Participant Perceptions of a Robotic Coach Conducting Positive Psychology Exercises: A Systematic Analysis

MINJA AXELSSON†, Department of Computer Science and Technology, University of Cambridge, UK
NIKHIL CHURAMANI†, Department of Computer Science and Technology, University of Cambridge, UK
ATAHAN ÇALDIR, Department of Computer Science, Özyeğin University, Turkey
HATICE GUNES, Department of Computer Science and Technology, University of Cambridge, UK

This paper provides a detailed overview of a case study of applying Continual Learning (CL) to a single-session Human-Robot Interaction (HRI) session (avg. 31 ± 10 minutes), where a robotic mental well-being coach conducted Positive Psychology (PP) exercises with (n = 20) participants. We present the results of a Thematic Analysis (TA) of data recorded from brief semi-structured interviews that were conducted with participants after the interaction sessions, as well as an analysis of statistical results demonstrating how participants’ personalities may affect how they perceive the robot and its interactions.

CCS Concepts: • Human-centered computing → User studies; Empirical studies in HCI.

Additional Key Words and Phrases: Human-Robot Interaction, Continual Learning, Well-being, Personality, Positive Psychology

ACM Reference Format:
Minja Axelsson, Nikhil Churamani, Atahan Çaldır, and Hatice Gunes. 2022. Participant Perceptions of a Robotic Coach Conducting Positive Psychology Exercises: A Systematic Analysis. 1, 1 (September 2022), 37 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION
The need for mental health interventions is increasing with more individuals turning towards different well-being practices such as mindfulness [4, 96] and Positive Psychology (PP) [33]. While mindfulness provides a meditation-based tool for alleviating anxiety and depression levels [68], PP is a branch of psychology that aims to enhance well-being by focusing particularly on the positive experiences of people and positive individual traits [33], [84]. Such positive reflection has been shown to increase feelings of positive affect and future expectancy in individuals [76].

With the recent COVID-19 pandemic, the general population has been severely impacted, resulting in negative mental health outcomes [87]. People have found it particularly difficult to seek mental health advice and treatment due to social distancing regulations imposed [52]. As a result, digital forms of healthcare have been applied to assist individuals in need [66]. These include telehealth services via video calls, online therapies, and self-help resources through mobile and web apps. As more people become accepting of robots as mental health support partners [21], Socially Assistive

†Both authors contributed equally to this research.

Authors’ addresses: Minja Axelsson, minja.axelsson@cl.cam.ac.uk, Department of Computer Science and Technology, University of Cambridge, UK; Nikhil Churamani, nikhil.churamani@cl.cam.ac.uk, Department of Computer Science and Technology, University of Cambridge, UK; Atahan Çaldır, atahan.caldir@ozu.edu.tr, Department of Computer Science, Özyeğin University, Turkey; Hatice Gunes, hatice.gunes@cl.cam.ac.uk, Department of Computer Science and Technology, University of Cambridge, UK.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2022 Association for Computing Machinery.

Manuscript submitted to ACM
Robots (SARs) are being explored to provide mental health coaching and assistance [78, 82]. Examples include robots aimed at improving the psychological well-being of individuals by providing mindfulness coaching [2, 10], instilling behavioural changes to achieve weight loss [48], and administering PPs [43]. In the absence of close access to experienced mental well-being practitioners, such SARs could support individuals and help prevent mental health deteriorating to the point where medical help is needed, by providing accessible and timely interventions.

However, for SARs to be effective, they need socio-emotional adaptability that allows them to not only generalise to real-world interactions with individuals but also provide personalised interaction experiences by adapting to individual context and behaviours [19]. They need to be sensitive to specific user verbal and non-verbal behaviours [18] while learning to respond dynamically and appropriately, adhering to the context and evolution of the interaction [24, 32].

Extending our previous work [16] on embedding Continual Learning (CL)-based personalised affect perception and interaction capabilities in the social robots (see Section 2.4 for more details), this work presents an autonomous Robotic Coach (RC) that aims to help promote mental well-being amongst individuals using PP. We employ an iterative design [69] approach, together with a professional practising psychologist, to develop conversational and gestural capabilities in the RC that extend the study design presented in [16] to multiple well-being coaching rounds and exercises. With this enhanced interaction design that involves the participants repeatedly talking to the RC and undergoing three different well-being exercises, we investigate the differences in participant perceptions of the RC by conducting a within-subjects user study with N = 20 participants. We examine the following research questions:

RQ1: What are the participants’ perceptions of the RC and the interaction?
RQ2: What are the participants’ emotional experiences during the interaction?
RQ3: What is the impact of Continual Learning on the participants’ perceptions and experiences during the interaction?
RQ4: Do the participants’ personality traits influence their perceptions and experiences over time?

In order to examine RQ1 and RQ2, we collect quantitative self-measures from participants before and after each of the three interaction rounds. Additionally, we triangulate these results with qualitative data gathered from semi-structured post-interaction interviews, the analysis of which is discussed in Section 6. To examine RQ3, we compare three interaction rounds within participants, in which the robot randomly chooses from either following a static and scripted robot behaviour (C1), or using affect-based adaptation to modulate RC behaviour (C2), or embeds continual personalisation [16] in the RC (C3) to adapt to participant behaviour. Finally, to examine RQ4, we conduct a longitudinal analysis of quantitative perception and experience measures, in order to examine whether and how participant personality contributed toward the development of the participants’ experience during the interaction.

Our results demonstrate that participants had positive experiences of the Positive Psychology exercises the robot was conducting, and were accepting of the robot as a potential coach. Participants preferred the RC with continual personalisation across most evaluations, although they did still find that the robot’s skills needed improvement. We found that the participants’ impressions of the robot, as well as their emotional experiences, were affected by their personality when examined over time. Significant effects were found with regards to participants’ level of Extraversion and Neuroticism, as well as Agreeableness and Openness.
2 RELATED WORK

2.1 Well-being Coaching and Positive Psychology

Coaching for well-being is a non-clinical approach that places an emphasis on the present, as well as how the coachee may thrive in the future and increase their fulfilment in life and at work [39]. Contrary to psychological therapy, which is intended to manage mental illness, coaching aims to boost the coachee’s mental well-being, optimism, and goal-striving [37]. Coaching can emphasise various psychological practices, such as brief-solution focused practise (helping the coachee focus on how their problem is already being addressed in small ways) and cognitive behavioural therapy (focus on the relationship between thoughts, feelings, and actions), as well as positive psychology (paying more attention to the positive aspects of one’s life). Different styles of coaching emphasize different psychological practices, such as cognitive behavioural coaching (examining the relationship between the coachee’s actions, feelings, and thoughts), or brief-solution focused coaching (encouraging the coachee to notice and plan small steps that they are taking to address their problem) [37].

Positive psychology coaching is another style of coaching, which places emphasis on the positive aspects of the coachee’s life, and fostering a positive attitude toward things that happen in their life) [83]. Positive psychology practices can include e.g. focusing on positive experiences or positive individual traits [28]. Examples of positive psychology exercises are writing down things the person is grateful for daily (in order to increase gratitude), and identifying a person’s key strengths that they used a particular situation [64]. Positive psychology has been widely used e.g. in organizations to enhance employee well-being and performance [64], and in schools to improve student well-being, relationships, and academic performance [91].

2.2 Well-being in HRI

Robots have recently been explored to address participant well-being. Positive psychology was previously explored in a study where students interacted with the robot Jibo over seven sessions at their home [43]. Participants showed improvements in their mood, well-being, and readiness to change. Another longitudinal study examined the use of a robot as a mindfulness coach, in comparison to a human coach [10]. Both coaches received positive feedback, however the human coach rated significantly higher than the robotic coach. Both these studies noted the participants’ personality traits of Neuroticism and Conscientiousness having an influence on the results of the study. Jeong et al. [43] found that people with higher Neuroticism and lower Conscientiousness had a lower well-being increase response to the robot’s interventions, in comparison to people with lower Neuroticism and higher Conscientiousness. Bodala et al. [10] found that people with higher Conscientiousness gave low overall ratings for Robot Motion, and people with high Neuroticism enjoyed the mindfulness sessions less. These studies show that participants’ personality traits can have an influence on their experiences and outcomes during a robotic well-being coaching session.

A study by Alimardani et al. [2] also examined the use of robot-led mindfulness meditation, finding that the participants’ mood improved after the interaction. Robots have also been used to improve well-being by instilling behavioural changes to help with weight loss [47], and to improve participants’ mood by self-disclosing to a robot conversational partner [1]. Duan et al. [27] compared self-disclosure to a robot and in a journal, and found that people who felt stronger negative emotions benefited more from talking to the robot in comparison to the journal. Finally, Karim et al. [46] explored a community-based social robot that could help by displaying the users’ mood data, and to raise community awareness about the emotions people feel.
2.3 Affect-based Adaptation for Personalised Interactions

Endowing social robots with such affect perception mechanisms has been explored in several Human-Robot Interaction (HRI) studies [61] aimed at improving individuals’ experiences interacting with the robots. Robots are made to estimate user affective states by analysing their facial expressions [54], body gestures [49, 71] and speech [81] during interactions. This may enable them to adapt their own behaviours during interactions to appropriately reflect the users’ affective state, engaging them in meaningful conversations [17, 32]. However, this may not be easy in all situations owing to the unpredictability of human-centred environments resulting from changing environmental conditions and interaction contexts and the inherent uncertain and subjective nature of human affective expression [30]. Most HRI evaluations make use of off-the-shelf affect perception mechanism that, despite providing state-of-the-art results on benchmark datasets, are not able to adapt to the dynamics of real-time interactions [19]. Enabling personalised affect perception requires robots to adapt their perception towards individual users, accounting for individual differences in expression and other characteristic attributes [18]. Using such personalised models to encode the users’ affective state as a contextual affordances, robots may dynamically model and adapt their interactions with the users instead of following the same static and scripted interaction for each user [15, 40]. Such personalised interactions are expected to enhance the users’ experience interacting with the robot, offering a naturalistic interaction experience.

Lifelong or Continual Learning (CL) [74, 88]-based methods have been shown to be sensitive to dynamic shifts in data distributions and can balance novel learning in robots, as they interact with their environments, with the preservation and consolidation of knowledge across of past learning. For social robots, this entails learning during affective interactions with individual users, adapting their perception models to be sensitive to individual expressions [7, 18, 35]. This allows for robots to offer personalised interactions [16] sensitive to individual behaviours, while retaining their generalisation abilities [85] for a widespread application across different individuals.

2.4 The Proof-of-Concept Interaction Framework

Our earlier proof-of-concept work [16] presented a multi-modal framework for enabling personalised HRI capabilities in affective robots, especially in the context of PP-oriented well-being coaching. Employing a CL-based personalised affect perception model [18], the proposed framework allowed for robots to be sensitive to individual affective behaviour during interactions and adapt the interaction based on such personalised perception. [16] also presents a proof-of-concept user study comparing the proposed CL-based personalised interaction capabilities in the Pepper robot to using static interaction scripts or an off-the-shelf, deep Machine Learning (ML)-based affect perception model [5, 6]. The preliminary results indicate the participants’ preference for CL-based personalised interaction capabilities in the robot, rating it high on anthropomorphism, animacy and likeability as well as on offering warm and more comfortable interactions, compared to static interactions, by being sensitive to the participants’ expressions. This work extends the user-study conducted in [16], from a single interaction round to 3 interaction rounds, in order to evaluate whether the participants still prefer the CL-based interactions. We also extend the work to examine whether the participants’ personality influence how they perceive the robot across these 3 interaction rounds as well as how the participants experience the interactions, using a qualitative analysis.

3 MATERIAL AND METHODS

In this section, we introduce how we designed and set-up the study, who our participants were, what the experiment conditions were, what our experiment protocol was, and the capabilities of the robot Pepper as the well-being coach.
3.1 Iterative Design

We employed an iterative design approach [69] when designing the robot’s behaviour. Firstly, we considered relevant works that studied what participants would prefer in a robotic well-being coach (for example, [3]), to create a sense of empathy in robot behaviour toward participants through utterances [51]. Secondly, to create a salient script for the RC, we collaborated with a professional psychologist (similar to [78]). The initial script was written by adopting PP exercises appropriate for a one-off session [72, 80], with the assumption that the robot would have only a limited conversational understanding. The script was improved with the psychologist, by removing phrases that were deemed to be invalidating or patronising, improving the explanations for the exercises, clarifying prompts for participants to share their experiences, and determining the appropriate amount of conversation. This script was implemented on the Pepper robot, which the psychologist interacted with in a 1-hour session. During this session, they took notes on further improvements which were implemented for the final version used in the study.

3.2 User Study

3.2.1 Set-up. As seen in Fig. 1, the participant sat in front of the RC on a chair with a low table separating the two. A distance of $\approx 1.0m$ was kept between the two to ensure that the RC did not violate the participants’ ‘personal space’ [62, 67]. Two cameras were placed facing the participant and the robot, respectively, to record the entire interaction. An external microphone was placed on the table in front of the participant to record their speech. Additionally, a tablet was placed on the table for the participants to fill out the questionnaires.

