



Laban, G., Kappas, A., Morrison, V. and Cross, E. S. (2022) User Experience of Human-Robot Long-Term Interactions. In: 10th International Conference on Human-Agent Interaction (HAI '22), Christchurch, New Zealand, 05-08 Dec 2022, pp. 287-289. ISBN 9781450393232.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

© The Authors 2022. This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in the Proceedings of the 10th International Conference on Human-Agent Interaction (HAI '22), Christchurch, New Zealand, 05-08 Dec 2022, pp. 287-289. ISBN 9781450393232. <https://doi.org/10.1145/3527188.3563927>.

<https://eprints.gla.ac.uk/288226/>

Deposited on: 21 December 2022

User experience of human-robot long-term interactions

Guy Laban

Guy.Laban@Glasgow.ac.uk
Institute of Neuroscience and Psychology,
University of Glasgow
Glasgow, United Kingdom

Val Morrison

v.morrison@bangor.ac.uk
School of Human and Behavioural Sciences,
Bangor University
Bangor, United Kingdom

Arvid Kappas

a.kappas@jacobs-university.de
Department of Psychology and Methods,
Jacobs University
Bremen, Germany

Emily S. Cross

emily.cross@mq.edu.au
Institute of Neuroscience and Psychology,
University of Glasgow
Glasgow, United Kingdom
Department of Cognitive Science,
Macquarie University
Sydney, New South Wales, Australia

ABSTRACT

Since interactions with social robots are novel and exciting for many people, one particular concern in this specific area of human-robot interaction (HRI) is the extent to which human users will experience the interactions positively over time, when the robot's novelty is particularly salient. In the current paper, we investigated users' experience in long-term HRIs; how users perceive the ongoing interactions and the robot's ability to sustain it over time. Therefore, here we examine the effect of the repeated measures (10 testing sessions) and the discussion theme (Covid-19 related vs general) on the way participants experienced the interaction quality with a social robot and perceived the robot's communication competency over time. We found that despite individual differences between the participants, over time participants found the interactions with Pepper to be of higher quality and that Pepper's communication skills got better. Nevertheless, our results also stressed that the discussion theme has no meaningful nor significant effect on the way people perceive Pepper and the interaction.

KEYWORDS

Human-Robot Interaction, HRI, Social Robot, Self-Disclosure, User Experience, Interaction Quality, Communication Competency, Communication, Long-term Interactions

ACM Reference Format:

Guy Laban, Arvid Kappas, Val Morrison, and Emily S. Cross. 2022. User experience of human-robot long-term interactions. In *Proceedings of the 10th International Conference on Human-Agent Interaction (HAI '22)*, December 5–8, 2022, Christchurch, New Zealand. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/3527188.3563927>

1 INTRODUCTION

Social robots can elicit socially meaningful behaviours and emotions from humans across several experimental and real-world contexts [5, 6, 9]. Nevertheless, one of the challenges to human-robot interaction (HRI) research is replicating and extending lab-based findings to better understand how short, constrained laboratory manipulations might translate to real-world scenarios. Since interactions with social robots are novel and exciting for many people, one particular concern in this specific area of HRI is the extent to which human users will experience the interactions positively over time, when its novelty is particularly salient [12]. In the current paper, we were particularly interested in users' experience in long-term HRIs; how users perceive the interaction and the robot's ability to sustain it over time. Therefore, here we examine the effect of the repeated measures (10 testing sessions) and the discussion theme (Covid-19 related vs general) on the way participants experienced the interaction with a social robot and perceived the social robot Pepper (SoftBank Robotics). To evaluate the way participants perceived and experienced the interaction and the robot we are using measurements of interaction quality and communication competency.

2 METHODS

The study methodology followed an experimental design protocol for mediated online experimental design with a social robot [10]. For a detailed description of the experimental design, stimuli, task, procedure and measurements, please see the experimental design

protocol [10]. All study procedures were approved by the research ethics committee of the University of Glasgow. A 2 (Discussion Theme: Covid-19 related or general) by 10 (chat sessions across time) between-groups repeated measures experimental design was followed with 39 participants. Participants were randomly assigned to one of the two discussion topic groups, according to which they conversed with the robot Pepper (SoftBank Robotics) via Zoom video chats about general everyday topics (e.g., social relationships, work-life balance, health and well-being; see [10]). One group's conversation topics were framed within the context of the Covid-19 pandemic (e.g., social relationships during the pandemic, sustaining mental health during the pandemic, etc.), whereas the other group's conversation topics were similar, except that no explicit mention of the Covid-19 pandemic was ever made. Each interaction consisted of the social robot Pepper asking the participant 3 questions (x3 repetitions). The topic of each interaction was assigned randomly before the experimental procedure started, as was the order of the questions. Participants were scheduled to interact with the robot twice a week for five weeks during prearranged times. Each interaction with the robot lasted between 5 to 10 minutes, and another 10-20 minutes for completing questionnaires. After each interaction participants answered several short questionnaires.

