Well done, Robot! – The importance of praise and presence in human-robot collaboration

Juliane Reichenbach, Christoph Bartneck and Julie Carpenter

Abstract—**This study reports on an experiment in which participants had to collaborate with either another human or a robot (*partner*). The robot would either be present in the room or only be represented on the participants' computer screen (*presence*). Furthermore, the participants' partner would either make 20% errors or 40% errors (*error rate*). We automatically measured the praising and punishing behavior of the participants towards their partners and also asked the participant to estimate their own behavior. The participants unconsciously praised Aibo more than the human partner, but punished it just as much. Robots that adapt to the users' behavior should therefore pay extra attention to the users' praises, compared to their punishments.**

I. INTRODUCTION

THE United Nations (UN), in a recent robotics survey, identified personal service robots as having the highest expected growth rate [1].These robots are envisaged to help the elderly [2], support humans in the house [3, 4], improve communication between distant partners [5] and provide research vehicles for the study of human-robot communication [6, 7].

In the last few years, several robots have been introduced commercially and have received widespread media attention. Popular robots (see Figure 1) include Aibo [8], Nuvo [9] and Robosapien [10]. Robosapien has sold approximately 1.5 million units times by January 2005 [11].

Figure 1: Popular robots – Robosapien, Nuvo and Aibo

Aibo was discontinued in January 2006, which might indicate that the function of entertainment alone may be an insufficient task for a robot. In the future, robots that cooperate with humans in working on relevant tasks might become increasingly important. As human-robot interaction increases, human factors are clearly critical concerns in the design of robot interfaces to support collaborative work; human response to robot teamwork and support are the subject of this paper.

Human-computer interaction (HCI) literature recognizes the growing importance of social interaction between humans and computers (interfaces, autonomous agents or robots), and the idea that people treat computers as social actors [12], preferring to interact with agents that are expressive [13, 14]. As robots are deployed across domestic, military and commercial fields, there is an acute need for further consideration of human factors.

In actions and situations where people interact with robots as co-workers, it is necessary to define human-robot collaboration as opposed to human-robot interaction: collaboration is working with others, while interaction involves action on someone or something else [15]. The focus of our research is the exploration of human relationships with robots in collaborative situations.

Limitations in terms of access, cost and time often prohibits extensive experimentation with robots. Therefore, simulating the interaction with robots through screen characters is often used. Such simulations might provide insight, focus future efforts and could enhance the quality of actual testing by increasing potential scenarios. It has been shown, for example, that screen characters can express emotions just as good as robots [16]. Using static pictures focuses responses on exterior design issues, whereas a real robot's overall physical presence may enhance or detract their anthropomorphic appearance artificially if movement is purposely restrained. On the other hand, robots may have more social presence than screen-based characters, which might justify the additional expense and effort in creating and maintaining their physical embodiment in specific situations, such as collaborative activities.

In this study we report on an experiment in which human subjects collaboratively interact with other humans, a robot and a representational screen characters of a robot on a specific task. The resulting reaction of the subjects was measured, including the number of punishments and praises given to the robot, and the intensity of punishments and praises.

Juliane Reichenbach is with the Technische Universität Berlin, Marchstr. 12, 10587 Berlin, Germany (phone: +49 (0)30 314 2559; email: juliane.reichenbach@ilr.tu-berlin.de).

Christoph Bartneck is with the Department of Industrial Design, TU/e, Den Dolech 2, 5600 MB Eindhoven, Netherlands (christoph@bartneck.de)

Julie Carpenter is with the University of Washington, Department of Technical Communication, 14 Loew Hall, Seattle WA, Box 352195, 98195 USA (julie4@u.washington.edu)

II. RESEARCH QUESTIONS

In human-human teams, people tend to punish team members that do not actively participate, that benefit from the team's performance without own contribution, or even compromise the team's performance with their failures. Fehr and Gaechter [17] showed that subjects who contributed below average were punished frequently and harsh (using money units), even if the punishment was costly for the punisher. The overall result was the less subjects contributed to team performance, the more they were punished.

If computers and robots are treated as social actors, we would expect that they were punished for benefiting from a team's performance without or with only little own contribution. It has already been demonstrated that subjects get angry and punish not only humans, but also computers when they feel the computer has treated them unfairly in a bargaining game [18].

In order to not lead the participants in one direction, we also offered the possibility of praise in the experiment reported here. The research questions that follow from this line of thought are related to the use of praise and punishment:

1) Are robots punished for benefiting from a team's performance without own contribution?

2) Are robots praised for good performance?

3) Are robots punished and praised equally a) often and b) intense as humans?

4) Does the extent of taking advantage without own contribution (low vs. high error rate) have an effect on the punishment behavior?