3.2.2 Participants. The user study was conducted as a within-subjects study with $N = 20$ participants (12 female, 5 male, 3 not disclosed) aged 26.70 ± 3.68 years from 12 different nationalities, recruited amongst the students and members of the university. The majority of the participants ($N = 13$) indicated little to no prior experience with humanoid robots. To specifically target a non-clinical population, some of the enlisted participants were screened out if they were undertaking any mental health treatment or medication. All participants were asked to fill in the General Anxiety Disorder (GAD7) [86] and Participant Health Questionnaire (PHQ9) [50] before they were enrolled in the experiments to make sure no participants were experiencing high anxiety or depression. Before the study, participants were also asked to watch two videos of the Pepper robot showing its interaction capabilities, in order to familiarise them with the robot. After the screening process, all participants provided informed consent for their participation and the usage of...
their data for post-study analyses. The participants were compensated in the form of Amazon vouchers. The consent form, study-design and the experimental protocol was approved by the Departmental Ethics Committee.

3.2.3 Experiment Conditions. During the study, each participant experienced three different conditions across three different interaction rounds. First the robot asked them to talk about events in their recent past, then about their present, and finally it asked them to imagine possible future scenarios for themselves. To investigate RQ3 (see Section 1), we implemented three different conditions with regards to the robot's level of personalisation.

(C1) Static and Scripted Interaction: In this condition, the interaction between the participant and the robot follows the pre-defined script where the robot's responses are always the same and do not take into account the participants' affective responses towards the robot. The robot always responds in an anodyne manner (control condition).

(C2) Affect-based Adaptation without Personalisation: To evaluate if embedding adaptation in the robot influences its interactions with the participants as a well-being coach, in this condition, we employed an off-the-shelf and state-of-the-art facial affect prediction model, to determine the affect expressed by the participant (see Section 3.3.3 for details). Starting from the same initial state in the dialogue manager as C1, the affect perception outcome is used to modulate robot responses to adapt to the interaction.

(C3) Affect-based Adaptation with Continual Personalisation: In this condition, instead of the pre-trained off-the-shelf model, a CL-based affect perception model is employed to determine the affective state expressed by the participant during the interactions (see Section 3.3.4 for details). As the interaction progresses, the model continually personalises towards the individual participant’s facial expressions, customising its learning. CL model’s real-time affect prediction is used to modulate the robot’s responses towards each participant. In each interaction round (past, present or future), the robot randomly ‘administers’ one of the conditions (C1, C2 or C3), randomising the order in which the participants witness the three conditions. For instance, the RC may randomly choose C3 while talking about the past, C2 while talking about the present and C1 while talking about the future. The participants were not informed of any differences between the rounds.

3.2.4 Experiment Protocol. The overall duration of the study was designed to be a maximum of 1-hour, depending on participant responses (average length was 50 ± 11 minutes, including the questionnaires after each interaction round).

For each interaction round, the participants held similar interactions with the RC, with only slight variations focusing on their past, present, and future, in that order. This ordering reflects the emphasis that well-being interventions often place on imagining optimistic futures, which has been shown to reduce pessimism, negative affect, and emotional exhaustion [56]. Expressly, for the past (first interaction round) the RC asked the participants to recall events from the past few weeks, for the present (second condition); from the same day or week, and for future (third condition); they were asked to imagine situations that may arise in the coming weeks.

(1) Introduction: The RC first greeted the participant by introducing itself, and explaining the experiment protocol. During the introduction, it asked yes/no questions making sure that the participant understood the instructions and were willing to interact with it. This allowed the participants to get comfortable with talking to the RC.

(2) Administering the Experiment Conditions: After the introduction, the RC randomly assigned the three conditions (C1, C2 and C3) talking about the past, present and future, respectively. Past (11 ± 4 minutes), present (9 ± 3 minutes), future (10 ± 3 minutes), and overall interactions (31 ± 10 minutes). Depending upon the condition administered, the RC aimed to sympathise with the participant, with empathetic utterances [51] in accordance with the predicted participant affect. Each interaction round consisted of three exercises:
i. **Two Impactful Things:** In this exercise, the RC asked the participants to talk about 2 impactful things or events that either happened in the past two weeks (past), happened or are expected to happen today (present) or are expected to happen in the coming two weeks (future). The RC also asked the participants to think about why these events happened and how they made them feel.

ii. **Two Things the Participant is Grateful For:** This exercise focused on developing gratitude, a concept emphasised during PPs, to increase positive affect, subjective happiness and life satisfaction [23, 31]. The RC asked the participant to recall or imagine (depending on past, present or future interactions) 2 things that they felt or might feel grateful for.

iii. **Two Accomplishments:** In this exercise, the RC asked the participant to think about past, present or future accomplishments, focusing on self-esteem, which has been applied to increase well-being and ameliorate depressive symptoms [34]. The RC asked the participant to describe past, present or potential accomplishments, strengths the participant applied or may apply to accomplish these [58], and how these accomplishments make the participant feel.

(3) **Feedback:** After finishing each round, the RC asked the participant for verbal and survey-based feedback on the interaction. All participants partaking in the study completed the Positive and Negative Affect Schedule (PANAS) questionnaire [92] (measuring interest, positive and negative affect in individuals) before the experiment. During the experiment, after each interaction round, that is talking about past, present or future, the participants filled in the PANAS, the Godspeed [8, 93] (measuring robot anthropomorphism, animacy, likeability, perceived intelligence and perceived safety), and the Robotic Social Attributes Scale (RoSAS) [14] (18-item scale to measure people’s perceptions of social robots based on warmth, competence, and discomfort) questionnaires along with customised questions on whether the RC understood what they said or how they felt and adapted its behaviour during the interactions, recording the participants’ impressions of their interaction with the RC. As each interaction ‘administered’ one of the three experiment conditions (C1, C2 or C3), user evaluations were used to understand how a particular condition is evaluated by the participants. After finishing the questionnaire for that interaction, the process was repeated for the rest of the interactions.

(4) **Interview and Debriefing:** After finishing all three interactions with the RC a semi-structured interview was conducted with the participants, documenting their feedback and general impressions about the interactions. This enabled us to gather qualitative data on the participant’s perceptions of the robot and debrief them on the objectives of the study and how the conditions were administered. The interview structure can be seen in Appendix A.

### 3.3 Pepper as the Well-being Coach

Similar to other related works [10, 77], we employed the Pepper Humanoid Robot[^1] to provide well-being coaching to the participants. Pepper is a 1.2m tall 20-DoF humanoid robot, equipped with multi-modal sensing capabilities to model social interactions with humans. For our experiments, we implemented different functionalities of the Pepper robot as Robot Operating System (ROS) modules as described below. A detailed description of the different components of the interaction framework developed for the Pepper robot to deliver PP-oriented wellbeing coaching can be found in [16].

[^1]: https://www.softbankrobotics.com/emea/en/pepper
[^2]: http://wiki.ros.org/melodic
3.3.1 Dialogue Management. To effectively interact with the participants, we modelled a bi-directional interaction listening to what the participants say while responding appropriately by generating naturalistic and congruous responses. For this, we implemented a Finite-state Machine (FSM)-based dialogue manager using the SMACH [11] library where each interaction consisted of 20 different dialogue states with user-responses parsed and mapped to these states. Based on the current state and the participant’s response, state transitions were designed to generate robot responses.

Speech Recognition. An external microphone was used to accurately capture the participants’ voice due to the intrinsic fan-noise causing disturbances while using Pepper’s on-board microphone. The Speech Recognition Python Library [98] was used for recognition. For the Yes/No questions, we implemented keyword-spotting to capture variations of affirmative and negative user responses. These responses were used by the dialogue manager to model state-transitions. Other user-responses to the different exercises (see Section 3.2.4) were captured for the qualitative analyses but did not effect the robot’s decision making.

Natural Language Generation (NLG). For each of the 20 states in the dialogue manager, pre-defined sentence dictionaries were used by the NLG module to generate corresponding robot responses. A total of 150 robot responses were scripted, split into several sentence dictionaries, each mapped to a given dialogue state. Pepper’s in-built Text-to-Speech (TTS) python library\(^3\) was used to generate robot responses. The on-board speakers, located on the left and right of its head, were used to communicate the generated utterances.

3.3.2 Gestures. During the interactions, the robot was made to generate certain upper-body gestures by manipulating its joints (head, shoulders, elbows, wrists and hands) to make the interactions more naturalistic [57, 65]. These gestures were performed as the robot welcomed the participant, asked a question, responded to their affective expressions (only in C2 and C3), and at the end of the experiment to say goodbye. All gestures (see Fig. 2 for examples) were pre-defined using Choregraphe\(^4\), after consulting the psychologist in order to mitigate potential negative feelings in the participants caused by the gestures.

\(^3\)http://doc.aldebaran.com/2-5/naoqi/audio/altexttospeech.html
\(^4\)http://doc.aldebaran.com/2-4/software/choregraphe/index.html

Manuscript submitted to ACM
3.3.3 Real-time Facial Affect Recognition. We used Pepper’s on-board RGB-camera (on the forehead) to record participants’ behaviour at 30 FPS with a resolution of 640 × 480. OpenCV Face Detection\(^5\) was used to detect and crop-out the participants’ face from the recorded images.

For affect-based adaptation without personalisation (C2), these face images were passed to the FaceChannel [5, 6], a light-weight off-the-shelf neural model, to encode the participants’ facial expressions in terms of valence, representing the positive or negative nature of their expression and arousal, representing the intensity, normalised to \(\in \left[{-1}, {1}\right]\). Model predictions were averaged for the last 5 seconds (5 × 30 = 150 frames) of the participants’ responses, and used to modulate state-transitions accordingly. Based on arousal-valence predictions, dialogue states were mapped to the four quadrants (Q1: positive valence & positive arousal, Q2: negative valence & positive arousal, Q3: negative valence & negative arousal, and Q4: positive valence & negative arousal) of the Circumplex Model of affect [79], to determine robot responses. An additional Neutral dialogue-state was mapped for valence \(\in \left[{-0.10}, {0.10}\right]\). The FaceChannel was implemented as a ROS module publishing frame-based predictions as ROS parameters to be processed by the state manager.

3.3.4 Continual Personalisation. We base our implementation of embedding continual personalisation (C3) in the robot on the CLIFER [18] approach by extending the imagination model presented by the authors to dimensional facial expression editing [55]. Instead of the FaceChannel, here, we randomly sample 10 face images of the participant, extracted using OpenCV Face Detection, and for each of these images, simulate 49 additional facial images representing the participants’ face corresponding to arousal and valence values ranging from \(-0.75\) to \(0.75\). These simulated images are used by the model to update its learning, personalising towards the participants’ expressions. All hyper-parameters of the model were kept the same as presented in [18]. Since the participants see the C3 condition for only one interaction round (past, present or future), only episodic memory results were used to summarise the arousal-valence predictions over the last 5 seconds of each dialogue. Similar to the FaceChannel, model predictions are published as ROS parameters.

3.4 Analyses

To evaluate the results and validate our research questions, we employ triangulation by combining multiple data sources, both quantitative and qualitative, for a comprehensive evaluation of Pepper as a well-being coach providing PPs [75]. While the quantitative data evaluates how participants rate the robot during the interactions as well as how their affect scores evolve, qualitative analysis delves deeper into their overall impressions of Pepper and provides the basis for our discussions.

Preliminary results from our study, evaluating whether continual personalisation capabilities in Pepper improve participants impressions of the robot were presented in [16]. In this work, we present an in-depth analysis of the extended user study conducted to evaluate the different factors at play that robotic well-being coaches, in the manner designed and presented here, need to focus on. As such, we present, in the following three sections, (i) a quantitative analysis of the different interaction rounds (Section 4), (ii) a longitudinal analysis of how participants’ personality dimensions may have influenced the different evaluation measures (Section 5), and (iii) a qualitative analysis of participant responses from the post-session interviews (Section 6). For all the results presented in figures, * represents \(p < 0.05\) and ** represents \(p < 0.01\).

\(^5\)https://github.com/opencv/opencv
4 QUANTITATIVE ANALYSIS OF THE DIFFERENT INTERACTION ROUNDS

This section examines the participants’ perceptions of the robot and interaction (RQ1), their emotional experiences (RQ2), and how these perceptions and experiences were impacted by the level of Continual Learning (RQ3). For the quantitative analyses, we compared the self-evaluation scores using the pairwise Wilcoxon Signed-Rank Test [94] as these evaluations are recorded under repeated-measures settings where, irrespective of the conditions assigned, the participants provide these ratings for all three interaction rounds, one after the other. We compared the scores after each round to measure any significant changes. For the robot scores, while comparing how these ratings change over the different rounds using the pairwise Wilcoxon Signed-Rank Test [94], we also evaluated how they vary within each round depending upon the condition assigned. For this we employed the Mann-Whitney U Test [60] as this comparison represents independent-measures settings. Here we present the main outcomes of these evaluations.

4.1 Self-evaluation Scores

To understand whether the interactions with Pepper as the RC had any influence on changing their affective evaluations of themselves [22], similar to [2, 89], we recorded their responses to the PANAS questionnaires before the start of the experiment as well as after each interaction round (past, present and future). As can be seen in Fig. 3, the positive affect values for the participants increases after talking about the past but not significantly. However, as the participants continue the rounds, these values significantly decrease after the present ($W = 140.5, p = 0.008$) and the future ($W = 169.0, p = 0.001$) rounds, compared to the past. The negative affect scores also see a significant decline after the present ($W = 120.0, p = 0.019$) and future ($W = 119.5, p = 0.020$) rounds, compared to before starting the
Fig. 5. GODSPEED [8] scores for past, present and future rounds across all participants.

Fig. 6. GODSPEED [8] Scores under C1, C2 and C3 for past interactions.

experiments, showing that the interactions with the robot caused a decrease in negative affect scores. In particular, we see participants significantly lose interest ($W = 78.0, p = 0.0008$), become less enthusiastic ($W = 123.0, p = 0.011$), less alert ($W = 108.0, p = 0.002$) and less attentive ($W = 83.0, p = 0.025$) comparing their scores from before the start of the user study to after the future interaction round, despite feeling significantly more proud ($W = 15.0, p = 0.015$), less nervous ($W = 64.0, p = 0.021$), and less jittery ($W = 92.0, p = 0.003$) (see Fig. 4).