2.1 Measurements

2.1.1 Interaction quality. This scale was aimed at capturing how participants perceived and evaluated the interaction with Pepper using an adapted and adjusted version by [3] for a scale by [2]. Each interaction included two random items out of six, except for the mid-session (session 5) and the last session (session 10) which included all six items of the scale. These items were evaluated on a seven-point scale ranging from 1 (not at all) to 7 (extremely). Accordingly, a mean scale was constructed ($M = 5.45$, $SD = 1.58$) which was found to be reliable (Cronbach's $\alpha = .96$).

2.1.2 Communication competence. This scale was aimed at capturing how participants experienced and evaluated Pepper's communication competency using an adapted and adjusted version by [3] for a scale by [4]. The scale included three items that were evaluated on a seven-point scale ranging from 1 (not at all) to 7 (extremely). Accordingly, a mean scale was constructed ($M = 5.76$, $SD = 1.18$) which was found to be reliable (Cronbach's $\alpha = .93$).

3 RESULTS

3.0.1 Interaction quality. We used and lme4 [1] for R to perform a linear mixed effects analysis of the effect of session number, discussion theme and their interaction term on participants' perception of the interaction quality. As fixed effects, we entered the session order, the discussion theme and their interaction term into the model. To control for the part of the overall variance that can be accounted by the participant themselves (rather than the previous exposure and topic) we included a random effect intercept for the participants; which will not be further analysed but increases the power of the analysis. Significance was calculated using the lmerTest package [7], which applies Satterthwaite's method to estimate degrees of freedom and generate p-values for mixed models. The model explains 66.3% (Pseudo $R^2 = .663$) of the variance in participants' perceptions of the interaction quality, whereas the fixed effects in

the model explain 4% (Pseudo $R^2 = .039$) of the variance in participants' perceptions of the interaction quality. The results stress that despite the variance between the participants ($SD = 1.26$), the session number has a significant positive fixed effect on participants' perceptions of the interaction quality ($\beta = .10$, $SE = .02$, $p < .001$). Nevertheless, there were no significant fixed effects in terms of the discussion theme ($\beta = -.16$, $SE = .45$, $p = .730$), and the interaction term of the session number and discussion theme ($\beta = .03$, $SE = .03$, $p = .394$).

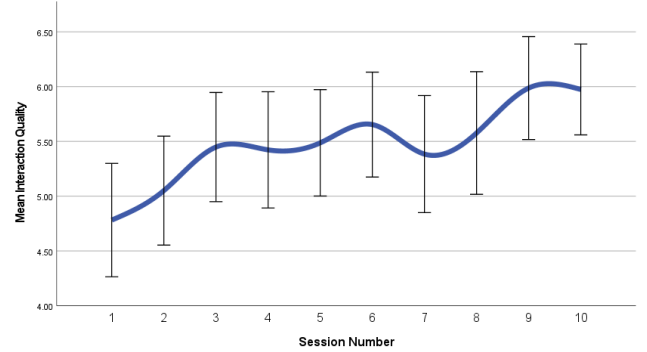


Figure 1: Mean scores of participants' perceptions of the interaction quality by the fixed effect of session number. Error bars: 95%CI.

3.0.2 Communication competence. We used and lme4 [1] for R to perform a linear mixed effects analysis of the effect of session number, discussion theme and their interaction term on participants' perception of Pepper's communication competence. As fixed effects, we entered the session order, the discussion theme and their interaction term into the model. To control for the part of the overall variance that can be accounted by the participant themselves (rather than the previous exposure and topic) we included a random effect intercept for the participants; which will not be further analysed but increases the power of the analysis. Significance was calculated using the lmerTest package [7], which applies Satterthwaite's method to estimate degrees of freedom and generate p-values for mixed models. The model explains 70% (Pseudo $R^2 = .697$) of the variance in participants' perceptions of Pepper's communication competency, whereas the fixed effects in the model explain 1.2% (Pseudo $R^2 = .012$) of the variance in participants' perceptions of Pepper's communication competence. The results stress that despite the variance between the participants ($SD = 1.00$), the session number has a significant positive fixed effect on participants' perceptions of Pepper's communication competence ($\beta = .03$, $SE = .02$, $p = .046$). Nevertheless, there were no significant fixed effects in terms of the discussion theme ($\beta = -.16$, $SE = .35$, $p = .641$), and the interaction term of the session number and discussion theme ($\beta = .02$, $SE = .02$, $p = .448$).