We are also interested in how the participants perceive their own praise and punishment behavior afterwards, and how they evaluate their own and the partner's performance.

5) Do subjects misjudge their praise and punishment behavior when asked after the game?

6) Do subjects judge their praise and punishment behavior differently for humans and robots?

7) Is the partner's and the participant's own performance estimated correctly?

III. METHOD

A. Design

The experiment was a 2 (partner) x 2 (error rate) x 2 (presence) design. The within subject factors were partner (human or AIBO) and error rate (high: 40% or low: 20%) and the between subject factor was presence (physical robot present or absent).

B. Measurements

The experiment software automatically recorded the following measurements:
• Frequency of praises and punishments: Number of incidences in which the participant gave plus points or minus points.
• Intensity of praises and punishments: Average number of plus points or minus points given by the participant, ranging from 1 to 5.
• Subject and partner errors: Number of errors made by the participant and the partner.

During the experiment, questionnaires were conducted, recording the following measurements:
• Self- evaluation of praise and punishment behavior: Self-reported frequency and intensity of the praises and punishments given by the participant.
• Self- evaluation of own and partner's performance: Self-reported number of errors made by the participant and the partner.
• Satisfaction: Participant's satisfaction with his/her own and the partner's performance after task completion, rated on a 5 point rating scale.

A post-test interview measured the following:
• Believability task: Did participants believe that the robot was able to do the task.
• Believability robot: Did participants believe that he/she interacted with a real robot.

C. Participants

25 Master's and Ph.D. students, mainly in Industrial Design and Computer Science, participated in the experiment, 19 of them were male, and 6 were female. The mean age of the participants was 24.9 years, ranging from 19 to 33. Twelve subjects participated in the robot present condition, 13 took part in the robot absent condition. Participants received a monetary reward for their participation.

D. Procedure

The participants were told that they would participate in a tournament. They would form a team with either another human player or the Aibo robot. The performance of both team players would equally influence the team score. To win the competition both players had to perform well. This setup ensured that the performance of the partner mattered to the participants.

The participants were then introduced to the task and examples were shown. The participants were told that these tasks might be easy for humans but that it would be much more difficult for the robot. To guarantee equal chances for all players and teams, the task had to be on a level that the robot could perform. Afterwards, a demonstration of Aibo's visual tracking skill was given, using a pink ball. Next, the participants would be seated.

After the instruction, the participants completed a brief demographic survey, and conducted an exercise trial. Subjects had the opportunity to ask questions before the tournament started.

The experiment was set up as a tournament, in which humans and robots played together in couples. The tournament consisted of two rounds. The participant would either be teamed up in the first round with a human and then

in the second round with a robot or the other way around. The participants were randomly assigned to either of the two possible sequences. One round consisted of two trials in which the participants' partners would either make 20% or 40% errors. The orders of the trials were counterbalanced. Each trial consisted of 20 tasks. The participants' task was to name or count objects that were shown on the computer display (see Figure 4).

After the participants entered their answer on the computer the result was shown. It was indicated if the participants and their partner's answer were correct. If the partner's answer was wrong, the participant could give minus points. If the participant decided to do so, he/she had to decide how many minus points to give. If the partner's answer was correct, the participant could choose if and how many plus points he/she wanted to give to the partner. Subjects were told that for the team score, correct answers of both the participant and the partner were counted. A separate score for each individual was kept for the number of plus and minus points. At the end, there would be a winning team, and a winning individual.

After each trial, the participant had to estimate how many errors the partner had made, how often the participant had punished the partner with minus points and how often the participant had praised the partner with plus points. In addition, the participants had to judge how many plus and minus points they had given to the partner.

After each round the participants were asked for their satisfaction with the partners' performance and their own performance. After the first round the participants' partner would change for the second round.

After the tournament, the participants were asked about the sympathy towards the robot. In an interview, the participants were asked if they believed that they played with a real robot and if they thought the task was solvable for robots. Finally, the participants were debriefed. The experiment took approximately 20 minutes.

E. Materials

For the experiment, we used the robot ERS-7 Aibo (Sony) and a picture of the robot (see Figure 2). In the robot present condition, Aibo was sitting on a table in front of a computer screen so participants could see it when they entered the room. The participants were seated back to back with Aibo (see Figure 3). In the robot absent condition Aibo would not be present, instead a picture of Aibo was presented on the computer screen in front of the participant.

Figure 2: Aibo picture used in the experiment

Figure 3: Setup of the experiment

For the tasks, 120 pictures with one or several objects on them were used. Examples are shown in Figure 4. The objects had to be either named or counted.

Figure 4: Example objects, naming (L) and counting (R).

IV. RESULTS

We calculated a 2 (partner) x 2 (error rate) x 2 (presence) ANOVA with partner and error rate as within subjects factors and presence as between subjects factor.