4.2 Robot Perception Scores

Participants’ impressions of the robot are evaluated using the following instruments:
within each interaction round, we witness significant differences between the conditions only for the past compared to the present interaction round (see Fig. 5). In particular, the robot was rated significantly kinder \((W = 31.5, p = 0.016)\) for the past interactions than for the present round. Furthermore, it is also rated significantly more responsible \((W = 46.0, p = 0.022)\) and intelligent \((W = 66.0, p = 0.010)\) for the past compared to the future round. These scores indicate an overall decline in robot scores as the interaction progresses through the different rounds.

Exploring the underlying dimensions, we see significant differences between the conditions for several dimensions for the past interactions while present and future rounds also see a couple of significant results. For the past interactions (see Fig. 6b), C3 is rated significantly higher than C1 on human-like \((U = 4.0, p = 0.017)\), conscious \((U = 2.5, p = 0.007)\), lifelike \((U = 4.5, p = 0.022)\), liking \((U = 4.5, p = 0.024)\), kind \((U = 4.0, p = 0.015)\), pleasant \((U = 2.0, p = 0.005)\) and nice \((U = 2.0, p = 0.005)\) dimensions. Additionally, C3 is also rated significantly higher than C2 on pleasant \((U = 5.5, p = 0.004)\) and nice \((U = 8.5, p = 0.011)\) dimensions while C2 is rated significantly more conscious \((U = 5.0, p = 0.008)\) than C1. For the present interactions, however, C1 is rated as significantly more conscious \((U = 45.5, p = 0.015)\) than C2 while C2 is rated as significantly more relaxed \((U = 18.0, p = 0.024)\) than C3. No other significant differences are found for the present interaction round. Similarly, for the future interaction round, the conditions are not rated to be significantly different for most dimensions except for friendly where C1 is rated as significantly higher than C2 \((U = 40.0, p = 0.001)\) and for kind where C1 is rated significantly higher than both C2 \((U = 22.0, p = 0.014)\) and C3 \((U = 36.0, p = 0.010)\).

4.2.2 RoSAS. Comparing participant ratings at the end of the three interaction rounds, a one-tailed Wilcoxon Signed-Rank Test measured significant decline in warmth ratings (see Fig. 7). The robot is rated significantly higher for the past compared to both present \((W = 123.0, p = 0.013)\) and future \((W = 143.0, p = 0.006)\) interaction rounds. Exploring the underlying dimensions, the robot was rated to be significantly better at feeling \((W = 40.5, p = 0.010)\) for the past interactions compared to the future interactions. Additionally, it is also rated significantly more reliable \((W = 21.0, p = 0.011)\), albeit stranger \((W = 98.0, p = 0.013)\). When comparing past and present interactions, the robot is rated significantly more awkward \((W = 36.0, p = 0.004)\) for past interactions.

Similar to the GODSPEED evaluations, when investigating the impact of the assigned condition to robot evaluation within each interaction round, we witness significant differences between the conditions only for the past interactions. A one-tailed Mann-Whitney-U test found the robot under C3 to be rated significantly higher for warmth compared to C1 \((U = 2.0, p = 0.010)\), as seen in Fig. 8a. Additionally, it was also rated significantly lower on discomfort ratings \((U = 29.0, p = 0.006)\) compared to C1, indicating the interactions to be much more comfortable for the participants under C3. Furthermore, C2 was also rated significantly higher \((U = 4.5, p = 0.009)\) on warmth ratings compared to C1.
Investigating the underlying dimensions constituting warmth, competence and discomfort, significant differences between the conditions were witnessed across several dimensions for past interactions while future interactions differed only for two dimensions. No significant differences were found for the present interactions. For the past interactions (see Fig. 8b), the robot under C3 was rated to be significantly less awkward ($U = 30.0, p = 0.003$), more feeling ($U = 2.0, p = 0.008$), more compassionate ($U = 4.0, p = 0.017$), more emotional ($U = 4.0, p = 0.021$) and more organic ($U = 4.0, p = 0.017$) than the robot under C1. Additionally, it was rated as significantly less awkward ($U = 45.5, p = 0.014$) than the robot under C2. The robot under C2 was also rated to be significantly more responsive ($U = 5.5, p = 0.011$), more feeling ($U = 4.5, p = 0.008$), more compassionate ($U = 6.0, p = 0.012$), and more organic ($U = 7.5, p = 0.020$) than under C1.

No significant differences were found between the conditions for the present interactions while for the future interactions, we see a reversed effect compared to the past interactions. Here, C1 is rated as significantly less awkward ($U = 4.5, p = 0.015$) and more compassionate ($U = 35.0, p = 0.017$) than C3.

4.2.3 Customised Questions. The customised questions are used to measure specific aspects of the interactions with the robot coach such as its ability to understand what the user said, how they felt and whether it was adapting towards the participant. We see no significant differences in the participant ratings when comparing the different interaction rounds (see Fig. 9). However, when investigating how these ratings were influenced by the conditions within individual interaction rounds, we see significant differences for the past and present interaction rounds while no significant differences between the conditions are witnessed for future interactions.
Fig. 9. Customised question scores for the past, present and future interactions across all participants.

Fig. 10. Customised question scores under C1, C2 and C3 for past and present interactions.

For past interactions (see Fig. 10a), the robot under C3 was rated significantly higher ($U = 1.0, p = 0.005$) on its ability to understand how the participant felt compared to C1. Additionally, C2 was also rated significantly higher ($U = 5.5, p = 0.009$) on the same, compared to C1. For the present interactions, however, we see a reverse effect (see Fig. 10b) where C1 is rated to be significantly higher than C3 on the robot’s ability to understand what the participants said ($U = 37.0, p = 0.023$) and how they felt ($U = 38.5, p = 0.012$). C1 is also rated significantly higher than C2 ($U = 45.5, p = 0.016$) on the robot’s ability to understand how the participants felt.

5 LONGITUDINAL QUANTITATIVE ANALYSIS OF PERSONALITY

This section examines the participants’ perceptions of the robot and interaction (RQ1), their emotional experiences (RQ2), and whether and how these perceptions and experiences were influenced by the participants’ personality traits (RQ4). Before the interaction with the robot, all participants completed the mini-IPIP [26], a 20-item personality questionnaire based on the Big Five (also known as OCEAN) model of personality [36, 70]. The OCEAN model introduces five personality traits: Openness (inventive and curious), Conscientiousness (efficient and organized), Extraversion (outgoing and energetic), Agreeableness (friendly and compassionate), and Neuroticism (sensitive and nervous). An overview of average personality traits of the study population can be seen in Figure 11a, and the overall personality compositions within the study population can be seen in Figure 11b.

Our longitudinal analysis examined the relationship of the five OCEAN personality traits (Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism) to the development of participants’ PANAS scores measured before the study and as the study progressed (over four measurements). We also examined scores measured after interaction each round (over three measurements): Godspeed and RoSAS scores for the robot, as well as how the participants rated...
custom questions about the robot’s capabilities (whether it understood what they said, felt, and whether it adapted. We also examined how these personality traits affected how much time participants spent on each interaction round.

5.1 Methods

We performed an analysis with a random-intercept model following the approach illustrated by Bliese and Ployhart [9] in order to examine how measures of participants’ experience develop over time. We examine these measures in relation to each individual personality trait (Section 5.2) and participants’ inclusion in a personality group (Section 5.3). As fixed effects, we entered either the OCEAN personality trait or the participants’ personality group, as well as the interaction round (as they progressed through time). As a random effect, we had an intercept for the individual participants.

As we are interested specifically in how the participants’ experiences change over time during the interaction, we report the significant effects with regards to slope, i.e. how a personality trait or inclusion in a personality group may have affected the change of a given measure over time. With these significant results regarding slope, we also report the related statistically significant effects on intercepts (an indicator of initial overall values) related to these slopes. In the personality group analysis, we conducted post-hoc independent t-tests for each interaction round to study which weeks showed significant differences between the groups with respect to experience measures, however found no significant differences.

We attempted to examine the study conditions’ effect on participants’ measures, by utilizing a random slope model. This model would have allowed experience ratings to have different slopes to model the effect of the changing study condition on these measures. However, such a model would have fewer observations than random effects, and would thus be unreliable. As such, all longitudinal analyses presented here do not take into account the effect of individual study conditions. We take the assumption that while participants’ experiences did also differ statistically significantly between conditions, the progress of time and the knowledge gained during the interaction is a more significant contributor to developing experiences when measured over time.
5.2 Longitudinal Analysis of the Effect of Personality Traits

We examined the effect of the 5 OCEAN personality traits (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism) on the measures on participant emotions, and their perceptions of the robot over time.

5.2.1 Openness. No significant effect was found on the PANAS emotion measures by the personality trait Openness. On the Godspeed measures regarding participants’ perceptions of the robot, we found a significant negative effect of the personality trait Openness on the Godspeed Anthropomorphism measure Lifelike on the slope ($t = -2.173, p = 0.036$), and significant positive effects on the measures of perceived Safety of Relaxed on the slope ($t = 2.185, p = 0.035$), and Calm on the slope ($t = 2.066, p = 0.046$). On RoSAS measures, we found a significant negative effect of Openness on the Competence measures of Interactive on the slope ($t = -2.247, p = 0.031$), and Responsive on the slope ($t = -2.187, p = 0.035$), as well as the Discomfort measure Strange on the slope ($t = -2.292, p = 0.028$). Openness also had a negative effect on the custom variable of perceiving the robot to adapt to what the participant said and did on the slope ($t = -2.129, p = 0.040$).

These results suggest that high Openness can have a positive effect on the participants’ perceptions of the robot’s sociality (perceiving it as more Lifelike, Relaxed, and Calm, and less Strange). However, this personality trait has negative effects on perceptions of the robot’s technical competence (Interactive and Responsive), as well as the participants’ perception of the robot’s capability to adapt to what they were saying. This suggests that while people with high Openness may perceive the robot as a social other, they may be disappointed in the robot’s technical competence to sustain the social interaction.

5.2.2 Conscientiousness. A significant positive effect over time was found on the negative PANAS measure of Jittery by the personality trait Conscientiousness on the slope ($t = 2.012, p = 0.049$). A significant positive effect was found on the Godspeed Likeability measure of Kind on the slope ($t = 2.928, p = 0.006$). A significant positive effect was found on the RoSAS Competence measure of Competent on the slope ($t = 2.181, p = 0.036$) and Warmth measure Social on the slope ($t = 2.115, p = 0.041$). A significant negative effect was found on the RoSAS Discomfort measure of Aggressive on the slope ($t = -2.671, p = 0.011$).

These results suggest that high Conscientiousness can have a negative effect on the participants’ emotional experience, causing them to become more Jittery over time. However, highly Conscientious participants also perceived the robot as higher on sociable measures (perceiving it to be more Kind, Competent and Social and less Aggressive), and more Competent over time. These results suggest that highly Conscientious people may become nervous or uncomfortable during the interaction with the robot, but they perceive it as an effective and safe social other.

5.2.3 Extraversion. A significant positive effect over time was found on the negative PANAS measure of Afraid by the personality trait Extraversion on the slope ($t = 2.523, p = 0.015$). A significant negative effect was found on the Godspeed measures of the robot’s Animacy for the measure Responsive on the slope ($t = -2.306, p = 0.027$), its Likeability measure of how much they Liked the robot on the slope ($t = -2.584, p = 0.014$), and the average of Godspeed Likeability measures on the slope ($t = -2.220, t = 0.033$). Significant negative effects of Extraversion were also found for the RoSAS measures of Competence of Knowledgeable on the slope ($t = -2.429, p = 0.020$), Interactive on the slope ($t = -3.150, p = 0.003$), and the average of Competence measures on the slope ($t = -2.863, p = 0.007$). Finally, a significant positive effect was found for the RoSAS Discomfort measure of Awful on the slope ($t = 2.158, p = 0.038$), and the average of Discomfort measures on the slope ($t = 2.113, p = 0.042$).

Manuscript submitted to ACM
These results suggest that participants with high Extraversion tend to perceive the interaction with the robot more negatively over time, both in their emotional experience (a positive effect on Afraid), and how they perceive the robot (negative effects on Likeability, Knowledgeable, Interactive, Competence, Emotional, and a positive effect for Awfulness and overall Discomfort). This suggests that people who have high Extraversion, who are described as highly social, may be expecting more from the robot during a social interaction.

5.2.4 Agreeableness. Significant negative effects over time were found by the personality trait Agreeableness on the PANAS measures of negative emotional experience of Distressed on the slope \( t = -2.005, p = 0.050 \), Scared \( t = -2.147, p = 0.036 \), and Afraid on the slope \( t = -2.638, p = 0.012 \). No significant effects were found on the participants’ perceptions of the robot.

These results suggest that participants with high Agreeableness have a more positive emotional experience over time while interacting with the robot. Agreeable people have been described as empathetic and having a high capacity to work with others. This suggests that people with this trait may become more accepting of the robot over time.

