Robotic Mental Well-being Coaches for the Workplace: An In-the-Wild Study on Form

Micol Spitale
ms2871@cam.ac.uk
University of Cambridge, UK

Minja Axelsson
minja.axelsson@cl.cam.ac.uk
University of Cambridge, UK

Hatice Gunes
hatice.gunes@cl.cam.ac.uk
University of Cambridge, UK

ABSTRACT

The World Health Organization recommends that employers take action to protect and promote mental well-being at work. However, the extent to which these recommended practices can be implemented in the workplace is limited by the lack of resources and personnel availability. Robots have been shown to have great potential for promoting mental well-being, and the gradual adoption of such assistive technology may allow employers to overcome the aforementioned resource barriers. This paper presents the *first* study that investigates the deployment and use of two different forms of robotic well-being coaches in the workplace in collaboration with a tech company whose employees (26 coachees) interacted with either a QTrobot (*QT*) or a Misty robot (*M*). We endowed the robots with a coaching personality to deliver positive psychology exercises over four weeks (one exercise per week). Our results show that the robot form significantly impacts coachees' perceptions of the robotic coach in the workplace. Coachees perceived the robotic coach in *M* more positively than in *QT* (both in terms of behaviour appropriateness and perceived personality), and they felt more connection with the robotic coach in *M*. Our study provides valuable insights for robotic well-being coach design and deployment, and contributes to the vision of *taking robotic coaches into the real world*.

CCS CONCEPTS

• **Human-centered computing** → *User studies*; **HCI design and evaluation methods**.

KEYWORDS

human-robot interaction, robotic coach, mental well-being, robot form, robot personality, longitudinal study, in the wild

ACM Reference Format:
Micol Spitale, Minja Axelsson, and Hatice Gunes. 2023. Robotic Mental Well-being Coaches for the Workplace: An In-the-Wild Study on Form. In *Proceedings of the 2023 ACM/IEEE Int'l Conference on Human-Robot Interaction (HRI '23), March 13–16, 2023, Stockholm, Sweden.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3568162.3577003

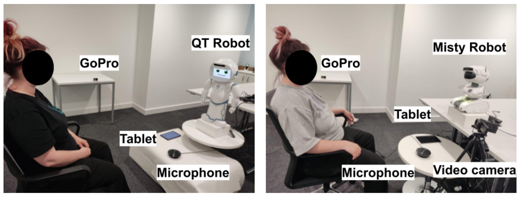
1 INTRODUCTION

According to the World Health Organization (WHO) [40], work may help with protecting the mental well-being of employees, giving them purpose, a sense of achievement, and opportunities to feel part of a community. On the other hand, work can also pose risks to mental well-being, due to excessive workload, tight work schedules, and inadequate work-life balance. Therefore, the WHO recommends that employers take appropriate actions to prevent work-related mental well-being issues, and protect and promote mental well-being at work [40]. However, lack of resources and personnel availability limits the extent to which these recommendations can be put into practice. Human-robot interaction (HRI) research (e.g., [8, 16, 60]) suggests that robots have great potential for improving and sustaining human well-being, and their adoption in the workplace might enable the employers to overcome existing barriers. Past works have explored the use of robots as coaches for promoting human well-being in various contexts, e.g., promoting physical exercises for the elderly [16], and supporting mental well-being [1, 3, 4, 8, 12]. Most of these works are limited to lab settings [27] due to the problems associated with running studies in the real world, such as availability of a host organisation, ethical concerns, and technical set-up challenges. For instance, [4, 12] explored the use of Pepper as a robotic coach to deliver positive psychology exercises in the lab. Very few studies have investigated the use of robots to promote mental well-being in real-world contexts – e.g., the works by Jeong et al. [25] and Ostrowski et al. [41] focused on home settings. None of these works investigated the deployment or the use of robots as mental well-being coaches in the workplace. Various design aspects, such as robot form, influence how people perceive robots [22, 33, 51]. In the context of delivering mental well-being exercises, Axelsson et al. [3] reported on the interplay between different forms and the tasks undertaken by the robot, as participants expected the robot's form to match its functionality (i.e., they wouldn't expect a robotic dog to speak, but they would expect conversation abilities from a humanoid robot). None of these studies have investigated the influence of different robot forms on delivering mental well-being exercises in the wild. In this paper, we present the *first* study that investigates the deployment and use of two forms of robotic mental well-being coaches *in the workplace*. To this end, we collaborated with a tech company (Cambridge Consultants Inc. [1]) whose employees (26 coachees) interacted with either a QTrobot or a Misty robot over 4 weeks. The robots delivered four positive psychology exercises (one exercise/week). We designed the robot personality to reflect a well-being coach, in collaboration with two (human) well-being coaches and informed by relevant literature [14, 55]. We gathered quantitative data via standardized and specifically designed questionnaires, and qualitative data from in-person interviews and focus groups, combining multiple methods [45] to develop a comprehensive understanding of the use of robotic coaches to promote well-being in the workplace.