To get comparable numbers across the error conditions, the actual number of praises or punishments given by a participant was divided by the possible number of praises or punishments in the condition. This gives a number between 0 and 1. 0 means that no praises or punishments were given and 1 means that praises or punishments were given every time. Both human and robotic partner received praise and punishment, i.e. subjects used the chance to give extra plus or minus points.

Aibo received more praise than the human partner ($F(1,23)=7.056$, $p=0.014$). Figure 5 shows the frequency of praise in the partner and error rate conditions. Differences in intensity of praise were not significant. There was no effect of error rate or presence. There were no differences in frequency or intensity of punishment.
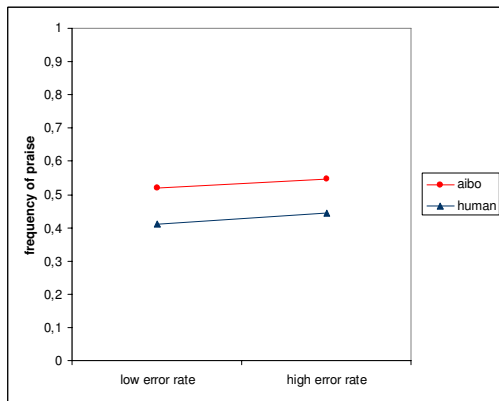

Figure 5: Frequency of praise.

After each trial subjects were asked for an estimation of their praise and punishment behavior. For the estimation of frequency of praise, we found an effect of partner ($F(1, 23)=4.957$, $p=0.036$) and an effect of error ($F(1,23)=5.941$, $p=0.023$). Participants underestimated their praises in low error rate conditions, and overestimated it in high error conditions. Estimations for Aibo were lower than for the human partner (see Figure 6). There was no effect of presence on the frequency of praise. There were no differences in the estimations of intensity of praise, frequency of punishment and intensity of punishment.
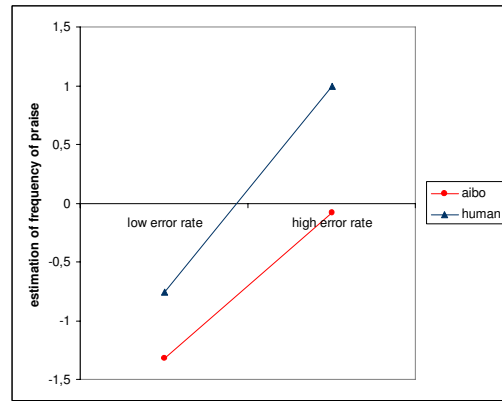

Figure 6: Estimated frequency of praise. 0 means correct estimation, a negative value is an underestimation and a positive value is an overestimation.

After each trial subjects gave an estimation how many errors they themselves and the partner made. For the estimation of partner performance, we found an effect of error ($F(1, 23)=23.633$, $p<0.001$). For low error rates, partner performance was slightly overestimated, for high error rates, the partner's performance was underestimated (see Figure 7). There was no effect of presence. There were no differences in the estimation of their own performance.
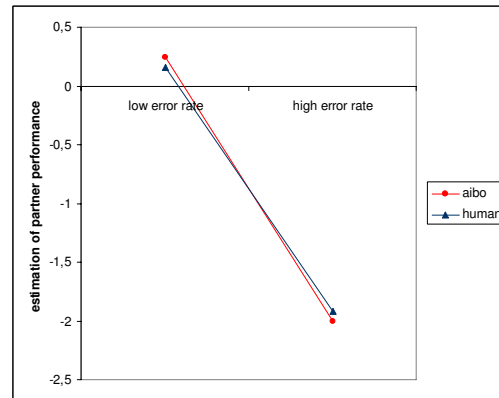

Figure 7: Estimated partner performance. 0 means correct estimation, a negative value is an underestimation and a positive value is an overestimation.

After playing the game with one partner, participants were asked how satisfied they were with their own and the partner's performance. Subjects were more satisfied with Aibo's performance than with the human partner's performance ($F(1, 23)=4.946$, $p=0.036$) (see Figure 8). There was no effect of presence. There were no differences in the satisfaction ratings for their own performance.
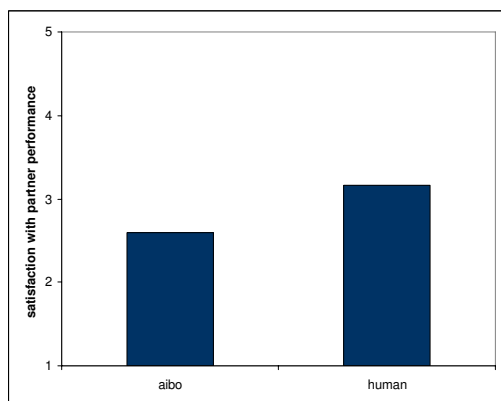
Figure 8: Satisfaction with partner's performance. 1 is very satisfied, 5 is dissatisfied.

