Accepted Manuscript

To appear at The 30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), 2021.

Teleoperated Robot Coaching for Mindfulness Training: A Longitudinal Study

Indu P. Bodala*, Nikhil Churamani*, and Hatice Gunes*

***Abstract*— Social robots are becoming incorporated in daily human lives, assisting in the promotion of the physical and mental wellbeing of individuals. To investigate the design and use of social robots for delivering mindfulness training, we develop a teleoperation framework that enables an experienced Human Coach (HC) to conduct mindfulness training sessions virtually, by replicating their upper-body and head movements onto the Pepper robot, in real-time. Pepper's vision is mapped onto a Head-Mounted Display (HMD) worn by the HC and a bidirectional audio pipeline is set up, enabling the HC to communicate with the participants through the robot. To evaluate the participants' perceptions of the teleoperated Robot Coach (RC), we study the interactions between a group of participants and the RC over 5 weeks and compare these with another group of participants interacting directly with the HC. Growth modelling analysis of this longitudinal data shows that the HC ratings are consistently greater than 4 (on a scale of $1-5$) for all aspects while an increase is witnessed in the RC ratings over the weeks, for the *Robot Motion* and *Conversation* dimensions. Mindfulness training delivered by both types of coaching evokes positive responses from the participants across all the sessions, with the HC rated significantly higher than the RC on *Animacy*, *Likeability* and *Perceived Intelligence*. Participants' personality traits such as *Conscientiousness* and *Neuroticism* are found to influence their perception of the RC. These findings enable an understanding of the differences between the perceptions of HC and RC delivering mindfulness training, and provide insights towards the development of robot coaches for improving the psychological wellbeing of individuals.**

I. Introduction

The need for mental wellbeing interventions for the general population is ever increasing [1], [2]. Mindfulness practice, in particular, is suggested as an efficient tool for alleviating anxiety and depression in individuals. For instance, research has shown mindfulness practice to improve resilience in students towards stress, fostering emotion regulation strategies and lowering depressive moods [3], [4]. However, learning and practising mindfulness can be challenging and inaccessible due to the lack of trained coaches and training programs, misconceptions about the methods involved and difficulties with establishing a regular guided practice. Virtual conversational agents, in the form of mobile applications and chat-bots, have helped with improving the accessibility of mindfulness training [5]. However, a majority of these agents are not interactive and even when they are, they rely on text-based and non-adaptive communication with very little human-feedback, resulting in individuals losing interest and dropping out from using such mental health interventions [6]. Social robots with multi-modal interaction capabilities such as speech, gestures and vision, may offer a promising solution by enabling adaptive interactions with users. Motivated by this, in this paper, we investigate the use of the Pepper Robot[1] as a mindfulness coach.

Despite their enhanced interaction capabilities, social robots face criticism with regards to the perceptions and expectations they evoke in the users and how they are managed [7]. Comparing human-human and human-robot interactions, under similar contexts, can help understand how user perceptions and expectations of these interactions change over time. Furthermore, previous studies have shown that participants' personality traits impact their interactions as well as the general impressions of social robots [8], [9]. Hence, in this paper, we conduct a longitudinal study to investigate these aspects, by comparing the interactions participants had with a teleoperated Robot Coach (RC) delivering mindfulness training, with the interactions another group of participants had with an experienced Human Coach (HC) directly. We aim to gain an understanding of the evolving user perceptions during longitudinal interactions with the coaches (HC and RC), that can in turn help us design effective strategies to develop autonomous robot coaches.

II. Related Work

A. Design of Social Robots for Psychological Wellbeing

Socially Assistive Robots (SAR) have been used for promoting general psychological wellbeing and behavioural changes in healthy populations [10]. For instance, Kidd et al. [11] developed robot coaches to instil behavioural change in people while dieting over a period of $4-6$ weeks. Jeong et al. [12] also developed a robotic platform to deliver positive psychology interventions to college students and found a significant improvement in their psychological wellbeing, mood and readiness to change behaviour across several sessions. Alimardani et al. [13] investigated changes in EEG frequency bands and self-reported affect scores during one-off robot sessions comparing meditative *vs.* non-meditative interactions. Despite achieving positive results in assistive interventions, SARs face a major challenge in terms of their acceptance by various stakeholders. A major concern regarding SARs comes from the psychologists' lack of confidence in using such robots for their practice [14]. It is important to investigate the differences in user perceptions when robots are used to deliver the interventions in place of expert human coaches or therapists. Moreover, design

*Department of Computer Science and Technology, University of Cambridge, UK. {*ipb29, nikhil.churamani, hatice.gunes*}*@cl.cam.ac.uk*. This work is supported by the EPSRC under grant ref. EP/R030782/1.