HRI '23, March 13–16, 2023, Stockholm, Sweden

ACM ISBN 978-1-4503-9964-7/23/03.
https://doi.org/10.1145/3568162.3577003

[1] https://www.cambridgeconsultants.com



(**QT**) QTrobot (**M**) Misty robot

Figure 1: Setup of the study. All sensors are located and used similarly in both conditions.

2 RELATED WORK

Robotic Coaches For Mental Well-being. Mental well-being coaching aims to help a mentally healthy coachee flourish in life or work (cf. psychological therapy, which aims to treat mental illness) [23]. General goals for coaching include increasing the coachee's hope, goal-striving, and general well-being [19]. Different styles of coaching may draw from different psychological practices, e.g., *Cognitive Behavioural* coaching focuses on the relationship between thoughts, feelings, and actions [19], and coaching based on *Positive Psychology* encourages the coachee to pay greater attention to the positive aspects of their life [52]. The success of the coaching practice also depends on the working alliance between the coachee and the coach [13]. This alliance relies on trust and is improved by transparency [21]. The quality of therapeutic working alliances have been previously examined by measuring the development of an affective bond between a therapist and their client, together with measurements of the agreement on the tasks and goals of therapy [36]. Very few works have explored the use of robotic coaches to promote mental well-being. Jeong et al. [25] studied the use of the Jibo robot that facilitated positive psychology interventions for students in home settings in a longitudinal study (7 days). Their results showed improvements in student well-being, mood, and readiness to change, and that participants built an alliance with the robot over the sessions. Bodala et al. [8] evaluated the participants' perceptions of a human versus a teleoperated robotic mindfulness coach in a 5-week study, reporting that while both coaches received positive feedback, the human coach was evaluated significantly higher on animacy, likeability and perceived intelligence. The participants' *neuroticism* and *conscientiousness* traits also affected how they perceived the robot. A recent study investigated the use of a robot to assess children's mental health [1]. Their findings showed that compared to the self-report and parent-report standard tests, the mental well-being evaluation using the robot appeared to be the most suitable for identifying well-being-related anomalies in children. None of these works investigated the use of robots as mental well-being coaches in the workplace.

Robot Form in HRI. Form influences how people perceive robots, namely through the *form function attribution bias* [22]. This phenomenon uses visual information as a cognitive shortcut to attribute certain capabilities and functionalities to robots. Past works have demonstrated the importance of robot form and how it could impact human-robot interactions [22, 33, 43, 51]. Most of these studies (e.g., [33, 47, 51]) investigated form by showing people static pictures of the different robotic platforms. Schaefer et al. [51] conducted a survey study involving university students who evaluated pictures of robots. Their results showed that physical form impacted the perceived trustworthiness of robots. Similarly, Li et al. [33] investigated how the robot form and motion influence human social attention by showing participants pictures and videos of different robotic platforms. Their results demonstrated that differences in the form of the agent (robot vs. android vs. human) impacted social attention, specifically how quickly the user could disengage attention from the robot and respond (the attentional capture). Only a small number of works undertook user studies where participants actually interacted with the different robotic platforms. For instance, Paetzel et al. [43] presented an empirical study that investigated the persistence of first impressions between varying levels of human-likeness of a Furhat robot across repeated sessions. Their results showed that perceptual differences between the human-likeness conditions of the robot persist across repeated interactions. As far as we know, none of these works investigated the variations in human perception when participants interacted with *different* robotic platforms (forms) for the *same* HRI task.

Robot Personality in HRI. People can accurately recognize robot personality based on verbal and nonverbal behaviours [32, 63]. Robot and virtual agent personality has been previously designed by varying behavioural variables such as the speed [65] and the frequency of gestures [7, 14] and word choices [55, 65]. However, users' perception of personality can also be influenced by other design dimensions, including a robot's form [9]. Robot personality has been shown to influence interaction outcomes. For example, matching extroverted people with extroverted robots (and vice versa) can improve motivation [2], and can influence preference [59] and social attraction [32], and conscientious robots can weaken uncanny feelings [44]. Despite these encouraging works, there is a lack of a systematic understanding of robot personality in HRI [50], and specifically in the context of HRI for health care [15]. Most commonly, robot personality is expressed through the OCEAN model [18, 38], also known as the Big Five, Personality Trait model (see Section 4.1.4 for further details) [15, 50]. Overall though, there is a lack of standardized, open, and commonly used tool for designing and measuring robot personality.