After the game was over, subjects were asked who their favorite partner was and which partner they thought contributed more to their team's performance. 12 participants named Aibo as their favorite partner and 13 participants chose the human partner. 13 participants considered Aibo to be the highest scorer while 12 participants considered the human partner to be the highest scorer. There were no differences between present and absent condition.

At the end of the experiment, subjects were asked if they believed that the robot could do the task, and if they believed that they played with a real robot. Only 5 subjects believed that they played with a real robot. 16 subjects believed that the robot was able to do the task. There were no differences between present and absent condition.

V. CONCLUSION

People's satisfaction with their own performance and with others' performance largely depends on their expectations. Apparently, the participants in this study did not expect too much of the Aibo robot and were positively surprised by its performance. After the experiment, 16 participants believed that the tasks were feasible for the robot, even though it would be technical impossible at this point in time. They praised Aibo more than the human player and were more satisfied with its performance. The participants were not aware of their bias, since they constantly underestimated their number of praises for Aibo compared to the number of praises for a human partner. Also, they did punish Aibo just as much as a human partner. The surpassing of the expectation appears to influence only praises, but not punishments.

More and more robots use feedback from the user to adapt their behavior. Our results show that it might be beneficial to focus the analysis of the users' behavior on praises rather than on punishments, since the relationship between the users' expectations and the robots performance appears to influences praises more than punishments.

The robot's presence did not appear to have much influence on our measurements. This insignificance of presence might strengthen the position that using screen representations of the robots could simulate experiments with robots. However, in our study the participants only interacted with their partner through a computer. Only five participants actually believed that they really played with the robot. Additional research would be necessary to investigate what happens if the participants would interact with the robot directly. The effect of the robot's social presence might be higher under such conditions.

REFERENCES

[1] UnitedNations, World Robotics 2005. 2005, Geneva: United Nations Publication.
[2] Hirsch, T., et al. The ELDeR Project: Social and Emotional Factors in the Design of Eldercare Technologies. in Conference on Universal Usability. 2000. Arlington: ACM Press.
[3] NEC. PaPeRo. 2001 [cited; Available from: http://www.incx.nec.co.jp/robot/PaPeRo/english/p_index.html.
[4] Breemen, A.J.N.v., X. Yan, and B. Meerbeek. iCat: an animated user-interface robot with personality. in Fourth International Conference on Autonomous Agents & Multi Agent Systems. 2005. Utrecht.
[5] Gemperle, F., et al. The Hug: A new form for communication. in Designing the User Experience (DUX2003). 2003. New York: ACM Press.
[6] Breazeal, C., Designing Sociable Robots. 2003, Cambridge: MIT Press.
[7] Okada, M. Muu: Artificial Creatures as an Embodied Interface. in ACM Siggraph 2001. 2001. New Orleans.
[8] Sony. Aibo. 1999 [cited; Available from: http://www.aibo.com.
[9] ZMP. Nuvo. http://nuvo.jp/nuvo_home_e.html 2005 [cited.
[10] WowWee. Robosapien. http://www.wowwee.com/robosapien/robo1/robomain.html 2005 [cited.
[11] Intini, J. Robo-sapiens rising: Sony, Honda and others are spending millions to put a robot in your house. . http://www.macleans.ca/topstories/science/article.jsp?content=20050718_109126_109126, 2005
[12] Nass, C. and B. Reeves, The Media equation. 1996, Cambridge: SLI Publications, Cambridge University Press.
[13] Koda, T., Agents with Faces: A Study on the Effect of Personification of Software Agents. 1996, MIT Media Lab: Cambridge.
[14] Bartneck, C. Interacting with and embodied emotional Character. in DPPI2003. 2003. Pittsburgh: ACM Press.
[15] Breazeal, C., et al. Working Collaboratively with Humanoid Robots. in EEE-RAS/RSJ International Conference on Humanoid Robots. 2004. Los Angeles.
[16] Bartneck, C., J. Reichenbach, and A.v. Breemen. In your face, robot! The influence of a character's embodiment on how users perceive its emotional expressions. in Design and Emotion 2004. 2004. Ankara.
[17] Fehr, E. and S. Gaechter, Altruistic punishment in humans. Nature, 2002. 415: p. 137-140.
[18] Ferdig, R.E. and P. Mishra, motional Responses to Computers: Experiences in Unfairness, Anger, and Spite. Journal of Educational Multimedia and Hypermedia, 2004. 13: p. 143-161.