[1] https://www.softbankrobotics.com/emea/en/pepper

of robots based on human interaction theories have the potential of meeting initial user expectations and influence the evoked behaviours in humans [15], [16]. Hence, in this paper, using teleoperation setup as an initial step, we address the questions of how user perceptions towards a robot coach change compared to the perceptions towards a human coach.

B. Longitudinal Studies in HRI

Longitudinal studies are useful for understanding changes in user behaviour and experiences over time. The motivation behind investigating long-term effects of Human-Robot Interaction (HRI) is that current robots and virtual agents lack social capabilities to engage users over extended periods of time. Early long-term studies show that the *novelty effect* quickly wears off and people lose interest and change their attitudes towards robots resulting in a decreased adherence to interventions they deliver [17], [18].

Selecting appropriate methodology to analyse longitudinal data, characterized by intra- and inter-individual variability across time, is also crucial. Most longitudinal studies in HRI analyse the effects of interaction across time using Generalised Linear Model (GLM) methods such as ANOVA [12], [19]. GLM methods primarily cater to understanding mean changes for groups of observations to check whether one group differs from another and are based on an assumption that residual errors are uncorrelated. This is unlikely to be the case when modelling changes over time. Hence, in contrast to such approaches, in this study, we use growth modelling for analysing longitudinal data. Growth modelling approaches can handle violations in error assumptions for univariate analysis as well as incorporate multiple change predictors that are either static or dynamic, for multivariate analysis [20]. Using these approaches, we investigate the evolving perceptions of the participants undertaking mindfulness training delivered by the HC and RC with time.

III. Research Questions

Social robots face challenges in terms of their acceptance by the stakeholders while delivering similar interventions as humans [21]. Hence, in our experiments, we aim to study the longitudinal differences in user perceptions and the overall effectiveness of mindfulness training, as experienced by participants, when delivered by the HC *vs.* the RC. For this purpose, we firstly look into the longitudinal differences between the perceptions of the HC and RC **(RQ1)** using the participants' ratings on a 'session experience' survey filled after each session. We also investigate how these ratings for the HC **(RQ2)** and RC **(RQ3)** change with time.

Personality traits of the users are seen to influence their interactions with robots as well as the expected outcomes [8], [22]. Thus, we also study the effect of participants' personality traits on longitudinal changes in their perceptions of the RC **(RQ4)**. Additionally, we study the efficacy of the mindfulness training delivered by the HC, with respect to participants' mental state, and whether a similar effect is witnessed in the RC sessions **(RQ5)**.

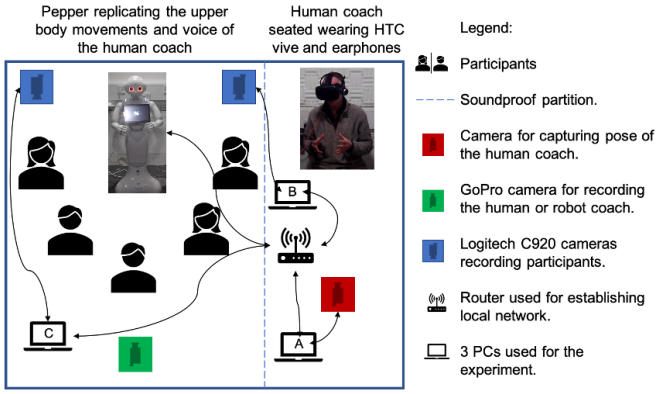


Fig. 1: Experiment setup during RC sessions.

IV. Methods

A. Experiment Design

*1) **Mindfulness Training Course**:* The user study consisted of 5 mindfulness training sessions, with one session administered each week for ≈ 40 minutes to a group of participants ($4-5$ participants per group). The sessions were delivered by an experienced human coach and the course structure was based on the 'Mindful Student Study' by Galante et al. [3]. The 5 training sessions, together, were designed to provide an introduction to mindfulness techniques and suggest how to integrate mindfulness into daily life, with each session focusing on a specific topic. These sessions were structured along several types of pedagogical strategies necessary for effective coaching of mindfulness meditation: *didactic*, to convey the conceptual basis of mindfulness; *experiential learning*, via guided meditations; and *dyadic interactions*, designed to maintain participant engagement via peer and coach dialogue customised for each session with the use of external props. Thus, the group sessions did not just contain meditations or breathing exercises but also included group interactions between the coach and the participants, including discussions on the importance of these practices, how the participants felt about them and whether they could be integrated into their daily life.