4 CONCLUSIONS

Here we evaluated users' experience of long-term human-robot interaction via mediated zoom online chats. We assessed participants'

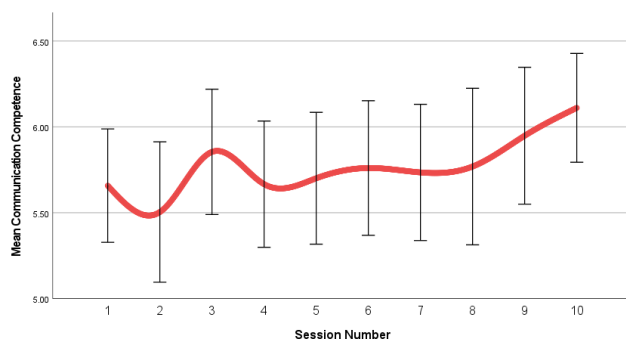


Figure 2: Mean scores of participants’ perceptions of Pepper’s communication competence by the fixed effect of session number. Error bars: 95%CI.

perceptions of the robot’s communication competency and interaction quality. We found that despite individual differences between the participants, over time participants found the interactions with Pepper to be of higher quality and that Pepper’s communication skills were better. These results are in line with some of our previous studies’ results (see [11]). Nevertheless, our results also stress that the discussion theme has no meaningful nor significant effect on the way people perceive Pepper and the interaction. These preliminary results provide important evidence concerning the potential of introducing social robots in real-life applications for long-term use and establishing meaningful relationships with these agents over time (see [8]).

ACKNOWLEDGMENTS

The authors gratefully acknowledge funding from the European Union’s Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie to ENTWINE, the European Training Network on Informal Care (Grant agreement no. 814072), the European Research Council (ERC) under the European Union’s Horizon 2020 Research and Innovation Programme (Grant agreement no. 677270 to EC), and the Leverhulme Trust (PLP-2018-152 to EC).

REFERENCES

- [1] Douglas Bates, Martin Mächler, Benjamin M. Bolker, and Steven C. Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67, 1 (10 2015), 1–48. <https://doi.org/10.18637/JSS.V067.I01>
- [2] Diane S Berry and Jane Sherman Hansen. 2000. Personality, Nonverbal Behavior, and Interaction Quality in Female Dyads. *Personality and Social Psychology Bulletin* 26, 3 (3 2000), 278–292. <https://doi.org/10.1177/0146167200265002>
- [3] Emmelyn A J Croes and Marjolijn L Antheunis. 2020. Can we be friends with Mitsuku? A longitudinal study on the process of relationship formation between humans and a social chatbot. *Journal of Social and Personal Relationships* 38, 1 (9 2020), 279–300. <https://doi.org/10.1177/0265407520959463>
- [4] Virginie Demeure, Radosław Niewiadomski, and Catherine Pelachaud. 2011. How Is Believability of a Virtual Agent Related to Warmth, Competence, Personification, and Embodiment? *Presence: Teleoperators and Virtual Environments* 20, 5 (10 2011), 431–448. <https://doi.org/10.1162/PRES.1a.00065>
- [5] Anna Henschel, Guy Laban, and Emily S Cross. 2021. What Makes a Robot Social? A Review of Social Robots from Science Fiction to a Home or Hospital Near You. *Current Robotics Reports* 2 (2021), 9–19. <https://doi.org/10.1007/s43154-020-00035-0>
- [6] Ruud Hortensius, Felix Hekele, and Emily S Cross. 2018. The Perception of Emotion in Artificial Agents. *IEEE Transactions on Cognitive and Developmental Systems* 10, 4 (2018), 852–864. <https://doi.org/10.1109/TCDS.2018.2826921>
- [7] Alexandra Kuznetsova, Per B. Brockhoff, and Rune H.B. Christensen. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82, 13 (12 2017), 1–26. <https://doi.org/10.18637/JSS.V082.I13>
- [8] Guy Laban, Ziv Ben-Zion, and Emily S. Cross. 2022. Social Robots for Supporting Post-traumatic Stress Disorder Diagnosis and Treatment. *Frontiers in psychiatry* 12 (2 2022). <https://doi.org/10.3389/FPSYT.2021.752874>
- [9] Guy Laban, Jean-Noël George, Val Morrison, and Emily S. Cross. 2021. Tell me more! Assessing interactions with social robots from speech. *Paladyn, Journal of Behavioral Robotics* 12, 1 (2021), 136–159. <https://doi.org/10.1515/pjbr-2021-0011>
- [10] Guy Laban, Arvid Kappas, Val Morrison, and Emily S. Cross. 2021. Protocol for a Mediated Long-Term Experiment with a Social Robot. *PsyArXiv* (2021). <https://doi.org/10.31234/OSF.IO/4Z3AW>
- [11] Guy Laban, Val Morrison, Arvid Kappas, and Emily S Cross. 2022. Informal Caregivers Disclose Increasingly More to a Social Robot Over Time. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22)*. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3491101.3519666>
- [12] Catharina Vesterager Smedegaard. 2019. Reframing the Role of Novelty within Social HRI: From Noise to Information. In *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI '19)*. IEEE Press, 411–420. <https://doi.org/10.1109/HRI.2019.8673219>