5.2.5 Neuroticism. Significant positive effects over time were found by the personality trait Neuroticism on the PANAS measures of positive emotional experiences of Strong on the slope \( t = 2.891, p = 0.006 \) and Active on the slope \( t = 2.722, p = 0.009 \). Significant positive effects were also found by the personality trait Neuroticism on the Godspeed Intelligence measure of Knowledgeable on the slope \( t = 2.030, p = 0.050 \), and the average of Intelligence measures on the slope \( t = 2.317, p = 0.026 \). Positive effects were also found for the RoSAS Competence measure of Interactive on the slope \( t = 2.804, p = 0.008 \), and the average of Competence measures on the slope \( t = 2.307, p = 0.027 \). A positive effect was also found for the RoSAS Warmth measure of Compassion on the slope \( t = 2.158, p = 0.038 \). Finally, significant negative effects by the personality trait Neuroticism were found for the RoSAS measure of Discomfort of Strange on the slope \( t = -2.932, p = 0.006 \), and the average of Discomfort measures on the slope \( t = -2.900, p = 0.006 \).

These results suggest that people with high Neuroticism have a better emotional experience of the interaction over time (positive effects on Strong and Active), as well as a more positive perception of the robot (positive effect on Warmth and negative effects on Strange and the average of Discomfort measures) and its capabilities (positive effects on Knowledgeable, the average of Intelligence measures, Interactive, and the average of Competence measures). These results suggest that while people with high Neuroticism are more predisposed to negative emotionality, they are able to gain a positive experience while interacting with the robot over time. This may suggest that people with high Neuroticism may be able to benefit from the interaction with the RC.

5.3 Longitudinal analysis of personality groups

In order to further our analysis, we aimed to examine not only how personality traits affect such variables our analysis, but whether people could be grouped on the basis of their personality, and whether different groupings had a different development in experience over time. Such groupings could help us determine how robots could adapt to individuals that fall within certain personality groups. We applied Pearson’s correlation coefficient to find significant correlations between personality traits, and grouped participants into two groups according to these correlated personality traits with a K-means clustering algorithm. In our study population, we found three groupings of people with significant correlations between personality traits: high Extraversion and low Neuroticism (in comparison to low Extraversion and high Neuroticism); high Agreeableness and high Openness (in comparison to low Agreeableness and low Openness); and high Extraversion and low Conscientiousness (in comparison to low Extraversion and high Conscientiousness).
5.3.1 Extraversion and Neuroticism. People with high Extraversion have been shown to experience positive emotions with higher frequency, duration, and intensity, while people with high neuroticism have been shown to experience negative emotions with higher frequency and duration [90]. The combination of high neuroticism and low extraversion has been associated with less positive and more negative average mood as well as greater mood variation, while high extraversion and low neuroticism was associated with a better and more stable mood [95].

We examine the group $E^+N^-$ (high Extraversion and low Neuroticism) in comparison to $E^-N^+$ and find a significant negative correlation in our study population between Extraversion and Neuroticism ($r = -0.482, p = 0.037$). We group participants with a K-means clustering algorithm ($k = 2$) into groups $N_{E+N^-} = 10$ and $N_{E^-N^+} = 10$. Comparing the two groups ($E^+N^-$ and $E^-N^+$), we found significant effects between groups for the following 7 measures (Figure 12):

1. A significant positive effect of inclusion in $E^+N^-$ on the PANAS measure Nervous on the slope ($t = 2.102; p = 0.040$) suggesting that people included in $E^+N^-$, when compared to the group $E^-N^+$, over time, gave higher ratings over four measures for Nervous.

2. A significant positive effect on the subject-wise average of Negative PANAS measures on the slope ($t = 2.089; p = 0.041$) suggesting that people included in $E^+N^-$, over time, gave higher ratings for the four measures for Negative PANAS measures.

3. A significant negative effect on the GODSPEED Intelligence measure, Knowledgeable, on the slope ($t = -2.100; p = 0.043$) suggesting that people included in $E^+N^-$, over time, gave lower ratings over the three measures for Knowledgeable.

4. A significant positive effect on the RoSAS Discomfort measure, Strange, on the slope ($t = 2.789; p = 0.008$) suggesting that people included in $E^+N^-$, over time, gave lower ratings over the three measures for Strange.
(5) A significant positive effect on the RoSAS Discomfort measure, Awful, on the slope \((t = 2.343; p = 0.025)\) suggesting that people included in E+N-, over time, gave lower ratings over three measures for Awful.

(6) A significant positive effect on the subject-wise average of RoSAS measures for Discomfort on the slope \((t = 2.622; p = 0.013)\) suggesting that people included in E+N-, over time, gave higher ratings over the four measures for Discomfort.

(7) A significant positive effect on the Time used for each interaction round on the slope \((t = 2.517; p = 0.016)\), and a significant negative effect on the intercept \((t = −2.725, p = 0.014)\). This suggests that people included in E+N-, over time, used more Time over the four interaction rounds, and also started at a significantly lower usage of time.

Inclusion in the E+N- group shows a negative statistical difference in robot measures related to the participants’ emotional experience: both the PANAS measure for Nervous, as well as the subject-wise average for Negative emotions, increase over time. Participants in group E+N- also rated the robot more negatively over time: the Godspeed measure Knowledgeable decreases, and the RoSAS measures Strange and Awful increase over time. The subject-wise average of RoSAS Discomfort measures also increases over time. Thus, members of the E+N- group seem to view the robot more negatively over time, and their emotional experience becomes more negative over time. Participants who have high Extraversion and low Neuroticism may be more comfortable in social interactions, and thus expect more from their social interaction partner in return. Members of this group may have been disappointed in the robot’s social capabilities, and thus had a more negative experience.

5.3.2 Agreeableness and Openness. High Agreeableness and Openness have been shown to be predictors of tolerance to human diversity with respect to beliefs, showing that agreeableness and openness are predictors of some forms of pro-social behaviour [13]. Openness was also found to be related to higher cultural intelligence (i.e. a person’s ability to deal effectively with people from different cultural backgrounds), when the agreeableness of the person was also high (but not when it was low) [53].

We examine the group A+O+ (high Agreeableness and high Openness) in comparison to A−O−. Applying Pearson’s correlation coefficient, we find a significant positive correlation in our study population between Agreeableness and Openness \(r = 0.55, p = 0.013\). We group participants with a K-means clustering algorithm \((k = 2)\) into groups \((N_{A+O+} = 12)\) and \((N_{A−O−} = 8)\). Between the two groups, we found significant effects between groups for the following 5 measures (Figure 13):

(1) A significant positive effect of inclusion in A+O+ on the GODSPEED Likeability measure, Friendly, on the slope \((t = 2.752, p = 0.009)\) suggesting that people included in A+O+, when compared to the group A−O−, over time, gave higher ratings over the three measures Likeability.

(2) A significant positive effect on the GODSPEED Safety measure, Relaxed, on the slope \((t = 2.122, p = 0.041)\), suggesting that people included in A+O+, over time, gave higher ratings over the three measures for Relaxed.

(3) A significant negative effect on the RoSAS Competence measure, Interactive, on the slope \((t = −2.848, p = 0.007)\), suggesting that people included in A+O+, over time, gave lower ratings over the three measures the Interactive measure.

(4) A significant negative effect on the RoSAS Competence measure, Responsiveness, on the slope \((t = −2.626, p = 0.013)\), suggesting that people included in A+O+, over time, gave lower ratings over the three measures for Responsiveness.
(5) A significant negative effect on the RoSAS Discomfort measure, Strange, on the slope ($t = -2.042, p = 0.049$), suggesting that people included in A+O+, over time, gave lower ratings over the three measures for Strange.

Inclusion in the A+O+ group shows a positive statistical difference in robot measures related to interpersonality: Godspeed’s Friendly and Relaxed have higher ratings over time, and the RoSAS measure Strange decreases over time. Thus, members of the A+O+ group seem to rate the robot as more personable over time. However, members of the A+O+ group also seem to expect more from the robot in terms of technical competence: members rate the robot lower over time in the RoSAS measures of Interactive and Responsiveness. Participants who are highly agreeable and who are open to new experiences may view the robot with a more open mind, thus being more accepting of the robot’s social skills. However, the same participants, who are characterized by imagination, might have imagined more developed capabilities for the robot, thus being disappointed in its technical competence.

5.3.3 Extraversion and Conscientiousness. In the context of job performance requiring interpersonal interaction, high Extraversion has been shown to result in better performance in high individuals with high Conscientiousness (but Manuscript submitted to ACM
lower performance in people with low Conscientiousness) [97]. Additionally, people with higher Extraversion have also been shown to anthropomorphize robots more [44].

We examine the group E+C− (high Extraversion and low Conscientiousness) in comparison to E−C+.

Applying Pearson’s correlation coefficient, we find a significant negative correlation in our study population between Extraversion and Conscientiousness ($r = -0.482, p = 0.037$). We group participants with a K-means clustering algorithm (k = 2) into groups ($N_{E+C−} = 9$) and ($N_{E−C+} = 11$). Yet, no significant differences were found between the two groups, for any of participants’ reported scores.

5.3.4 PANAS measures Over Time. We also examined how the average of negative and positive PANAS measures developed over time, and how these differed by personality groups (Figure 14). Applying a post-hoc t-test, we found significant differences between the first and the final measure of the average negative emotions in the whole population ($p = 0.050$), with negative emotions having significantly decreased. The same significant decrease in the average of negative emotions was found for the personality groups A+O+ ($p = 0.036$) and E−N+ ($p = 0.018$). No significant differences were found for the changes in the average of positive PANAS emotions.

6 QUALITATIVE ANALYSIS OF PARTICIPANT RESPONSES

This section examines the participants’ perceptions of the robot and interaction (RQ1), and their emotional experiences (RQ2), from a qualitative point of view. For a qualitative understanding participants experience of the interactions, as well as their impressions of the RC and the exercises conducted, we applied Thematic Analysis (TA) [20] and analysed the data gathered from the post-study semi-structured interviews (see Appendix A for the interview questions). Participants were asked about their perceptions of the robot, its performance as a well-being coach, and its behaviour. We coded the transcribed interviews into ten major themes (see Fig. 15) presented below.
6.1 PP Exercises

Participants made several comments on the Positive Psychology (PP) exercises (see Table 1), noting that the exercises themselves were successful and helped them reflect on the positives in their life, and they experienced the robot being helpful. P9 underlined the importance of the robot asking the “correct questions”, saying that working with a coach is important, and that the questions were well made in this case. P7 noted that they felt more grateful and positive after the exercise, and that the interaction was effective, especially those about the future. Participants noted that the exercises made them think of new things, but that some of them had difficulty understanding the tasks. This could be mitigated with the robot further explaining the structure of the exercise session in the beginning. Instructions on how to interact with the robot, as well as the ability to ask the robot to repeat questions were called for in future interactions.

Participants noted that improvements are needed on the robot’s responses (see Table 1). Participants also noted that the robot could be more responsiveness in its responses — e.g. P11 suggested that the robot could ask follow-up questions about telling someone about a thing they were grateful for, i.e. who they would like to tell. This is discussed further in the next section (Section 6.2).

6.2 Robot capabilities

Robot capabilities were also discussed outside of the PP exercise content itself (see Table 2). As noted in the context of PP exercises, participants wanted more responsiveness, noting that the robot was “inert” in its lack of suggestions (P3). P5 noted that they enjoyed the robot giving specific new responses such as “that sounds like a great accomplishment” and “you should be proud”, describing a sense of being listened to. This wording variety should be emphasised in future
Table 1. Quotes from participants ($P_i, (i = 1, 2, \ldots, 20)$) regarding the Positive Psychology exercises by the Pepper robot. $+/−$ signs indicate positive and negative statements, neutral statements do not contain signs.

| PP exercises | Quotes from participants |
|--------------|--------------------------|
| Experience of helpfulness | $P_{2+}$: “[It’s] an exercise we could do ourselves, but it felt more interactive because the robot was asking you questions.”  
$P_{6+}$: “I particularly liked it trying to make me think about a positive future. [...] That made me feel hopeful of what will come.”  
$P_{7+}$: ‘At the end I think I felt more grateful. [...] I’m feeling more positive than I was before the start of the exercise, in that respect it was effective.”  
$P_{12+}$: ‘If I [use it every day], it can potentially lift my mood. It’s like a pet, but I’m not responsible for it when I travel or something like that.”  
$P_{19+}$: “Pretty useful in terms of guiding me to reflect on certain things that I feel grateful for, it did have an impact on my present emotions, and I did feel better after the exercise.” |
| Correct questions | $P_{2+}$: “It was good that it explained the thought process behind the exercises [...] why thinking of the positive aspects was important.”  
$P_{9+}$: “I think the questions were really well made, and so because of that it does help you centre a little bit, into ‘What has been positive?’ ‘What has been negative?’ It really felt a little bit like a coaching [or] counselling session. In that sense, it helped a lot.”  
$P_{3+}$: “The questions really made me think, especially for the questions about future. The present and the past are usually something which we reflect on, but for the future it’s something which is nice. It made me think about things that I usually don’t.” |
| Thinking of new things | $P_{7+}$: “It was good, trying to make me think ‘oh why would I feel grateful for things’ so that was that was good because I don’t really think about that.”  
$P_{11+}$: “It definitely made me think of stuff that I hadn’t thought of before, it took me a little while to think of things, and once I was thinking of them, I was like ‘oh yeah, let’s listen to that’ so I think it did its job.”  
$P_{4+}$: “[...] It’s good to have the opportunity to think about these things. The difficulty sometimes is [being put] on the spot to list these things.”  
$P_{5}$: “It feels weird to talk to a robot about anything personal at all. So I’m not sure I would choose that if I had an option. But surprisingly, at the end I was like ‘Oh yeah, I am grateful for things’, so I don’t know. I guess it did work.”  
$P_{6+}$: “I felt guided through things and [it] structured my thinking about different things.” |
| Repeat questions | $P_{21}$: “An ability to go back or repeat the question would fix a lot of the problems I had.”  
$P_{6}$: ‘[...] I didn’t understand one of the questions [...] so I asked if it could repeat that and I think it registered that as a next [answer]. So I guess [being able to ask it to repeat questions] would be better in making it feel like a conversation.” |
| Difficulty understanding tasks | $P_{11}$: ‘I couldn’t quite see the structure ahead of time [...] . It asked what are you grateful for, and then when it talks about the accomplishments, I was like, ‘oh, I totally should have used that’. [...] I just feel like it would have been more natural with a person to be like, ‘oh, that one thing that I was just saying before’. And to a robot I kind of just felt like I said the same thing again as I did before, so I guess that was slightly weird.”  
$P_{16}$: “These three [rounds (past, present, and future)], they all seemed similar. […] It asks very similar questions. The first one was supposed to focus on the past and the second one on the present. And it asked similar questions like ‘recent accomplishments’ and that was confusing.” |
| Instructions to use robot | $P_{2}$: “Maybe I wasn’t speaking loud enough, because I felt like I needed to repeat the commands on the feedback portion of the interaction] a lot of times.”  
$P_{3}$: “I responded and it didn’t get it, but after that it was very accurate. I didn’t have to have any trouble, like having to wait for it. Sorry, you asked me to wait, but it didn’t feel very natural.” |

applications of a robotic coach. A sense of acknowledgement and active listening (a technique used in coaching [12]) could be emphasised further by picking up keywords from the participants’ speech, and adapting the robot’s responses specifically to these. P1 noted that this would “reinforce the person’s belief that the robot was understanding what they’re saying”.