*2) **Participants**:* We recruited staff and students across the university, splitting them in two groups; HC interactions, consisting of 2 males and 7 females guided by the human coach and RC interactions, consisting of 6 males and 3 females guided by the teleoperated robot coach. Our initial plan to recruit a higher number of participants per group was interrupted by the COVID-19 pandemic. The participants were offered an incentive in the form of Amazon vouchers upon attending all the sessions. 8 out of 9 participants in the HC group, and 5 out of 9 participants in the RC group attended all 5 sessions while others missed only one session each, at random. Despite our attempts to balance gender distribution across the groups, the voluntary sign-up and the blind and random group-allocation resulted in the above distribution (see V-A for further analysis of gender effect on ratings). The experiment design was approved by the departmental Ethics Committee. All participants provided informed consent for data collection. Participants were ad-

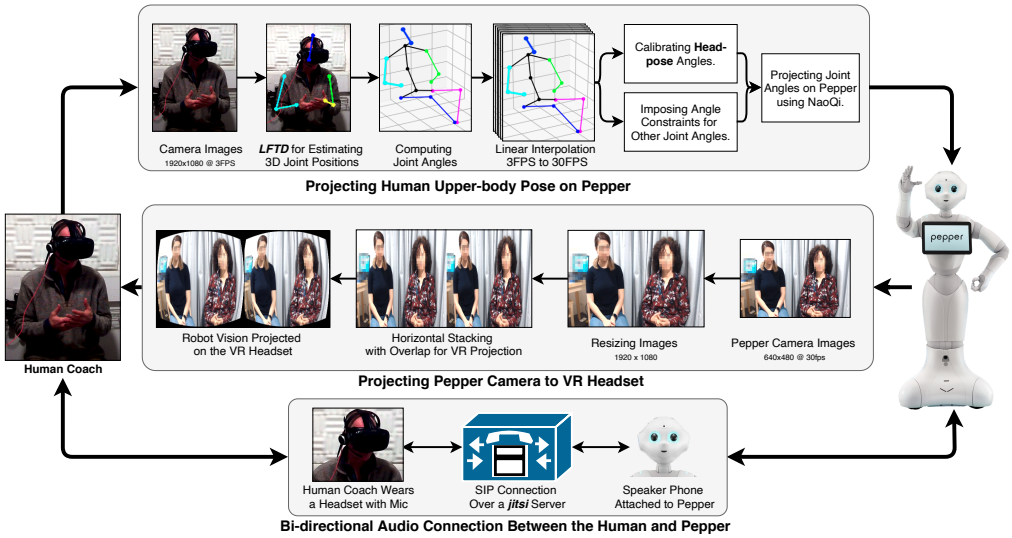


Fig. 2: Implementation of the Pose Replication, Video Projection and Bidirectional Audio pipelines during Teleoperation.

vised that they should not be undertaking other professional mental health related treatments or medication to take part in this study. We also asked the participants to fill the Participant Health Questionnaire (PHQ9) [23] and the General Anxiety Disorder (GAD7) [24] questionnaire to assess their depression and anxiety levels, respectively, to make sure no participant was experiencing high anxiety or depression levels as our target group was a non-clinical population.

*3) **Questionnaires**:* After each session, the participants filled a 'session experience questionnaire' evaluating their experience interacting with the HC or the RC. For the RC sessions, the questionnaire was adapted from a combination of the Godspeed [25] and the human-robot interaction questionnaires [26]. While the former evaluates the participants' impressions of a robot in terms of *Anthropomorphism, Animacy, Likeability* and *Perceived Intelligence*, the latter measures the interactions with items based on *Robot Motion, Conversation* and *Sensations*. A modified version, with the items irrelevant to HC (*Anthropomorphism* and *Robot Motion*) removed, was used for the HC sessions. Additionally, we also measured how the participants felt at the beginning and end of each session on the scales: *Anxious – Relaxed* and *Agitated – Calm*. Each participant also filled a 20-item 'personality questionnaire' [27]. After each session, the human coach also filled the NASA TLX questionnaire [28] to report the workload experienced during delivering the sessions directly and while teleoperating the robot.

B. Robot Teleoperation

For the RC sessions, the human coach was seated in a separate room and asked to deliver the mindfulness session to the participants by teleoperating the Pepper Robot (see Fig. 1). To enable this, a secure and encrypted local network was established that connected all the systems (3 PCs and the Pepper Robot) needed for transmitting video, audio and pose data between the robot and the human coach.

