grant No. 2012/07/N/ST7/03308 awarded by the National Science Centre of Poland.

Compliance with Ethical Standards

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

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Al Moubayed S, Skantze G (2011) Turn-taking control using gaze in multiparty human–computer dialogue: effects of 2D and 3D displays. In: Proceedings of the international conference on auditory-visual speech processing AVSP, Florence, Italy

2. Bartneck C, Kanda T, Mubin O, Al Mahmud A (2009) Does the design of a robot influence its animacy and perceived intelligence? *Int J Soc Robot* 1(2):195–204
3. Bartneck C, Nomura T, Kanda T, Suzuki T, Kennsuke K (2005) A cross-cultural study on attitudes towards robots. In: *HCI international*
4. Bertel LB (2011) Peers: persuasive educational and entertainment robotics. Ph.D. thesis, Aalborg University, Aalborg, Denmark
5. Bohus D, Horvitz E (2009) Dialog in the open world: platform and applications. In: *Proceedings of the 2009 international conference on multimodal interfaces, ICMI-MLMI '09*. ACM, New York, USA, pp 31–38. <https://doi.org/10.1145/1647314.1647323>
6. Bohus D, Horvitz E (2010) Facilitating multiparty dialog with gaze, gesture, and speech. In: *International conference on multimodal interfaces and the workshop on machine learning for multimodal interaction, ICMI-MLMI '10*. ACM, New York, USA, pp 5:1–5:8. <https://doi.org/10.1145/1891903.1891910>
7. Bohus D, Horvitz E (2011) Multiparty turn taking in situated dialog: study, lessons, and directions. In: *Proceedings of the SIGDIAL 2011 conference*. Association for Computational Linguistics, pp 98–109
8. Breazeal C (2003) Toward sociable robots. *Robot Auton Syst* 42(3–4):167–175
9. Bruce A, Nourbakhsh I, Simmons R (2002) The role of expressiveness and attention in human–robot interaction. In: *Proceedings 2002 IEEE international conference on robotics and automation, ICRA '02*, vol 4. IEEE, pp 4138–4142
10. Castellano G, Paiva A, Kappas A, Aylett R, Hastie H, Barendregt W, Nabais F, Bull S (2013) Towards empathic virtual and robotic tutors. In: *International conference on artificial intelligence in education*. Springer, pp 733–736
11. Chao C, Thomaz AL (2010) Turn taking for human–robot interaction. In: *AAAI fall symposium: dialog with robots*
12. Deshmukh A, Castellano G, Kappas A, Barendregt W, Nabais F, Paiva A, Ribeiro T, Leite I, Aylett R (2013) Towards empathic artificial tutors. In: *2013 8th ACM/IEEE international conference on human–robot interaction (HRI)*. IEEE, pp 113–114
13. Dougherty EG, Scharfe H (2011) Initial formation of trust: designing an interaction with Geminoid-DK to promote a positive attitude for cooperation. In: *International conference on social robotics*. Springer, pp 95–103
14. Duncan S (1972) Some signals and rules for taking speaking turns in conversations. *J Personal Soc Psychol* 23(2):283–292
15. Dziergwa M, Kaczmarek M, Kaczmarek P, Kędziński J, Wadasztyłowska K (2018) Long-term cohabitation with a social robot: a case study of the influence of human attachment patterns. *Int J Soc Robot* 10(1):163–176
16. Feil-Seifer D, Mataric MJ (2005) Defining socially assistive robotics. In: *9th international conference on rehabilitation robotics, 2005. ICORR 2005*. IEEE, pp 465–468
17. Foster ME, Gaschler A, Giuliani M (2017) Automatically classifying user engagement for dynamic multi-party human–robot interaction. *Int J Soc Robot* 9(5):659–674
18. Fraune M, Sabanovic S, Kanda T (2019) Human group presence, group characteristics, and group norms affect human–robot interaction in naturalistic settings. *Front Robot AI* 6:48
19. Fukuda H, Kobayashi Y, Kuno Y, Yamazaki A, Ikeda K, Yamazaki K (2016) Analysis of multi-party human interaction towards a robot mediator. In: *2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN)*. IEEE, pp 17–21