3 RESEARCH QUESTIONS

Our vision is to create fully autonomous mental well-being coaches that can be successfully used in various contexts in the real world, including the workplace. In this work, we undertake the *first study* that explores the use of two different forms of robotic well-being coaches in the workplace, focusing on factors driven by the literature review provided above. Form impacts how people perceive robots [22]. Past works [33, 47, 51] investigated how different robotic forms affect users' perception, mainly via viewing robot pictures, with some evaluating robot form via interaction studies [43]. No work to date has investigated whether and how human perceptions vary when they interact with *different* robotic platforms for the *same* HRI task. Therefore, in this work we investigate how coachees perceive the interaction with two different robotic platforms (forms) delivering positive psychology exercises, and how the *robot form* influences their perceptions of the robotic coach in the workplace (**RQ1**). Personality has a key role in the perceptions of a robot during human-robot interactions [50] . The design of an "optimal" personality profile for robots in a specific context is still a challenging problem. People's perceptions of a robot's behaviour depend on people's demographic background, and their own personality may influence their preference for certain robot personality traits [29]. Therefore, in this work we designed and created a robot personality that aims to reflect a well-being coach, and we evaluated if this personality was perceived the way it was intended and whether the perceptions differ due to other factors such as form (**RQ2**). A good relationship between the coach and the coachee is central to the success of a coaching practice [13]. Previous work has shown the importance of working alliance between a robotic coach and a coachee in the success of the well-being exercises [25]. Hence, we explored the perceptions of the *coachee-coach alliance* created during positive psychology exercises in the workplace and how that working alliance differed across two different robotic forms (**RQ3**).

4 THE STUDY

Our work is the *first study* that investigates the use of two different forms of robotic well-being coaches in the workplace via a between-subjects study where we compared the perceptions of 26 participants (coachees) who interacted with two different robots—a QTrobot and a Misty robot—endowed with the same coach personality that delivered the same positive psychology exercises over 4 weeks.

4.1 Materials and Methods

4.1.1 Setup. The study was conducted in a meeting room (see Figure 1) of the Cambridge Consultants Inc. headquarters. The set-up included a big table, a chair, two cameras—a GoPro (on the left-side of the room) and a video camera (placed near the robot)—a microphone (in front of the robot on a low table), and a tablet (close to the microphone). The stationary robot was placed on the big table and the coachee was on the chair about 1.2m from the robot (as in [57]). The COVID-19 sanitary guidelines were followed according to the Cambridge Consultants' office regulations.

4.1.2 Participants. We involved 26 participants in total, 6 women, 1 non-binary person, and 19 men, among which 7 were 18-25 years old, 11 were 26-35 years old, 4 were 36-45 years old, and 4 were 46-55 years old. The study was supported and promoted by Cambridge Consultants, and all participants were Cambridge Consultants employees. The gender distribution in our study reflects the employee gender distribution of the tech company (i.e., the company has more male than female / non-binary employees). The company advertised the study through their communication channels (e.g., newsletter, flyers in their canteen) and participation was voluntary (with no compensation). Our study aimed to involve healthy participants in the first instance, therefore we screened 41 participants and recruited 26. 15 participants were excluded based on their self-reported levels of anxiety and depression, scoring more than 5 (the threshold for mild anxiety disorder) in the Generalized Anxiety Disorder 7 (GAD-7) [58] and more than 5 (the threshold for mild depression severity) in Patient Health Questionnaire (PHQ-9) [34]. Participants had very little knowledge (on average 1.3 on a 5-point Likert scale) of robotic technology. All participants provided informed consent for their participation and agreed on the usage of their data for scientific research. The study design, the experiment protocol, and the consent forms were approved by the Computer Science and Technology Departmental Ethics Committee of the University of Cambridge. In the rest of the paper, we will refer to the participants as *coachees*.