Manuscript submitted to ACM
Table 2. Quotes from participants ($P_i$, ($i = 1, 2, ..., 20$)) regarding the Robot Capabilities. $+/−$ signs indicate positive and negative statements, neutral statements do not contain signs.

| Robot capabilities | Quotes from participants |
|--------------------|--------------------------|
| Responsive-ness    | P5+: “There were a few times when it did have a specific new response or it said ‘That sounds like a great accomplishment’ or ‘You should be proud’ and that was nice, I was like ‘Oh thank you robot for listening to me’.” | |
|                    | P18-: “I would like to feel a difference between the reactions to something positive or negative I’m saying. If I said ‘I’m very happy’ I wouldn’t like to say ‘That’s fine’. [I would like it to] engage more positively with my positive feelings.” | |
|                    | P10-: “Even though it was pretty much always the same response, I think it’s nice that she responded to everything I said.” | |
|                    | P1-: “It wasn’t really useful past the fact that it did ask the questions [...]. If there’s a complex robot in front of me, in order to do more than just readouts, it would be nice if it could respond.” | |
| Acknowledgement    | P14-: “You say something and then it continues along not really acknowledging what you said.” | |
|                    | P20-: “[...] [the thing that made] me feel that it wasn’t useful was when I didn’t feel that what I was saying was being reacted to in a way I expected.” | |
|                    | P13: “It did ask good questions, but then there wasn’t any room for follow up.” | |
|                    | P1: “If you did add acknowledgement to it while [someone is] talking, [...] it would reinforce the person’s belief in the robot actually understanding what you’re saying. As opposed to when it’s just standing, then I don’t know if it actually really understands what I’m saying.” | |
| Active listening   | P13: “It asked good questions that made me think. But it felt more like the robot has a good script than the robot is engaging in active listening.” | |
|                    | P19: “Sometimes I feel like [it’s] not really listening because [there are] very few responses, [and it’s] choosing between the available sentences. So [it] can say ‘Umm, I see. Thank you for sharing that with me’, with exactly the same [tone] every single time.” | |
|                    | P10: “More natural head movements and arm movements, while I’m talking, [...] If that’s enhanced, [it’s] useful for keeping me engaged. [...] Nodding or saying ‘Umm’, ‘OK’, or ‘Wow’.” | |
|                    | P8+: “I think the head nods were quite real and if it gets very in sync, those were one of the things that made me feel like I was being listened to.” | |
| Wording variety    | P16: “Maybe include more of a variety of responses, [so it’s] not always the same.” | |
|                    | P11: “I guess a wider repertoire of different things it might do [...]. There were certain phrases it would say lots of times. I think it was moving the same way every time.” | |
|                    | P6: “[To feel that it] understood how I felt, it would have required more of a range of replies.” | |
|                    | P4: “If there is a bit more variability, it would feel more natural. [...] Because then when you do this five times [...] it loses all sense of humanity. You just disconnect because you can see the automaton. Now you can see the machine there.” | |
| Adaptation         | P13:-: “[...] [it was] asking me questions that are making me think productively on my end, but it’s not really responding to what I’m saying.” | |
| Keywords           | P9: “[...] I don’t know how much it actually was just reacting to keywords, or was really sensing what I was feeling.” | |
|                    | P3: “[...] it can give some suggestions, like ‘try focusing more’.” | |
| Keywords           | P12: “I told the story, so maybe mentioning some keywords from my stories rather than using a template [...]” | |
| Keywords           | P17: “I thought it could do a little bit more personalization, if I say the the keywords, like ‘organisation’ or ‘events’ it could say ‘Oh, I hope you have a nice event’ or [...] ‘You have a nice organisational skills’ or ‘Seems like good organisational skills’, so it can pick up words and make it seem more human interactive I think.” | |

6.3 Robot features

Participants made notes on the robot’s features (Table 3), such as its form, voice, movements and behaviour. In appearance, P9 described the robot as an overall “positive figure” and “cute”. P21 noted that it was at the high end of humanoid before being creepy, while P15 called it “non-threatening”. P4 “approachable”, and P10 “animated” as it spoke. Overall, the Pepper robot was evaluated positively on appearance for this task. Participants made a few specific complaints...
Table 3. Quotes from participants ($P_i, (i = 1, 2,\ldots,20)$) regarding the robot’s features. $+/−$ signs indicate positive and negative statements, neutral statements do not contain signs.

| Robot features | Quotes from participants |
|----------------|--------------------------|
| **Behaviour**  | P2+: “Very supportive, especially once it starts to properly respond to the content of what we’re saying. [...] very helpful, it kind of guided you through the whole process.” |
|                | P2-: “I think at some points it felt a bit condescending. Maybe it’s just my ego talking, it just felt a bit weird to be told ‘well done’ by a robot.” |
|                | P5-: “It accomplished the task, but it was robotic [...]” |
|                | P16+: “A lot of listening and giving space to talk. I think that’s important.” |
|                | P9+: “It made me feel comfortable, so I was OK in talking about these things.” |
|                | P8+: “One really nice thing was that I felt heard. [...] and I felt like we were interacting.” |
|                | P16+: “It gave me a good feeling because it seemed compassionate.” |
|                | P20+: “I felt a bit stressed out because I knew that it would respond quickly and I’m not very fast when it comes to classifying emotions.” |
|                | P2+: “I liked that it stays calm. When someone talks to another person who isn’t a qualified professional, it’s very hard to get someone to stay completely calm in the face of some things.” |
|                | P4: “I like the friendliness, it’s a good thing. I like the positive reassurance that it gives when it talked to me about the achievements and then said ‘great you deserve this’, that’s good to hear.” |
| **Voice**      | P19+: “[The robot] speaks at a relatively slow pace, [...] it’s patient and listening.” |
|                | P9+: “The tone of voice was also really good.” |
|                | P14-: “I was very creeped out. The technical speech engine isn’t human.” |
|                | P20+: “The voice is pleasant. There’s some weird pronunciation [...], but otherwise it’s good.” |
|                | P21+: “I like the voice actually. I’m cool with this slightly robot-like inflection.” |
|                | P8: “I like the friendly like tone voice. That made me want to talk more with the robot.” |
|                | P18: “I like the voice. It doesn’t seem as mechanical as I thought it would.” |
| **Appearance** | P9+: “There’s a lot of things here that together all of it may be a very very positive figure. And it doesn’t feel very threatening at the moment. [...] It’s very cute.” |
|                | P21: “It’s just about is at the high end [of being humanoid] before it gets too creepy. [...] It didn’t have to be an entire robot, [it could be] a rectangle with eyes maybe.” |
|                | P10+: “The human form factor is quite nice.” |
|                | P21+: “It does have quite a friendly face, it’s non-threatening. It’s a good size.” |
|                | P7: “I think if it was more human-like, that would actually have a negative impact.” |
|                | P15+: “That it didn’t have any expressions [...] was actually good. It was like I’m talking to somebody, but at the same time that person is [...] just taking in what I’m saying, it’s not really even giving back. That’s actually a good thing.” |
|                | P20+: “Because the expression is the same, it’s very hard to see if the robot is empathising.” |
|                | P4+: “I like that it has a more positive than neutral face. It’s more approachable like that.” |
| **Movement**   | P17: “During speaking it can make some motions. [...] There are other ways of agreeing and not agreeing in body language terms. Tapping, clicking, thinking, putting their hands together in a more lifelike in a way.” |
|                | P14-: “There’s nothing that’s clearly missing, except in terms of movement. Maybe moving around while I’m talking as opposed to only when it’s responding. And acknowledging things as I talk.” |
|                | P14+: “The way it moves is actually not too odd. The movement seems to be fairly lifelike.” |
|                | P21+: “Because it was always mostly the same gesture, that gesture made it artificial.” |
|                | P19+: “[It’s] trying to show open, welcome, and friendliness, but we don’t really get the meaning. It’s always the same movement with exactly the same angle, so it’s easy to see the programming.” |
|                | P18: “Even though I like the expression of the hands, it could be a little bit repetitive.” |

about Pepper’s appearance, noting that they were sometimes distracted by the light in its eye, and were sometimes confused by the function of Pepper’s tablet displaying a screen saver animation. However, some participants noted that
Table 4. Quotes from participants ($P_i, (i = 1, 2, ... 20)$) regarding the advantages of a robotic coach. +/- signs indicate positive and negative statements, neutral statements do not contain signs.

| Robot advantages | Quotes from participants |
|------------------|--------------------------|
| Saying out loud  | P21+: “Being forced to do it out loud is helpful.”  
P14+: “It was good to kind of think about what’s going on at the moment”  
P15+: “[...] When you think about it and you say it out loud, it puts a lot of things in perspective, so I think the robot did help me do that.”  
P4+: “[...] you can still benefit from simply having to say things, and listening to them as well.” |
| Lack of judgement | P3+: “It’s easier to talk to a robot somehow than a real human being. You don’t face judgement from a robot, so that’s nice. It’s an open ended question so sometimes you feel stupid saying certain answers to a human than a robot.”  
P19+: “It does a good job in listening and not judging you.”  
P15+: “I felt like I wasn’t getting judged. [...] the fact that it didn’t have any expressions was actually a good thing. A lot of times when you talk about things you look at another [person], whether verbally or non-verbally [they react] [...] and the fact that it didn’t do it made me feel like I can say as much as I can.”  
P8+: “When you’re trying to do this with a person one more wall that you have to navigate, which I didn’t feel like in the beginning. [...] You’re not entering a room and sitting with a person, so there’s less social layers you have to navigate.” |
| Accessibility    | P9: “For coaching it could definitely help. Not for making a first assessment, because I think that still should be a human. But needing someone that you can interact with and actually go through the things you’re [thinking about], having the robot react to this, [...] that’s really good.” |
| Sense of presence| P1+: “It’s obviously not going to be on the same level as as a human [...] so the question really is, is it better than just something you can do to yourself? [...] If it were online then [...] you can just go to any question and just skim through it. Whereas with this, you actually have to sit down and it has to actually say each one to you. So you’re forced to consider each question and in that sense I suppose it’s better than just reading something.”  
P20: “It’s like seeing an online quiz, [...] but having a text to speech sort of thing and tell me, that and having a physical presence in the room with me. That establishes more of a connection. It’s not the same as if it were a human.” |

they enjoyed the animation and it made them feel relaxed. For future interactions, a robot with no eye light or tablet could be considered, if they are not relevant for the interaction.

Participants had more feedback on its movement, calling for more gestures to signal body language and turn-taking, making it clearer when they should be speaking. Gestures were also seen as “artificial” and “repetitive”. This could be mitigated by developing a more sophisticated method for back-channeling during conversations.

In its behaviour, the robotic coach was called “very supportive” (P2), “compassionate” (P16), and “giving space to talk” (P10), all important qualities for a robotic coach. However P2 noted that it also felt “condescending” at points, even if the robot telling them “well done” was intended to be friendly. The robot’s encouragement phrases should be further evaluated on when they are useful, and when they are seen as too invasive. This could vary according to participant, which makes examining how different personality traits affect perceptions of the robot important.

Participants also commented on the robot’s voice, appreciating the “slow pace” that made it appear “patient” (P19), its tone of voice (P9), “pleasant” (P20), and “friendly” (P7). P14 however said they were very “creeped out” by the voice, and P13 called it “weird”. This illustrates that there is no one-size-fits-all solution with regards to the voice of a robotic coach — rather that the voice should be selected in order to be congruous with the expectations of a wellbeing coach (i.e. patient, pleasant, and friendly.