20. Fukuhara Y, Nakano Y (2011) Gaze and conversation dominance in multiparty interaction. In: *2nd workshop on eye gaze in intelligent human machine interaction*, vol 9
21. Han J (2010) Robot-aided learning and r-learning services. In: *Human–robot interaction*. InTech
22. Hanington B, Martin B (2012) *Universal methods of design*. Rockport Publishers, Essex
23. Jones KS, Schmidlin EA (2011) Human–robot interaction: toward usable personal service robots. *Rev Hum Factors Ergon* 7(1):100–148
24. Kędziński J (2014) *System sterowania robota społecznego* (eng. social robot control system). Ph.D. thesis, Wrocław University of Science and Technology
25. Kędziński J, Kaczmarek P, Dziergwa M, Tchoń K (2015) Design for a robotic companion. *Int J Humanoid Robot* 12(01):1550007
26. Kędziński J, Muszyński R, Zoll C, Oleksy A, Frontkiewicz M (2013) EMYS—emotive head of a social robot. *Int J Soc Robot* 5(2):237–249. <https://doi.org/10.1007/s12369-013-0183-1>
27. Kim ES, Paul R, Shic F, Scassellati B (2012) Bridging the research gap: making HRI useful to individuals with autism. *J Hum Robot Interact* 1(1):26–54
28. Kondo Y, Takemura K, Takamatsu J, Ogasawara T (2013) A gesture-centric android system for multi-party human–robot interaction. *J Hum Robot Interact* 2(1):133–151
29. Kose-Bagci H, Dautenhahn K, Nehaniv CL (2008) Emergent dynamics of turn-taking interaction in drumming games with a humanoid robot. In: *RO-MAN 2008. The 17th IEEE international symposium on robot and human interactive communication*. IEEE, pp 346–353
30. Lee KM, Jung Y, Kim J, Kim SR (2006) Are physically embodied social agents better than disembodied social agents?: The effects of physical embodiment, tactile interaction, and people's loneliness in human–robot interaction. *Int J Hum Comput Stud* 64(10):962–973
31. Levinson SC (2016) Turn-taking in human communication—origins and implications for language processing. *Trends Cogn Sci* 20(1):6–14
32. Levinson SC, Torreira F (2015) Timing in turn-taking and its implications for processing models of language. *Front Psychol* 6:731
33. Li L, Xu Q, Tan YK (2012) Attention-based addressee selection for service and social robots to interact with multiple persons. In: *Proceedings of the workshop at SIGGRAPH Asia*. ACM, pp 131–136
34. Matsuyama Y, Akiba I, Fujie S, Kobayashi T (2015) Four-participant group conversation: a facilitation robot controlling engagement density as the fourth participant. *Comput Speech Lang* 33(1):1–24
35. Mori M, MacDorman KF, Kageki N (2012) The uncanny valley [from the field]. *IEEE Robot Autom Mag* 19(2):98–100
36. Mumm J, Mutlu B (2011) Human–robot proxemics: physical and psychological distancing in human–robot interaction. In: *Proceedings of the 6th international conference on human–robot interaction*. ACM, pp 331–338
37. Mutlu B, Shiwa T, Kanda T, Ishiguro H, Hagita N (2009) Footing in human–robot conversations: how robots might shape participant roles using gaze cues. In: *Proceedings of the 4th ACM/IEEE international conference on human robot interaction*. ACM, pp 61–68
38. Oreström B (1983) *Turn-taking in English conversation*, vol 66. Krieger Pub Co, Malabar
39. Pereira AT, Prada R, Paiva A (2014) Improving social presence in human-agent interaction. In: *Proceedings of the 32nd annual ACM conference on human factors in computing systems*. ACM, pp 1449–1458
40. Raux A, Eskenazi M (2009) A finite-state turn-taking model for spoken dialog systems. In: *Proceedings of human language technologies: the 2009 annual conference of the North American chapter of the association for computational linguistics, NAACL '09*. Association for Computational Linguistics, Stroudsburg, PA, USA, pp 629–637
41. Ribeiro T, Paiva A (2012) The illusion of robotic life: principles and practices of animation for robots. In: *2012 7th ACM/IEEE*