*1) **Replicating Pose**:* The upper-body and head movements of the human coach were captured using an RGB camera and processed using the *Lifting from the deep* (LFTD) algorithm [29], using a modified approach that estimates frame-by-frame $3D$ pose coordinates for joint positions corresponding to the head, shoulders, elbows and wrists of the human coach (see Fig. 2). Ondras et al. [9] found LFTD to be more robust to missing/occluded joints, providing significantly less jerky trajectories compared to other state-of-the-art pose estimation approaches. The estimated pose coordinates were interpolated to 30 frames-per-second to obtain even smoother joint movements and joint angles were estimated from the pose coordinates.

We imposed angle constraints to avoid dangerous (for robot joints) or unnatural upper-body movements of the robot resulting from spurious detection. Due to the frame interpolations made for pose estimation, the head-pose projections onto the robot resulted in a jittery camera feed from the robot, making it uncomfortable for the HC to teleoperate the robot. To correct this, head-pose pitch was thresholded based on the direction and magnitude of the HC's head movements and fixed to 5 positions[2]. All joint angles were projected onto the robot using NAOqi motion API[3].

*2) **Projecting Robot Vision**:* Along with replicating their pose, it was also important for the coach to be able to observe how the participants responded to the training session. Thus, a vision pipeline (see Fig. 2) was developed to project Pepper's vision onto an HTC Vive Head-Mounted Display (HMD) worn by the human coach, allowing them to *see through the robot's eyes*. Real-time frames were acquired from the camera on Pepper's forehead using the NAOqi vision API[4], recording at 30 FPS. Horizontal stacking with slight superposition (estimated empirically) was used to adapt the robot's vision to be projected onto the HMD.

*3) **Bi-directional Audio Connection**:* To enable the participants and the human coach to converse with each other,

[2]Head Pitch: **Centre**: [-10° $\leq \theta \leq$ 10°] $\mapsto$ 0°; **Right**: [-30° $\leq \theta \leq$-10°] $\mapsto$ -20° and [$\theta <$ -30°] $\mapsto$ -35°; **Left**: [10° $\leq \theta \leq$ 30°] $\mapsto$ 20° and [$\theta \geq$ 30°] $\mapsto$ 35°.

[3]http://doc.aldebaran.com/2-4/naoqi/motion/almotion.html

[4]http://doc.aldebaran.com/2-4/naoqi/vision/alvideodevice.html

we established a Session Initiation Protocol (SIP)-based communication server using *jitsi*[5] between two PCs over the local encrypted network (see Fig. 2). The human coach was able to interact with the participants during the RC sessions through the headphones and microphone connected to one PC while the participants conversed with the human coach using a speaker/mic system connected to Pepper.

C. Data Analysis

To investigate longitudinal changes in the perception ratings of HC **(RQ2)** and RC **(RQ3)**, we use a growth modelling for analysing the data with both inter- and intra-individual variability with missing data points. We follow the model building approach illustrated by Bliese et al. [30] and obtained the results from the random-intercept model which models the individual variability in overall perception ratings. A basic growth model for each individual participant j at time i is modelled as:

$$Y_{ij} = \pi_{0j} + \pi_{1j} Time_{ij} + r_{ij} \tag{1}$$

where π_{0j} is the intercept term (corresponding by initial overall values), π_{1j} is the slope term (corresponding to the rate of change with time) for participant j and r_{ij} is the residual error term. The intercept term, π_{0j} for a random-intercept model is modelled as:

$$\pi_{0j} = \beta_{00} + u_{0j} \tag{2}$$

Adding the residual term u_{0j} to Eq. 2 allows us to estimate the difference between an individual participant's intercept and the combined group intercept β_{00}.

We used a multi-variate analysis where participant-wise interaction effects from an Interaction Variable (C) on the random-intercept model are modelled by modifying the intercept term in Eq. 2 as follows:

$$\pi_{0j} = \beta_{00} + \beta_{01} C_j + u_{0j} \tag{3}$$

where the product of the interaction variable (C_j) of participant j and the coefficient associated with it (β_{01}) is added.

Using multivariate analysis, we studied (i) the between-group effect on the ratings by adding *group identity* variable, that is, 1 if a participant belongs to HC and 0 otherwise **(RQ1)**, (ii) how personality scores of the participants affect their longitudinal perception ratings **(RQ4)** by adding participant-wise personality scores corresponding to the five personality traits, and (iii) how participants felt before and after each mindfulness session **(RQ5)** by adding *instance*; 0 for beginning and 1 for end, as an interaction variable for *Anxious–Relaxed* and *Agitated–Calm* across all weeks. We also conducted post-hoc independent t-tests for each week to study which weeks showed significant differences between the groups with respect to these interaction variables.