Manuscript submitted to ACM
Table 5. Quotes from participants (\(P_i, (i = 1, 2, ..., 20)\)) regarding the disadvantages of a robotic coach. +/- signs indicate positive and negative statements, neutral statements do not contain signs.

| Robot disadvantages | Quotes from participants |
|---------------------|--------------------------|
| Timing              | P19: “It can be a distraction, when I haven’t finished the whole speech, but I’m just pausing a little bit, and [it] thought I finished and then [it] asked me questions.”  
   P10: “The long pauses were awkward, but you also don’t want to interrupt people.”  
   P10-: “While I narrate saying ‘Oh wait, I’m thinking’, [the robot] thinks I’ve said something and goes on to the next question. [...] I didn’t know what to do to give myself time to think.”  
   P9-: “I didn’t understand the interval it was giving me to talk, if it’s a fixed interval. Sometimes it would interrupt me or I almost couldn’t finish the sentence and it was asking the next question.”  
   P20: “There were times that I felt interrupted, so there was a point when I was thinking about something. [The robot says], ‘oh I couldn’t understand what you said’, so it’s that sort of speech processing and recognition. So at times it felt that it was trying to hurry me up.”  
   P18: “Sometimes you want to reply saying ‘thank you’ as well. And I don’t know if the robot is waiting for me to say that or not.” |
| Incorrect responses | P6-: “I was saying I managed to travel and it was the first time that I’ve done that in a while, and just spent time with people I care about. And the replies were. ‘That sounds like a tough experience’, so I don’t know what triggered it to say that, but it was the complete opposite of what I was feeling and it was making me feel obviously misunderstood. And also made me question what was I saying that was making this happen.”  
   P15: “I think at times I would say something positive and it would say ‘oh, I’m sorry you felt that way’. [...] At times, it didn’t understand what I was trying to say.” |
| Privacy issues      | P4: “It’s hard to choose what you even say, especially if you know you’re being recorded, you have to choose something that is not going to expose you extremely at the same time.”  
   P19: “I feel like more like I’m being recorded, than being heard, or [listened to].” |
| Lack of humanness   | P7-: “If this was an actual person asking me the same questions, maybe it would have had a better impact like because maybe I could have had a more personal connection with them. And the robot is quite limited in the depth of information that it can ask. You can’t program it to ask follow up questions. A real therapist or […] a coach [...] could have done better [...]”  
   P19: “It doesn’t have some of the bad things a human coach would have, like negative comments. [...] On the flip side [the robot] didn’t give me an organic response. It’s always the same response and sometimes it can be demotivating.” |

6.4 Robot Advantages, Disadvantages and Comparisons with Human or App-based Coaching

Participants detailed both robot advantages (see Table 4) and disadvantages (see Table 5). Participants also made favourable and negative comparisons of the robot to both a human coach and a hypothetical phone or computer application performing the same PP exercises (Table 6). These three themes are interrelated, and are all discussed in this section.

A major advantage that participants saw in using the robot was speaking out loud. Participants felt that verbalizing their thoughts helped them think of new things and put things into perspective, and benefited simply by hearing themselves talk. This is supported by literature, where self-talk is found to be related to motivation [38], and that the production effect indicates that words spoken out loud improve explicit memory in comparison to words simply read quietly [42, 59]. However, speaking out loud can come with the disadvantage of incorrect robot responses. On occasion, the robot would respond incongruousley to an emotional experience that a participant had described, making them feel misunderstood. This suggests that there is a delicate balance in using adaptive robot utterances, so that participants can benefit from speaking and being responded to, while mitigating the negative effects of incorrect responses. The same negative effects can be seen in the disadvantage of incorrect timing: some participants noted that long pauses were
Table 6. Quotes from participants ($P_i$, ($i = 1, 2, ..., 20$)) comparing the RC to a human or another type of application. $+/−$ signs indicate positive and negative statements, neutral statements do not contain signs.

| RC VS human or application | Quotes from participants                                                                                                                                 |
|---------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------|
| RC VS human               | P10: “Not too bad, [it’s] obviously not a replacement for [a human]. If there’s no human counsellors or coaches available, then it’s possible that it can have a good effect on people.”  
|                           | P9+: “I wonder how different this would be from a real human session. It made me feel comfortable, so I was OK in talking about these things. So in that sense, it’s definitely positive.”  
|                           | P5-: “It was appropriate in telling me what to think, but it didn’t seem like it was understanding things because it just said the same thing back. Which is fine, because it’s not actually my therapist, but probably would be irritating if you were really pouring your heart out.”  |
| RC VS application         | P20+: “I’ve tried this in the past as well, maintaining a gratitude diary. […] When someone puts you on the spot to think about it’s more effective, than trying to enforce it myself.”  
|                           | P21+: “The voice, the interactive part of it, that is different to doing it on a screen, or in self help books. So I think that is positive.”  
|                           | P10+: “It’s quite useful, and a better format than doing it on a laptop or something like that.”  
|                           | P9+: “We’ve seen some of these coaching exercises through a computer when you even have just a sentence, like let me think about the positive thing. So having a voice and a sort of reaction is a really good step, so I think in that sense as a coach, I think it was cool. It was very positive.”  |

awkward, or that the robot would begin responding when the participant hadn’t finished. This could be improved by enhancing turn-taking cues, e.g. with robot backchannelling via vocalizations and gestures.

Another such delicate balance in robotic coach design is the relationship between privacy and data collection, as well as adaptation and responsiveness. Multiple participants wished the robot would adapt to their speech in its verbal utterances, creating wording variety with picking up keywords from their speech. This could improve the impression of active listening, which they saw could be a potential advantage over a mobile phone or computer application. However, participants were concerned about being recorded while sharing their emotional experiences. The concern for lack of privacy could be mitigated through user education on what data exactly is being collected, and how it is being stored.

Participants also commented on both the robot’s lack of judgement, but also its lack of humanness. Participants noted feeling comfortable with the robot, and that there were less social layers to navigate than with a human coach. A participant also noted that it didn’t have some of the “bad things” that a human coach may have (such as negative comments), however that the responses weren’t organic, which could be demotivating. On the other hand, some participants experienced a lack of judgement from the robot, noting that it was easier to talk to, and that they “didn’t feel stupid” giving the robot answers, as they might feel with a human coach. A participant noted that the robot’s lack of facial expressions was a good thing, and that they felt like they could “say as much as they can” without reactions from the robot. This suggests that robotic coaches are useful for certain use cases, but could be irritating if they were “pouring their heart out”.

6.5 Users

Participants also made suggestions as to for whom such a robotic coach could be useful (Table 7). Participants saw that the robot worked on a basic level, and could work especially for users who didn’t have a lot of experience with PP exercises, and could benefit from learning them and the positive thinking from the robot. However, high expectations might lead to disappointment. Participants suggested that introverted users may benefit from the robot, if users were to feel shy or awkward talking to a human. A participant specifically mentioned the robot being useful at their home.
Table 7. Quotes from participants ($P_i$, ($i = 1, 2, ... 20$)) regarding the potential users of the robotic coach. $+/-$ signs indicate positive and negative statements, neutral statements do not contain signs.

| Users | Quotes from participants |
|-------|--------------------------|
| User expectations | P6: “I guess it’s more beneficial for people who have never done therapy or that kind of stuff. Because if you’re already used to laying things up like this and focusing on the positive aspects rather than the negative ones, then it feels very obvious, what it’s doing.”  
P3: “I think it also depends upon the expectations of the people. So if some days I’m feeling really crappy and if I’m expecting more, maybe then I wouldn’t get what I want.”  
P3: “It works to at the very basic level. It definitely works, but you don’t have the same day every day. So it depends upon that. What kind of expectations you have coming into this.” |
| COVID-19 | P12: “[With] COVID quarantine [...] if this is industrialised for people who are isolated at homes [...]. It’s a moving intelligence, and being able to understand your difficulties, and trying to help with some mental health problems. I think that’s really, really beneficial.”  
P12: “If we apply this to the home, to a domestic situation. I would say it’s very useful. [...] even in not Covid-19 times, I think it’s still very useful. And for these people who are at care homes or or these people who are not sociable, like in Japan, a lot of teenagers are absorbed in computer games and they only talk to virtual machines. So I would say this is a very good channel for them to open themselves back to society, like a transition period because the computer does interact. So this can bridge them from self isolate situation to a more [social life].” |
| Introverts | P12: “For an introverted person it’s really helpful [...] [you] could be acting very shy to a real person, but I feel quite relaxed to talk, because in your mind it’s still a machine.”  
P9: “It will be helpful for a lot of people who might have awkward interactions with humans.”  
P9: “[...] People who feel awkward when they want to say things that are very personal and private. If the robot can react, and help people go through their problems, it’s a huge thing.”  
P17: “Some people feel more confident talking on video call [...] this might cater to a particular part of the of people who who are more confident speaking to a robot.” |
| Nurseries and hospitals | P9: “[...] that could help a lot in nurseries and hospitals. You can’t have a nurse to every person, but you could have a robot. Having these small interactions, this could make a huge difference.”  
P9: “[...] my parents worked in the health sector and I know these can provide a huge help. [...] especially thinking about older generations who are alone [...]” |

during Covid-19 social distancing conditions. Another user suggested the robot could be useful in the *health sector*, such as in nurseries and hospitals, where there is a lack of staff, and having small interactions with a robot could make a positive difference. One participant mentioned that it “depends on the expectations of people”, suggesting that deploying these robotic coaches into the field needs to be carefully framed to participants as to their usefulness and capabilities.

### 6.6 User perceptions of conditions

Even though participants remarked both positively and negatively on every condition (see Table 8), they especially highlighted instances when the robot made mistakes with regards to empathetic utterances, during affect-based adaptation ($C_2$) or affect-based adaptation with personalisation ($C_3$). Here, participants sometimes preferred $C_1$, which employed no adaptation and thus did not cause the robot to make mistakes. This occasional preference for *neutrality* is in-line with the comments made by the consulting psychologist, who remarked that neutral statements may be preferred by some individuals as the RC making mistakes and saying things that do not align with participants’ expectations may become jarring (as seen under the themes in Table 2  *Responsiveness* and *Adaptation*).

Yet, some participants remarked being particularly pleased when the robot responded accurately, as seen in Table 2 with regards to *Responsiveness*. This supports the notion that *continual personalisation* may become even more relevant for long-term interactions, in the form of *active listening*, which would help them engage more with the robot. In
Table 8. Quotes from participants \((P_i, (i = 1, 2, \ldots, 20))\) regarding the study conditions. +/- signs indicate positive and negative statements, neutral statements do not contain signs.

| Conditions | Quotes from participants |
|------------|--------------------------|
| **Condition C1** | |
| P4+ (Present): "I think the robot was performing really well on the present (C1) task." |
| P6- (Present): "In general, I think the past (C3) was fine, the future (C2) was great, and the present (C1) was really bad." |
| **Condition C2** | |
| P17+ (Past): "Except the last round (C2) [...], it felt like mechanical sort of questions that came one upon the other. [...] In last round, I felt it did pick up on some things and it understood what’s a good event." |
| P3- (Present): "If I said some negative things it still didn’t react to the negativity. That made [it] feel a bit mechanical." |
| **Condition C3** | |
| P3+ (Future): "I think the last one (C3) was smoother. It was telling me OK, I hope you achieve these things. The last one was easier. It felt more meaningful but I don’t know. Maybe I got used to it." |
| P6- (Past): "I think for the first part about the past (C3), it was just the same sequence of replies that it would give me. Thank you for sharing that with me’ was said almost all the time." |

In particular, participants would hope to see the robot using "keywords" from their speech to signify that it understood what had been said, and provide acknowledgement via gesturing and intermittent back-channelling (such as "Huh" and "Okay" noises) during the participants’ self disclosure.

### 6.7 Novelty and habituation

In addition to the themes outlined above, the phenomena of **novelty and habituation**, and **suspension of disbelief** were evident in the analysis of qualitative data (Table 9).

Participants described their experience being described by **novelty and habituation** [41], both to the exercises and the robot itself. Participants noted that their experience changed throughout the sessions. Habituation was seen as a positive by P17, who understood the exercises more clearly throughout the sessions. P8 on the other hand saw this as both a positive and a negative, noting feeling more prepared but reflecting less due to being able to predict the interactions. Participants described feelings of **surprise** toward the RC when first encountering it. However, participants habituated to the robot over time, after “getting used to it”. These observations suggest that in order to create successful longitudinal (whether longer single sessions, or multiple repeated sessions) interactions with a robot, the robot should follow an interaction pattern to be understandable, but with variations introduced in order to keep the participant engaged in the exercises.

### 6.8 Suspension of disbelief

Suspension of disbelief is a concept that originates from the examination of the willing participation in a fiction by an audience [73]. It has been previously defined in HRI [29] as the person’s willingness to suspend their disbelief of what is living in physical social robotics, and that this may effect how the person perceives the robot. Suspension of disbelief is also relevant in well-being practices, as the optimism — a foundational concept of Positive Psychology — has been described as a “willing suspension of disbelief” [63].

While participants did not explicitly describe their experiences as a suspension of disbelief, but described feelings of strangeness and being conscious of how they responded to the robot (Table 9). For example, P2 said that being aware of the robot not actually understanding what they were saying made them less engaged. P2 described talking to an inanimate object as “shouting into the void”, as they weren’t sure what it was actually comprehending.