- international conference on human–robot Interaction (HRI). IEEE, pp 383–390
42. Ribeiro T, Vala M, Paiva A (2012) Thalamus: closing the mind-body loop in interactive embodied characters. In: International conference on intelligent virtual agents. Springer, pp 189–195
 43. Richter V, Carlmeyer B, Lier F, Meyer zu Borgsen S, Schlangen D, Kummert F, Wachsmuth S, Wrede B (2016) Are you talking to me?: Improving the robustness of dialogue systems in a multi party HRI scenario by incorporating gaze direction and lip movement of attendees. In: Proceedings of the fourth international conference on human agent interaction. ACM, pp 43–50
 44. Riek LD (2012) Wizard of Oz studies in HRI: a systematic review and new reporting guidelines. *J Hum Robot Interact* 1(1):119–136
 45. Robins B, Dautenhahn K, Te Boekhorst R, Billard A (2004) Effects of repeated exposure to a humanoid robot on children with autism. In: Keates S, Clarkson J, Langdon P, Robinson P (eds) *Designing a more inclusive world*. Springer, London, pp 225–236
 46. Rossi S, Staffa M, Tamburro A (2018) Socially assistive robot for providing recommendations: comparing a humanoid robot with a mobile application. *Int J Soc Robot* 10:1–14
 47. Sacks H, Schegloff E, Jefferson G (1974) A simplest systematics for the organization of turn-taking for conversation. *Language* 50(4, Part 1):696–735
 48. Shimada M, Kanda T, Koizumi S (2012) How can a social robot facilitate children collaboration? In: International conference on social robotics. Springer, pp 98–107
 49. Short E, Mataric MJ (2017) Robot moderation of a collaborative game: towards socially assistive robotics in group interactions. In: 2017 26th IEEE international symposium on robot and human interactive communication (RO-MAN). IEEE, pp 385–390
 50. Sidnell J, Stivers T (2013) *The handbook of conversation analysis*, vol 121. Wiley, New York
 51. Sidner CL, Lee C, Kidd CD, Lesh N, Rich C (2005) Explorations in engagement for humans and robots. *Artif Intell* 166(1–2):140–164
 52. Stivers T, Enfield NJ, Brown P, Englert C, Hayashi M, Heinemann T, Hoymann G, Rossano F, De Ruiter JP, Yoon KE et al (2009) Universals and cultural variation in turn-taking in conversation. *Proc Natl Acad Sci* 106(26):10587–10592
 53. Strauss PM, Minker W (2010) *Proactive spoken dialogue interaction in multi-party environments*. Springer, Berlin
 54. Strohkorb Sebo S, Traeger M, Jung M, Scassellati B (2018) The ripple effects of vulnerability: the effects of a robot’s vulnerable behavior on trust in human–robot teams. In: Proceedings of the 2018 ACM/IEEE international conference on human–robot interaction. ACM, pp 178–186
 55. Takayama L, Pantofaru C (2009) Influences on proxemic behaviors in human–robot interaction. In: 2009 IEEE/RSJ international conference on intelligent robots and systems, IROS. IEEE, pp 5495–5502
 56. Ter Maat M, Truong KP, Heylen D (2010) How turn-taking strategies influence users impressions of an agent. In: International conference on intelligent virtual agents. Springer, pp 441–453
 57. Trovato G, Zecca M, Sessa S, Jamone L, Ham J, Hashimoto K, Takanishi A (2013) Cross-cultural study on human–robot greeting interaction: acceptance and discomfort by Egyptians and Japanese. *Paladyn J Behav Robot* 4(2):83–93
 58. Vázquez M, Carter EJ, McDorman B, Forlizzi J, Steinfeld A, Hudson SE (2017) Towards robot autonomy in group conversations: understanding the effects of body orientation and gaze. In: Proceedings of the 2017 ACM/IEEE international conference on human–robot interaction. ACM, pp 42–52
 59. Vlachos E, Schärfe H (2012) Android emotions revealed. In: International conference on social robotics. Springer, pp 56–65
 60. Vlachos E, Schärfe H (2015) Towards designing android faces after actual humans. In: Jezic G, Chen-Burger YHJ, Kusek M, Sperka R, Howlett R, Jain LC (eds) *Agent and multi-agent systems: technologies and applications*. Springer, Berlin, pp 109–119
 61. Walters ML, Dautenhahn K, Te Boekhorst R, Koay KL, Kaouri C, Woods S, Nehaniv C, Lee D, Werry I (2005) The influence of subjects’ personality traits on personal spatial zones in a human–robot interaction experiment. In: ROMAN 2005. IEEE international workshop on robot and human interactive communication, 2005. IEEE, pp 347–352
 62. Walters ML, Dautenhahn K, Te Boekhorst R, Koay KL, Syrdal DS, Nehaniv CL (2009) An empirical framework for human–robot proxemics. In: *Procs of new frontiers in human–robot interaction*
 63. Yohanan S, MacLean KE (2012) The role of affective touch in human–robot interaction: human intent and expectations in touching the haptic creature. *Int J Soc Robot* 4(2):163–180
 64. Złotowski JA, Sumioka H, Nishio S, Glas DF, Bartneck C, Ishiguro H (2015) Persistence of the uncanny valley: the influence of repeated interactions and a robot’s attitude on its perception. *Front Psychol* 6:883

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

Mateusz Żarkowski Ph.D. Graduated from the Wrocław University of Science and Technology holding degrees in Robotics and Biomedical Engineering. Leader of Social Robot Control for Multiparty Human-Robot Interaction research project. His scientific research focuses on making machines think and behave like humans, mainly artificial intelligence, machine learning and human-robot interaction, with recent shifts towards decision intelligence.