V. Results

A. Perception Scores

To understand if gender had any effect on the results, we conducted an independent t-test on the ratings of all items on the 'session experience' questionnaire by separating them based on gender but did not find any significant differences in the ratings ($p > 0.05$). Nevertheless, we acknowledge that it is important to balance gender distributions across the groups for future studies.

*1) **Between-subject Analysis**:* A multivariate growth modelling analysis was conducted on the perception ratings with *time* as the within-subject factor and *group identity* as the between-group factor to investigate the longitudinal differences between the perceptions of the HC and RC (see Fig. 3b-3d,3f, 3g). A significant effect of *group identity* (that is, HC *vs.* RC) was found for *Animacy* $(t = -6.178, p = 1e^{-4})$; *Likeability* $(t = -6.329, p = 1e^{-4})$ and *Perceived Intelligence* $(t = -2.829, p = 0.012)$. Post-hoc independent t-tests were conducted to compare the HC and RC groups ratings for each week. *Animacy* and *Likeability* were found to be significantly higher $(p < 0.05)$ for the HC group as compared to the RC group for all weeks and *Perceived Intelligence* was found to be significantly higher $(p < 0.05)$ for the HC group for weeks 1 and 3.

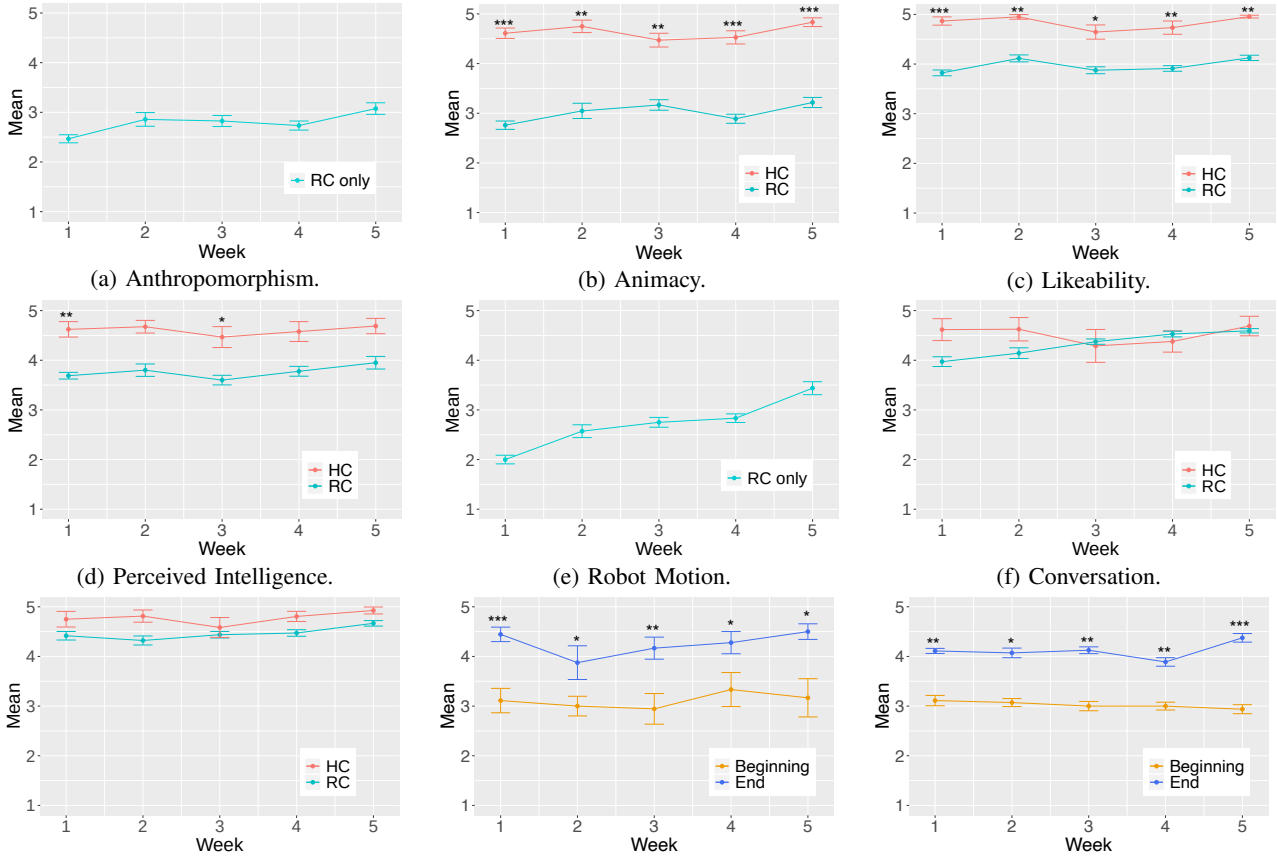
*2) **Longitudinal Interactions with the HC**:* The changes in the perception scores of HC over time are shown in Fig. 3b-3d, 3f, 3g. No significant changes with time were witnessed in the perception ratings of the HC. Moreover, ratings for all the items stayed consistently high ($mean > 4.0$) for all the sessions (on a scale of $1-5$). This suggests that the participants consistently rated high their experiences with the HC delivering mindfulness training sessions.