Participants also noted that the robot’s form contributed to their lack of suspension of disbelief, citing its hands as too human. P4, who has more extensive experience with robots than the majority of participants, noted that they were
Table 9. Quotes from participants ($P_i, (i = 1, 2, ... 20)$) regarding the phenomena of Novelty and Habituation, and Suspension of Disbelief during RC interaction. $\pm$ signs indicate positive and negative statements, neutral statements do not contain signs.

| Phenomenon | Quotes from participants |
|------------|--------------------------|
| Novelty and habituation | Exercises  
P17: “I felt the last round, [...] I got more used to the pattern and that is probably the bias here, but it might have also been that it understood things more clearly in the last round.”  
P8: “I think it was very useful, but after the first run, I could predict what exactly was going to happen, and what questions I would be asked, so I felt more prepared. But consequentially, that meant I wasn’t thinking or reflecting enough for that.”  
Robot  
P11: “[... the surprise], it was weird to be talking to a robot. It’s the surprise when it first talked and moved, and the first time it would say certain things, that were kind of funny as well. I think once I got used to it, and it didn’t take very long to get used to it, the weirdness wore off early. [...] Then I was like, ‘Oh OK, cool that’s what it’s trained to do’, but it caught me off guard.”  
P20: “Initially I was a bit creeped out by the smile, but it’s actually fine. It seemed completely fine, maybe a bit surprising at first when there’s some weird arm movement. But otherwise easy to get used to. And even though it was weird at the beginning, I got very used to the appearance after just a few minutes.”  
P11: “At the beginning I kept laughing and [thinking], this is not the purpose of [the interaction, if] I’m just sitting here laughing [...]. After I got over laughing I started paying closer attention.” |
| Suspension of disbelief | P2: “The fact that this is a study might have skewed my response towards the robot [...]. The experience, especially it being in such a controlled environment [is] a bit strange [...] I felt a lot more observed. I think that definitely did play into how open I was with responses.”  
P2: “I think initially it felt a bit strange to be talking to a robot, but I think it really did get a lot better through the exercises. It felt more natural somehow, once you get over the initial strangeness of the whole thing. [...] The entire concept, [...] talking to an inanimate object about your feelings, and your thought process. [...] So it feels like ‘am I just saying this, am I just shouting things out into the void’? Or is it actively responding to me?”  
P9: “It was quite strange because you don’t really know how much the robot is reacting to what you’re actually saying, or to keywords. [...] Even though you have someone or something there to listen to you, it’s actually not listening to you. [...] You start questioning, ‘what type of interaction am I having’? But it does help to talk about it, even if it is out loud.”  
P4: “I know the domain so I was conscious. I was trying to reverse engineer how Pepper works.”  
P9+: “It’s quite well made in terms of creating the environments that you might need for a coaching session.” |

7 DISCUSSION AND CONCLUSION

Our analyses extensively examined how participants experienced the interaction with the RC as well as the PP exercises (RQ1), along with their emotional experiences (RQ2), from both a quantitative (Sections 4 and 5) and qualitative (Section 6) perspective. Additionally, we examined how impressions of the RC corresponded to the level of emotional adaptation in the robot (RQ3, Section 4), and how these impressions developed longitudinally (over time) according to personality traits and personality groups (RQ4, Section 5).

With regards to their impressions of the RC (RQ1) and their emotional experience (RQ2), participants described feelings of helpfulness, and noted that the robot asked the correct questions during the exercises. Participants especially...
appreciated *thinking of new things* during the exercises, and having the opportunity to focus on the positives throughout the exercises. Additionally, the participants’ negative affect (PANAS) scores decreased significantly from before the overall interaction, to after. These results indicate that the participants, overall, found the exercises to be useful in themselves.

Participants enjoyed the *future* exercises the most, noting that the future is something that they do not usually reflect on. However, participants spent the most time on the *past* exercises, describing that it was the easiest to think of things to say for these exercises. From *past to future* exercises, a decrease, albeit insignificant, was seen in participants’ ratings of the robot, as well as their positive affect scores. These, along with qualitative data, may indicate a *novelty and habituation* [41] effect, although the same results could be explained by the participants feeling more able to find things to say during the *past* exercises. While we did show images and video of the robot to participants before the study to habituate them with the robot’s appearance and functional abilities, future research should aim to further reduce these effects via warm-up sessions before interactions.

The participants’ preference for the RC with *Continual Personalisation* (RQ3), further emphasises the relevance of future research on *responsiveness and adaptation*. The preference for adaptation, and the need for future improvements, was supported by the qualitative results. While participants appreciated the robot when it adapted to them, participants would have wanted the robot to respond even more to what they were saying, and to acknowledge their emotions. Along with facial affect, future studies should also consider undertaking long-term studies integrating more advanced Natural Language Processing (NLP) capabilities, strengthening the active listening capabilities in the RC.

With regards to the influence of participants’ personality traits on their perceptions and experiences (RQ4), a significant decrease in negative affect was witnessed for all participants, especially for those belonging to personality groups *A+O+* and *E-N*+ (in comparison to the groups *A-O−* and *E+N−*). Additionally, people with higher *Openness, Conscientiousness, Agreeableness* and *Neuroticism* all perceived the robot more socially positively over time. This is especially promising in the case of *Neuroticism*, in which people are predisposed to higher negative emotionality, a trait which can pose challenges for the usefulness of practices such as Positive Psychology. These results suggest that the robot could be suitable for people high in all these four personality traits.

In contrast, people with higher *Extraversion* perceived the robot more *negatively* over time, and also had a more negative emotional experience. This could be due to extraverted people being highly social, and thus expecting more from a social interaction. The same effect is seen in the analysis of the personality group *E+N−*, where this group viewed the robot more negatively over time in terms of technical capabilities and levels of discomfort (in comparison to the group *E−N*+). These results are supported by participants describing that the robot could be especially useful for *introverted users*. Introverted people should be considered as a relevant user group for Robotic Coaches and interactions with such groups needs further research. For this user group, the robot could provide the advantages of being *accessible*, helping the participants *speak out loud*, and provide a *lack of judgement* and *sense of presence*. The impact of personality traits should continue to be explored, especially keeping in mind the potential advantages the robot could have for introverted users.

7.1 Limitations

Our study has a limited number of participants (*N* = 20). Additionally, participants interact with the robot repeatedly over multiple exercises in a single session. While this one-session study with 20 participants is an necessary examination of how robotic coaches could operate, future research is needed with more participants, where the participants would
interact with the robot over a longer term in multiple sessions. Such a study would enable the examination of whether the results discussed here still hold under a longer time frame.

Additionally, future studies should aim to adapt the robot based on other social signals from the user, in addition to facial affect. For example, body language [25] and tone of voice [45] are other important indicators of emotional experience. Furthermore, using the participants’ personality traits to further personalise the interactions towards each participant can help improve their individual experience interacting with the robot. Additionally, the robot should also adapt its behaviour, along with the interaction, by generating appropriate gestural responses as well as modulating its verbal responses towards the participants. This may be achieved by using back-channeling, hand and head movements as well as other physical gestures in order to improve its social behaviours.

Acknowledgments

We would like to thank Katherine Parkin for her valuable help with Positive Psychology exercises. We are thankful to all the study participants.

Funding: N. Churamani is funded by the EPSRC under grant EP/R513180/1 (ref. 2107412). M. Axelsson is funded by the Osk. Huttunen foundation and the EPSRC under grant EP/T517847/1. H. Gunes is supported by the EPSRC under grant ref. EP/R030782/1. A. Çaldır contributed to this work while undertaking a summer research study at the Department of Computer Science and Technology, University of Cambridge.

Data Access Statement: Overall statistical analysis of research data underpinning this publication is contained in the manuscript. Additional raw data related to this publication cannot be openly released as the raw data contains videos and transcripts of the participants’ interaction with the robot, which were impossible to anonymise.

References

[1] Takuto Akiyoshi, Junya Nakanishi, Hiroshi Ishiguro, Hidenobu Sumioka, and Masahiro Shiomi. 2021. A Robot That Encourages Self-Disclosure to Reduce Anger Mood. IEEE Robotics and Automation Letters 6, 4 (2021), 7925–7932.
[2] Maryam Alimardani, Linda Kemmeren, Kazuki Okumura, and Kazuo Hiraki. 2020. Robot-assisted mindfulness practice: Analysis of neurophysiological responses and affective state change. In 2020 9th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). IEEE, 683–689.
[3] Minja Axelsson, Indu P. Bodala, and Hatice Gunes. 2021. Participatory Design of a Robotic Mental Well-being Coach. In 30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). To Appear.
[4] Nicholas Barnes, Patrick Hattan, David S Black, and Zev Schuman-Olivier. 2017. An examination of mindfulness-based programs in US medical schools. Mindfulness 8, 2 (2017), 489–494.
[5] Pablo Barros, Nikhil Churamani, and Alessandra Sciutti. 2020. The FaceChannel: A Fast and Furious Deep Neural Network for Facial Expression Recognition. SN Computer Science 1, 6 (Oct. 2020).
[6] Pablo Barros, Nikhil Churamani, and Alessandra Sciutti. 2020. The FaceChannel: A Light-weight Deep Neural Network for Facial Expression Recognition. In Proceedings of the 15th International Conference on Automatic Face and Gesture Recognition (FG). 652–656.
[7] Pablo Barros, German Parisi, and Stefan Wermter. 2019. A Personalized Affective Memory Model for Improving Emotion Recognition. In Proc. ICML PMLR, 485–494.
[8] Christoph Bartneck, Dana Kulic, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. International journal of social robotics 1, 1 (2009), 71–81.
[9] Paul D Bliese and Robert E Ployhart. 2002. Growth modeling using random coefficient models: Model building, testing, and illustrations. Organizational Research Methods 5, 4 (2002), 362–387.
[10] Indu P. Bodala, Nikhil Churamani, and Hatice Gunes. 2021. Teleoperated Robot Coaching for Mindfulness Training: A Longitudinal Study. In 30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). 939–944.
[11] Jonathan Bohren. 2018. SMACH. http://wiki.ros.org/smach
[12] Frank Bresser and Carol Wilson. 2010. What is coaching. Excellence in coaching: The industry guide 2 (2010).
[13] Ninawa Butrus and Rivka T. Witenberg. 2013. Some personality predictors of tolerance to human diversity: The roles of openness, agreeableness, and empathy. Australian Psychologist 48, 4 (2013), 290–298.

Manuscript submitted to ACM
[14] Colleen M Carpinella, Alisa B Wyman, Michael A Perez, and Steven J Stroessner. 2017. The robotic social attributes scale (rosas) development and validation. In Proceedings of the ACM/IEEE International Conference on human-robot interaction. 254–262.

[15] Nikhil Churamani, Paul Anton, Marc Brügger, Erik Fleißwasser, Thomas Hummel, Julius Mayer, Waleed Mustafa, hwee Geok Ng, Thi Linh Chi Nguyen, Quan Nguyen, Marcus Soll, Sebastian Springenberg, Sascha Griffiths, Stefan Heinrich, Nicolás Navarro-Guerrero, Erik Strahl, Johannes Twiefel, Cornelius Weber, and Stefan Wermter. 2017. The Impact of Personalisation on Human-Robot Interaction in Learning Scenarios. In Proceedings of the 5th International Conference on Human Agent Interaction (Bielefeld, Germany) (HAI '17). ACM, New York, NY, USA, 171–180. https://doi.org/10.1145/3125739.3125756

[16] Nikhil Churamani, Minja Axelsson, Atahan Çaldır, and Hatice Gunes. 2022. Continual Learning for Affective Robotics: A Proof of Concept for Wellbeing. In 2022 10th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW). IEEE.

[17] Nikhil Churamani, Pablo Barros, Hatice Gunes, and Stefan Wermter. 2022. Affect-Driven Learning of Robot Behaviour for Collaborative Human-Robot Interactions. Frontiers in Robotics and AI 9 (2022).

[18] Nikhil Churamani and Hatice Gunes. 2020. CLIFER: Continual Learning with Imagination for Facial Expression Recognition. In Proceedings of the 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG). IEEE, 322–328.

[19] Nikhil Churamani, Sinan Kalkan, and Hatice Gunes. 2020. Continual Learning for Affective Robotics: Why, What and How?. In 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). 425–431.

[20] Victoria Clarke, Virginia Braun, and Nikki Hayfield. 2015. Thematic analysis. Qualitative psychology: A practical guide to research methods 222 (2015), 248.

[21] Oracle Corporation. 2020. Global Study: 82% of People Believe Robots Can Support Their Mental Health Better Than Humans. https://www.oracle.com/news/announcement/ar-at-work-100720.html

[22] John R. Crawford and Julie D. Henry. 2004. The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. British Journal of Clinical Psychology 43, 3 (Sept. 2004), 245–265. https://doi.org/10.1348/0144665031752934

[23] Lüzie Fofonka Cunha, Lucas Campos Pellanda, and Caroline Tozzi Reppold. 2019. Positive psychology and gratitude interventions: A randomized clinical trial. Frontiers in Psychology 10 (2019).

[24] Kerstin Dautenhahn. 2004. Robots we like to live with? - a developmental perspective on a personalized, life-long robot companion. In IEEE International Workshop on Robot and Human Interactive Communication (Ro-Man). 17–22. https://doi.org/10.1109/ROMAN.2004.1374720

[25] Beatrice De Gelder. 2016. Emotions and the Body. Oxford University Press.

[26] M Brent Donnellan, Frederick L Oswald, Brendan M Baird, and Richard E Lucas. 2006. The mini-IPIP scales: tiny-yet-effective measures of the Big Five factors of personality. Psychological assessment 18, 2 (2006), 192.

[27] Yunfei Duan, Myung Yoon, Zhixuan Liang, and Johan Ferdinand Hoorn. 2021. Self-Disclosure to a Robot: Only for Those Who Suffer the Most. Robotics 10, 3 (2021), 98.

[28] A Lee Duckworth, Tracy A Steen, and Martin EP Seligman. 2005. Positive psychology in clinical practice. Annual review of clinical psychology 1, 1 (2005), 629–651.

[29] Brian R Duffy and Karolina Zawieska. 2012. Suspension of disbelief in social robotics. In 2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication. IEEE, 484–489.