4.1.3 Robotic Platforms. We used the QTrobot by LuxAI S.p.A. and Misty II robot by Misty Robotics as they have been used in previous HRI studies [28, 42, 57]. The QTrobot is a 90 cm tall, tabletop child-like robot. The body components are white and include a screen face (with a white background, green eyes, and a black line-mouth) embedded into a robot-like head (2 DOF neck), a human-like upper body (4 DOF full arms, i.e., shoulders, elbows, and hands), and a human-like lower body (static legs). Its overall score of human-like appearance is 45.65 according to the Anthropomorphic Robot Database[2]. The Misty II is a 36 cm tall, toy-like robot. The body components are white and include a screen face (with a black background, green eyes, and a white line-mouth) embedded into a robot-like head (3 DOF neck), an upper chest (1 DOF half-arms, i.e., no elbows and hands), and a navigation base. We chose those robots because they are both equipped with a face screen and we could control facial appearance across the forms, e.g., facial expressions, lip-sync etc. Additionally, the robots are smaller, making them portable in a workplace environment, and affordable, making them a realistic investment for a workplace to purchase and use. We collaborated with a well-being professional to pick the robots' voices and gestures. We used the synthesised AWS Polly's Amy voice, and Amazon Polly visemes to synchronise (lip-sync) the robot's mouth positions with the spoken voice (the same for both platforms). In addition, we designed movements for the robot's head (e.g., nodding when listening to the coachee) and arms (e.g., lifting and waving the right arm to greet the coachee in the beginning of the interaction). The robots also transcribed the speech of the coachee using a local automatic speech recognition (ASR) module (DeepSpeech). The interaction flow was pre-scripted and we did not equip the robot with any natural language processing capability because of the current limitations of the local ASR modules. If the interaction flow depended on what the coachee said and the ASR failed in transcribing the coachee's speech, it could have led to a very negative and frustrating experience for the coachee, especially during a positive psychology exercise. In addition, the literature [5] recommends adhering to the script of the specific practice for successfully delivering mental well-being interventions. We defined the robots' level of autonomy using the framework in [6] as follows: *sense* (fully autonomous. i.e., the robot uses its microphone and camera to collect video and audio signals, and was able to transcribe the user speech automatically), *plan* (autonomous, i.e., using pre-programmed decisions that didn't change upon the coachee behavior), and *act* (fully autonomous). In the rest of the paper, with the term *robot form* we will refer not only to the appearance of these two robotic platforms but also to the other robot-related characteristics (e.g., motor noise, degree, speaker characteristics etc.) that are platform-dependent and could to a certain extent impact the coachees' perceptions.

[2]https://www.abotdatabase.info/collection

4.1.4 Robot Personality. Joosse et al. [26] argued that robot personality should be designed according to what users would expect in the context of a task. We designed the robot's personality to be appropriate for a well-being coach. We examined the literature on preferred personality in human coaches, and interviewed two practicing coaches on their thoughts about appropriate coach personality. We defined the appropriate personality in terms of the OCEAN model of personality (defined by the personality traits of Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism) [18, 38]. As a result of our literature survey and interviews with two coaches, we determined that the most important coach traits were high Openness (i.e. listening non-judgmentally) [17, 64], and high Conscientiousness [64] (i.e. being dependable). Additionally, the coach should have a medium level of Extraversion (so as to be conversational but not take up too much space from the coachee) [17], medium to high Agreeableness (a mix of validation and moving the person towards change) [17, 64], and low Neuroticism (a self-conscious coach would distract from the coaching) [17]. To reflect these personality traits, we consulted previous HRI literature on robot behaviour design (i.e., aspects of the robot's voice, gestures and facial expressions) with respect to personality. The collated behaviours can be seen in Figure 3. We implemented behaviours following these choices for both robotic platforms (i.e., the platforms shared the same pace and pitch of synthesised voice, same frequency and style of gesturing, etc.).

4.1.5 Exercises. We scripted 4 positive psychology (PP) exercises with the help of a well-being coach professional, because PP practice has been shown to be successfully delivered by robots in previous studies [3, 25]. The exercises were adapted from 4 existing PP exercises to the robot. The 4 PP exercises were delivered on a weekly basis in the following order. *Savouring exercise* (week 1) consists of asking coachees to choose a small moment to fully feel and appreciate positive experiences that one normally hurries through (adapted from [54]). *Gratitude exercise* (week 2) consists of asking coachees to recall two things that they felt grateful for during the past week (adapted from [20]). *Accomplishments exercise* (week 3) consists of asking coachees to talk about two accomplishments achieved during the last week (adapted from [20]). *Future optimism exercise* (week 4) consists of asking coachees to imagine their optimistic future and the steps along the way to get there (adapted from [53]).

4.1.6 Study Conditions. The study was conducted as a between-subjects study where each coachee was randomly assigned to one of the two conditions that corresponded respectively to the two different robotic platforms with varying levels of anthropomorphism: the child-like **QTrobot** (**QT**, 14 coachees: 2 women, 1 non-binary person, and 11 men) and the toy-like **Misty II** (**M**, 12 coachees: 4 women and 8 men). Over 4 weeks, each coachee engaged in 4 exercises (one for each week) with the robotic coach assigned to them. The exercises were the same for both conditions.