*3) **Longitudinal Interactions with the RC**:* The changes in the perception ratings of RC over time are shown in Fig. 3a-3g. The robot is rated low on *Anthropomorphism* and *Animacy* but high on *Likeability* and *Perceived Intelligence*. This suggests that our design for the RC evoked positive responses from the participants despite the interaction being *robot-like*. We did not see any significant changes for *Anthropomorphism*, *Animacy*, *Likeability* and *Perceived Intelligence*. This is expected as the appearance and the functionality of the robot does not change across 5 sessions. However, we found a significant effect for *Robot Motion* ($t = 4.71; p = 0.0001$) and *Conversation* ($t = 2.57; p = 0.015$), with both scores increasing over time. This suggests that participants' perception of the robot's movement and conversation consistently improved over 5 weeks.

B. Interaction between Perception and Personality Scores

To investigate the effect of participants' personality on their perception ratings, we conducted a multivariate analysis by including each of the 5 personality traits (*extroversion, agreeableness, conscientiousness, neuroticism* and *openness*), into the univariate random intercept model of each attribute of the 'session experience' questionnaire. Only *Conscientiousness* and *Neuroticism* showed significant effects on the perception ratings. We found a significant negative effect of *Conscientiousness* on the intercept variability of *Robot Motion* ($t = -2.618; p = 0.032$). This suggests that people with higher conscientiousness scores gave low overall ratings for *Robot Motion*. Additionally, we found a

[5] https://github.com/jitsi/jitsi-meet/blob/master/doc/sipgw-config.md



(a) Anthropomorphism. (b) Animacy. (c) Likeability.

(d) Perceived Intelligence. (e) Robot Motion. (f) Conversation.

(g) Sensation. (h) Beginning *vs.* End Feelings in HC. (i) Beginning *vs.* End Feelings in RC.

Fig. 3: Means and Standard Errors for Session Experience Ratings for HC and RC groups. Weeks with significant differences (post-hoc t-tests) between compared conditions are indicated by * for $p < 0.05$, ** for $p < 0.01$ and *** for $p < 0.001$.

significant negative effect of *Neuroticism* (or low emotional stability) on the intercept ($t = -2.789; p = 0.027$) and slope ($t = -2.968; p = 0.021$) variability of *Sensation*, influencing participants' overall experience of the mindfulness training.

C. Mindfulness Sessions Promote Relaxation and Calm

Using multivariate growth modelling, we found a significant effect of *instance* (that is, beginning *vs.* end) on *Anxious–Relaxed* and *Agitated–Calm* ratings for both HC ($t = 7.5; p = 0.003$; see Fig. 3h) and RC ($t = 9.6; p = 0.002$; see Fig. 3i) groups. Post-hoc paired t-tests conducted on each week's data found these self-reported scores to be significantly higher ($p < 0.05$) after the sessions for both HC and RC, suggesting that both type of sessions made the participants more relaxed and calm.

D. Workload comparison

The NASA TLX scores collected from the human coach in both HC and RC sessions indicate an increased perceived workload during the RC sessions. A *large* effect size (*Hedges'* $g = 4.48$) [31] is witnessed while comparing the workload for the human coach during the RC and HC sessions across the 5 weeks. Post-experiment discussions with the coach attributed this increase to a direct consequence of having to adapt to the use of additional equipment (such as the VR headset) and the added cognitive load of *teleoperating* the robot during the RC sessions.