[30] Michael Dzięgwa, Mirela Kaczmarek, Paweł Kaczmarek, Jan Kędzierski, and Karolina Wadas-Szydłowska. 2018. Long-Term Cohabitation with a Social Robot: A Case Study of the Influence of Human Attachment Patterns. JISR 10, 1 (01 Jan 2018), 163–176. https://doi.org/10.1007/s12366-017-0439-2

[31] Robert A Emmons and Charles M Seligman. 2002. Gratitude and the science of positive psychology. Handbook of positive psychology 18 (2002), 459–471.

[32] Maurizio Ficocelli, Junichi Terao, and Goldie Nejat. 2016. Promoting Interactions Between Humans and Robots Using Robotic Emotional Behavior. IEEE Transactions on Cybernetics 46, 12 (2016), 2911–2923.

[33] Shelly L Gable and Jonathan Haidt. 2005. What (and why) is positive psychology? Review of general psychology 9, 2 (2005), 103–110.

[34] Fabian Gander, René T Proyer, and Willibald Ruch. 2016. Positive psychology interventions addressing pleasure, engagement, meaning, positive relationships, and accomplishment increase well-being and ameliorate depressive symptoms: A randomized, placebo-controlled online study. Frontiers in psychology 7 (2016), 686.

[35] John Gideon, Soheil Khorram, Zakaria Aldeneh, Dimitrios Dimitriadis, and Emily Mower Provost. 2017. Progressive Neural Networks for Transfer Learning in Emotion Recognition. In Proc. Interspeech. 1098–1102. https://doi.org/10.21437/Interspeech.2017-1637

[36] Lewis R Goldberg. 1993. The structure of phenotypic personality traits. Psychological assessment 7, 1 (2005), 81–97.

[37] LS Green, LG Oades, and AM Grant. 2006. Cognitive-behavioral, solution-focused life coaching: Enhancing goal striving, well-being, and hope. The Journal of Positive Psychology 1, 3 (2006), 142–149.

[38] James Hardy. 2006. Speaking clearly: A critical review of the self-talk literature. Psychology of Sport and Exercise 7, 2 (2006), 181–191.

[39] Vicki Hart, John Blattner, and Staci Leipinc. 2001. Coaching versus therapy: A perspective. Consulting Psychology Journal: Practice and Research 53, 4 (2001), 229.

[40] Jacqueline Hemminghaus and Stefan Kopp. 2017. Towards Adaptive Social Behavior Generation for Assistive Robots Using Reinforcement Learning. In Proceedings of International Conference on Human-Robot Interaction, ACM/IEEE (Vienna, Austria). 332–340. https://doi.org/10.1145/2999824.3020217

Manuscript submitted to ACM
[41] Guy Hoffman and Xuan Zhao. 2020. A primer for conducting experiments in human–robot interaction. _ACM Transactions on Human-Robot Interaction (THRI)_ 10, 1 (2020), 1–31.

[42] Michal Icht, Yaniv Mama, and Daniel Algom. 2014. The production effect in memory: Multiple species of distinctiveness. _Frontiers in psychology_ (2014), 886.

[43] Sooyeon Jeong, Sharifa Alghowinem, Laura Aymerich-Franch, Kika Arias, Agata Lapedriza, Rosalind Picard, Hae Won Park, and Cynthia Breazeal. 2020. A Robotic Positive Psychology Coach to Improve College Students’ Wellbeing. In 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) IEEE, 187–194.

[44] Alexandra D Kaplan, Tracy Sanders, and Peter A Hancock. 2019. The relationship between extroversion and the tendency to anthropomorphize robots: A Bayesian analysis. _Frontiers in Robotics and AI_ (2019), 135.

[45] Arvid Kappas, Ursula Hess, and Klaus R Scherer. 1991. Voice and emotion. (1991).

[46] Raida Karim, YuFei Zhang, Patricia Alves-Oliveira, Elin A Björling, and Maya Cakmak. 2022. Community-Based Data Visualization for Mental Well-being with a Social Robot. In _IRB_ 839–843.

[47] Cory D Kidd and Cynthia Breazeal. 2007. A robotic weight loss coach. In _Proceedings of the national conference on artificial intelligence_, Vol. 22. Menlo Park, CA, Cambridge, MA, London: AAAI Press; MIT Press; 1999, 1985.

[48] Cory D Kidd and Cynthia Breazeal. 2008. Robots at home: Understanding long-term human-robot interaction. In _IEEE/RSJ International Conference on Intelligent Robots and Systems_. 3230–3235.

[49] Andrea Kleinsmith and Nadia Bianchi-Berthouze. 2013. Affective Body Expression Perception and Recognition: A Survey. _IEEE Transactions on Affective Computing_ 4, 1 (2013), 15–33. https://doi.org/10.1109/T-AFFC.2012.16

[50] Kurt Kroenke, Robert L Spitzer, and Janet BW Williams. 2001. The PHQ-9: validity of a brief depression severity measure. _Journal of general internal medicine_ 16, 9 (2001), 606–613.

[51] Iolanda Leite, André Pereira, Samual Mascarenhas, Carlos Martinho, Rui Prada, and Ana Paiva. 2013. The influence of empathy in human–robot relations. _International Journal of Human-Computer Studies_ 71, 3 (2013), 250–260.

[52] Luming Li. 2020. Challenges and Priorities in Responding to COVID-19 in Inpatient Psychiatry. _Psychiatric Services_ 71, 6 (2020), 624–626.

[53] Ming Li, William H Mohley, and Aidan Kelly. 2016. Linking personality to cultural intelligence: An interactive effect of openness and agreeableness. _Personality and Individual Differences_ 89 (2016), 105–110.

[54] S. Li and W. Deng. 2020. Deep Facial Expression Recognition: A Survey. _IEEE Transactions on Affective Computing_ (March 2020).

[55] A. Lindt, P. Barros, H. Siqueira, and S. Wermter. 2019. Facial Expression Editing with Continuous Emotion Labels. In _Proceedings of the 14th IEEE International Conference on Automatic Face and Gesture Recognition (FG)_ IEEE, 1–8.

[56] Hadassah Littman-Ovadia and Dina Nir. 2014. Looking forward to tomorrow: The buffering effect of a daily optimism intervention. _The Journal of Positive Psychology_ 9, 2 (2014), 122–136.

[57] Katrin Solveig Lohan, Hagen Lehmann, Christian Dondrup, Frank Broz, and Hatice Kose. 2016. _Enriching the Human-Robot Interaction Loop with Natural, Semantic, and Symbolic Gestures_. Springer Netherlands, Dordrecht, 1–21. https://doi.org/10.1007/978-94-007-7194-9_136-1

[58] Shane J Lopez, Jennifer Teramoto Pedrotti, and Charles Richard Snyder. 2018. Positive psychology: The scientific and practical explorations of human strengths. Sage publications.

[59] Colin M MacLeod, Nigel Gopie, Kathleen L Hourihan, Karen R Neary, and Jason D Ozubko. 2010. The production effect: delineation of a phenomenon. _Journal of Experimental Psychology: Learning, Memory, and Cognition_ 36, 3 (2010), 671.

[60] H B Mann and D R Whitney. 1947. On a test of whether one of two random variables is stochastically larger than the other. _The Annals of Mathematical Statistics_ 18 (1947), 50–60. https://doi.org/10.1214/aoms/1177730491

[61] Derek McColl, Alexander Hong, Naasaki Hatakeyama, Goldie Nejat, and Beno Benhabib. 2015. A Survey of Autonomous Human Affect Detection Methods for Social Robots Engaged in Natural HRI. _Journal of Intelligent & Robotic Systems_ 82, 1 (Aug. 2015), 101–133. https://doi.org/10.1007/s10846-015-0259-2

[62] Ross Mead and Maja J Mataric. 2016. Autonomous Human-Robot Proxemics: Socially aware Navigation based on Interaction Potential. _Autonomous Robots_ 41, 5 (June 2016), 1189–1201.

[63] Joseph Melnick and Sonia March Nevis. 2005. The willing suspension of disbelief: Optimism. _Gestalt Review_ 9, 1 (2005), 10–26.

[64] M Christina Meyers, Marianne van Woerkom, and Arnold B Bakker. 2013. The added value of the positive: A literature review of positive psychology interventions in organizations. _European Journal of Work and Organizational Psychology_ 22, 5 (2013), 618–632.

[65] Blanca Müller and David Feil-Seifer. 2016. _Embodiment, Situatedness, and Morphology for Humanoid Robots Interacting with People_. Springer, 1–23. https://doi.org/10.1007/978-3-319-130-130-1

[66] Carmen Moreno, Til Wykes, Silvana Galderisi, Merete Nordentoft, Nicolas Crosseley, Nev Jones, Mary Cannon, Christoph U Correll, Louise Byrne, Sarah Carr, et al. 2020. How mental health care should change as a consequence of the COVID-19 pandemic. _The Lancet Psychiatry_ (2020).

[67] Jonathan Mumma and Bilge Mutlu. 2011. Human-Robot Proxemics: Physical and Psychological Distancing in Human-Robot Interaction. In _Proceedings of the 6th International Conference on Human-Robot Interaction (HRI)_ ACM, Lausanne, Switzerland, 331–338.

[68] NHS UK. 2018. Mindfulness. https://www.nhs.uk/mental-health/self-help/tips-and-support/mindfulness/

[69] Jakob Nielsen. 1993. Iterative user-interface design.

[70] Warren T Norman. 1963. Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. _The journal of abnormal and social psychology_ 66, 6 (1963), 574.

Manuscript submitted to ACM
Arielle AJ Scoglio, Erin D Reilly, Jay A Gorman, and Charles E Drebing. 2019. Use of social robots in mental health and well-being research. 

Martin EP Seligman and Mihaly Csikszentmihalyi. 2014. Positive psychology: An introduction. In Commun. ACM.

Björn W. Schuller. 2018. Speech Emotion Recognition: Two Decades in a Nutshell, Benchmarks, and Ongoing Trends. The Journal of Positive Psychology.

LA Witt. 2002. The interactive effects of extraversion and conscientiousness on performance. Journal of management.

Biometrics Bulletin.

Frank Wilcoxon. 1945. Individual Comparisons by Ranking Methods. Biometrics Bulletin.

Astrid Weiss and Christoph Bartneck. 2015. Meta analysis of the usage of the godspeed questionnaire series. In The Australian Educational and Developmental Psychologist.

James A Russell. 1980. A circumplex model of affect. Journal of personality and social psychology.

DG Williams. 1990. Effects of psychoticism, extraversion, and neuroticism in current mood: A statistical review of six studies. Personality and Individual Differences.

Anouk van Maris, Alexander Sutherland, Alexandre Mazel, Sanja Dogramadzi, Nancy Zook, Matthew Studley, Alan Winfield, and Praminda Caleb-Solly. 2020. The Impact of Affective Verbal Expressions in Social Robots. In Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (Cambridge, United Kingdom) (HRI ’20). Association for Computing Machinery, New York, NY, USA, 508–510. https://doi.org/10.1145/3371382.3378358

Philippe Verduny and Karen Brans. 2012. The relationship between extraversion, neuroticism and aspects of trait affect. Personality and Individual Differences 52, 6 (2012), 664–669.

Lea Waters. 2011. A review of school-based positive psychology interventions. The Australian Educational and Developmental Psychologist 28, 2 (2011), 75–90.

David Watson, Lee Anna Clark, and Auke Tellegen. 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales. Journal of personality and social psychology 54, 6 (1988), 1063.

Astrid Weiss and Christoph Bartneck. 2015. Meta analysis of the usage of the godspeed questionnaire series. In 2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). IEEE, 381–388.

Frank Wilcoxon. 1945. Individual Comparisons by Ranking Methods. Biometrics Bulletin 1, 6 (Dec. 1945), 80. https://doi.org/10.2307/3001968

DG Williams. 1990. Effects of psychoticism, extraversion, and neuroticism in current mood: A statistical review of six studies. Personality and Individual Differences 11, 6 (1990), 615–630.

Lena Wimmer, Lisa von Stockhausen, and Silja Bellingrath. 2019. Improving emotion regulation and mood in teacher trainees: Effectiveness of two mindfulness trainings. Brain and behavior 9, 9 (2019).

LA Witt. 2002. The interactive effects of extraversion and conscientiousness on performance. Journal of management 28, 6 (2002), 835–851.

A. Zhang. 2017. Speech Recognition (Version 3.8) [Software]. https://github.com/Uberi/speech_recognition
A POST-INTERACTION INTERVIEW

The questions asked during the semi-structured post-interaction interview:

1. Robot
   - What do you think was good about the robot’s appearance? What did you like?
   - What do you think was bad about the robot’s appearance? What would you change?
   - How appropriate do you think the robot was for helping you focus on the positive aspects in your life? How appropriate do you think the robot was for helping you foster a positive attitude toward things that happen in your life?
   - How appropriate do you think the robot was for these coaching tasks?
   - How useful do you think the robot was for these coaching tasks?
   - How beneficial do you think the robot was for these coaching tasks?

2. Behaviour
   - What do you think was good about the robot’s behaviour? What did you like?
   - What do you think was bad about the robot’s behaviour? What would you change?
   - How appropriate do you think the robot’s behaviour was for these coaching tasks?
   - How useful do you think the robot’s behaviour was for these coaching tasks?
   - How beneficial do you think the robot’s behaviour was for these coaching tasks?
   - Do you think the robot understood what you said?
   - Do you think the robot understood how you felt?
   - Do you think the robot adapted to what you said and did?
   - Did you notice any differences?