VI. Discussion and Conclusion

In this paper, we report on our research that bridges two important enablers for improving psychological wellbeing: telepresence robotics and mindfulness training. This study is a part of our vision for iterative design process of autonomous robot coach for mindfulness training, where the first version, as presented in this work is implemented using teleoperation with multi-modal interaction capabilities (voice and upper-body gestures). We expect that this process will enable us to better understand the context and expectations from the participants as well as the challenges that need to be addressed prior to designing autonomous robot coaches.

We compared the participants' perceptions of the mindfulness training sessions delivered by a HC and a teleoperated RC. The participants rated the HC significantly higher than the RC for *Animacy*, *Likeability* and *Perceived Intelligence*. These findings reveal the key dimensions where the teleoperated robot differs from the human coach. The HC interactions were found to be more lively and natural with the participants finding the HC more friendly, pleasant and intelligent compared to the RC. Within the RC ratings, higher values were observed for *Likeability and Perceived Intelligence* but the RC received low ratings for *Anthropomorphism* and *Animacy*. High *Likeability* ratings imply that the robot was found to be *pleasant* and *likeable* in appearance [32] while high *Perceived Intelligence* ratings can be the result of

the robot exhibiting voice-based interactive communication and synchronous movements as it was being teleoperated by the human coach [33]. Interestingly, we did not observe any *novelty effect* or a drop in RC ratings over time. This is crucial as it addresses a critical concern regarding the use of social robots to deliver longitudinal interventions, without losing interest in the robot over time [18].

We also found a longitudinal increase in the interaction ratings for *Robot Motion* and *Conversation* indicating that these aspects required some time for the participants to get used to. Post-study open-ended discussions with the participants from the RC sessions revealed a preference for a robot that could move synchronously while speaking and maintaining eye-contact during conversations. Our future work will investigate customized generation of gestures where there is synchrony between the (robotic) speech and movements of the robot similar to Ondras et al. [9]. The participants further appreciated that they could maintain a coherent conversation with the robot where it checked on them periodically and they could talk about how they were doing. In future, we will also explore natural language generation and adaptive communication during interactions [34].

Participants' *Conscientiousness* negatively influenced the overall ratings for *Robot Motion* suggesting that highly conscientious people were more critical of the efficiency of robot movements. The link between *Conscientiousness* and *Robot Motion* was also discussed by Ondras et al. [9] where high conscientious people preferred more robust models for synthesizing robot motion. The dimension, *Neuroticism* negatively influenced overall *Sensation* ratings, that is, whether the participants liked the sessions and would like to attend more such sessions. Other wellbeing related studies have also found Neuroticism and Conscientiousness to be key factors in predicting behavioural outcomes [12]. These findings suggest that participants with different personalities have different expectations from the design of the robot coach and therefore, a need for personality-specific customization in the design and delivery of robotic interventions.

Participants also suggested additional functionalities that can be incorporated into the robot coach such as customization for topic selection and duration of mindfulness sessions as well as progress-based feedback and guidance. These are in line with the participatory design studies that we are currently undertaking to further investigate the expectations of the various stakeholders through focus group discussions.

References

[1] N. Barnes *et al.*, "An examination of mindfulness-based programs in us medical schools," *Mindfulness*, vol. 8, no. 2, pp. 489–494, 2017.

[2] K. M. Sutcliffe *et al.*, "Mindfulness in organizations: A cross-level review," *Ann. Rev. of Org. Psych. and Org. Behav.*, pp. 55–81, 2016.

[3] J. Galante *et al.*, "A mindfulness-based intervention to increase resilience to stress in university students (the mindful student study): a pragmatic randomised controlled trial," *The Lancet Public Health*, vol. 3, no. 2, pp. e72–e81, 2018.

[4] L. Wimmer *et al.*, "Improving emotion regulation and mood in teacher trainees: Effectiveness of two mindfulness trainings," *Brain and behavior*, vol. 9, no. 9, p. e01390, 2019.

[5] E. Hudlicka, "Enhancing mindfulness-based cognitive therapy with a virtual mindfulness coach," in *Career Paths in Telemental Health*. Springer, 2017, pp. 171–182.

[6] J. Torous *et al.*, "Dropout rates in clinical trials of smartphone apps for depressive symptoms: A systematic review and meta-analysis," *J. of Affective Disorders*, vol. 263, pp. 413 – 419, 2020.

[7] E. S. Cross *et al.*, "From social brains to social robots: applying neurocognitive insights to human–robot interaction," 2019.

[8] P. A. Bremner *et al.*, "Personality perception of robot avatar teleoperators in solo and dyadic tasks," *Front. in Robotics and AI*, vol. 4, p. 16, 2017.

[9] J. Ondras *et al.*, "Audio-driven robot upper-body motion synthesis," *IEEE Trans. on Cybernetics*, 2020.

[10] A. A. Scoglio *et al.*, "Use of social robots in mental health and well-being research: systematic review," *JMIR*, vol. 21, p. e13322, 2019.

[11] C. D. Kidd *et al.*, "Robots at home: Understanding long-term human-robot interaction," in *IEEE/RSJ IROS*, 2008, pp. 3230–3235.

[12] S. Jeong *et al.*, "A robotic positive psychology coach to improve college students' wellbeing," in *IEEE RO-MAN*, 2020, pp. 187–194.

[13] M. Alimardani *et al.*, "Robot-assisted mindfulness practice: Analysis of neurophysiological responses and affective state change," in *IEEE RO-MAN*, 2020, pp. 683–689.

[14] D. Conti *et al.*, "Are future psychologists willing to accept and use a humanoid robot in their practice? italian and english students' perspective," *Front. in psychology*, vol. 10, p. 2138, 2019.

[15] M. Salem *et al.*, "To err is human (-like): Effects of robot gesture on perceived anthropomorphism and likability," *Int'l J. of Soc Robotics*, vol. 5, pp. 313–323, 2013.

[16] A. Edwards *et al.*, "Initial expectations, interactions, and beyond with social robots," *Comp. in Human Behavior*, pp. 308–314, 2019.

[17] R. Gockley *et al.*, "Designing robots for long-term social interaction," in *IEEE/RSJ IROS*. IEEE, 2005, pp. 1338–1343.

[18] K. Winkle *et al.*, "Social robots for engagement in rehabilitative therapies: Design implications from a study with therapists," in *Proc. ACM/IEEE HRI*, 2018, pp. 289–297.

[19] A. Van Maris *et al.*, "Designing ethical social robots—a longitudinal field study with older adults," *Front. in Robotics & AI*, vol. 7, 2020.

[20] R. E. Ployhart *et al.*, "The "quick start guide" for conducting and publishing longitudinal research," *J. of Business and Psychology*, vol. 26, no. 4, pp. 413–422, 2011.

[21] T. S. Behrend *et al.*, "Similarity effects in online training: Effects with computerized trainer agents," *Comp. in Human Behavior*, 2011.

[22] O. Celiktutan *et al.*, "Automatic prediction of impressions in time and across varying context: Personality, attractiveness and likeability," *IEEE Trans. on Affective Computing*, vol. 8, no. 1, pp. 29–42, 2015.

[23] K. Kroenke *et al.*, "The PHQ-9: validity of a brief depression severity measure," *J. of general internal medicine*, pp. 606–613, 2001.

[24] R. L. Spitzer *et al.*, "A brief measure for assessing generalized anxiety disorder: The GAD-7," *Arch. of Int. Med.*, pp. 1092–1097, 2006.

[25] C. Bartneck *et al.*, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *Int'l J. of Social Robotics*, pp. 71–81, 2009.

[26] A. Romero-Garcés *et al.*, "Testing a fully autonomous robotic salesman in real scenarios," in *IEEE ICARSC*, 2015, pp. 124–130.

[27] E. Topolewska-Siedzik *et al.*, "The short ipip-bfm-20 questionnaire for measuring the big five," *Ann. of Psych.*, pp. 385–402, 2014.

[28] S. G. Hart *et al.*, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in *Advances in psychology*. Elsevier, 1988, vol. 52, pp. 139–183.

[29] D. Tome *et al.*, "Lifting from the deep: Convolutional 3D pose estimation from a single image," in *Proc. IEEE CVPR*, 2017.

[30] P. D. Bliese *et al.*, "Growth modeling using random coefficient models: Model building, testing, and illustrations," *Organizational Research Methods*, vol. 5, no. 4, pp. 362–387, 2002.

[31] D. Lakens, "Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and anovas," *Frontiers in Psychology*, vol. 4, p. 863, 2013.

[32] N. Chatterji *et al.*, "Effectiveness of robot communication level on likeability, understandability and comfortability," in *IEEE RO-MAN*, 2019, pp. 1–7.

[33] A. Weiss *et al.*, "Meta analysis of the usage of the godspeed questionnaire series," in *IEEE RO-MAN*. IEEE, 2015, pp. 381–388.

[34] M. E. Foster, "Natural language generation for social robotics: opportunities and challenges," *Phil. Trans. of the Royal Soc. B*, 2019.