4.1.7 Experiment Protocol. Two weeks prior to the study, we asked the coachees to fill out 6 standardized questionnaires reported in detail in Section 4.1.9. The study consisted of 4 sessions that were conducted by two researchers. In each session, one of the researchers welcomed the coachee and asked them to enter the meeting room and sit on the chair in front of the robot. The other researcher started the recordings of the session (see Figure 1). Then, both researchers left the room, leaving the coachee alone with the robot. The one-to-one interaction with the robot lasted for about 10 minutes and consisted of the following steps. (1) The robot welcomed the coachee. (2) The robot introduced the exercise of the day. For example, in week 1, the robot described the savouring exercise (E1), highlighting the benefits of this practice. (3) The robot asked the coachee about the exercise. Again for E1, this includes questions like *"What is it about the experience that you find so positive? Please share with me"*. (4) The robot listened to the coachee's answers spoken aloud. (5) The robot concluded the session by asking the coachee to fill out a questionnaire on the tablet before leaving the room, thanking them, and reminding them of the following week's session. This protocol was repeated for all the sessions. At the end of the study, coachees were asked to fill out a set of questionnaires that included the same 6 questionnaires used before the study, as well as additional ones that are reported in Section 4.1.9. After that, the researchers conducted a semi-structured interview with each coachee (as described in Section 4.1.9).

4.1.8 Implementation. The robot interactions were implemented using the open-source HARMONI framework [56]. HARMONI enables the composition of interactions, using a set of open-source modules (e.g., DeepSpeech speech-to-text), and porting the same interaction code into different robotic platforms. Within the HARMONI framework, we designed and implemented study behavior trees (using the python py-tree library[3]) to plan for and code the interactions. Note that the behavior trees implemented were the same for both robotic platforms, and they were only handling the interaction flow and not modifying the content of the exercise.

4.1.9 Measures. We contacted the coachees via e-mail to fill out the 6 *standardized questionnaires* two weeks before the study took place: a demographic form (asking their age, gender, and previous experience with well-being practices as well as robots), the Short Big 5 personality test (IPIP-BFM-20 [61] to assess coachees' personality traits as in [4, 25]), the Negative Attitude Towards Robots Scale (NARS [37] to measure coachees' negative attitudes towards robots before interacting with them as in [57]), Ryff's Psychological Well-being Scale (RPWS [62] to assess coachees' mental well-being as in [25]), the Satisfaction with Life Scale ([46] to measure how much coachees were satisfied with their lives as in [51]), and the mood and readiness to change scale ([10] to measure coachees' willingness to change as in [25]). At the end of the study, coachees were asked to fill out the same 6 questionnaires used before the study, as well as the following *standardized* and *specifically designed questionnaires*: the Working Alliance Inventory Short Revised covering *task*, *goal*, and *bond* (WAI-SR [36] - 12 items on a 5-point Likert scale to measure the alliance between the coachee and the robot during the well-being interventions as in [25]), the Robotic Social Attributes Scale questionnaire covering *warmth*, *competence*, and *discomfort* (RoSAS [11] - 18 items on a 7-point Likert scale to evaluate coachees' perception of the robot as in [31]), questions about the robot's behavior (5-point Likert scale, e.g., *"The robot's voice is high pitched"*), and questions about the perceived personality of the robot (5-point Likert scale, e.g., *"Would you describe the robot*

[3]https://py-trees.readthedocs.io/en/devel/

as Extroverted? Extroverted people are outgoing and energetic - as opposed to solitary and reserved."). In the final session, each coachee was also interviewed using additional questions (see the Supplementary Material). A week after the study ended, the company organized two internal *focus groups* (one for each condition group) where the employees (coachees) freely expressed their opinion on using robotic coaches in the workplace without the researchers' involvement. We requested the company to ask their employees what they thought was important rather than prompting them to collect data according to our specifications. They designed the focus group internally and then shared Miro boards where employees provided their opinions on the chosen topic (e.g., interaction expectations). To support anonymity and privacy, these sessions were not recorded, instead, the company provided us with a collated document.

4.2 Data Analysis

For data analysis, we adopted a mixed-method approach in which we analyzed both quantitative (i.e., standardized and specifically designed questionnaires) and qualitative (i.e., interviews and focus groups) data to obtain a comprehensive understanding on robots as mental well-being coaches in the workplace. We analyzed the quantitative data from the pre- and post-study questionnaires using Python statistical libraries. We conducted non-parametric tests because our samples do not follow a normal distribution. In particular, we ran a Mann–Whitney U test to compare measurements in the two conditions (QT and M), while we used a Wilcoxon signed-rank test to compare the measurements pre- and post-study with the Bonferroni correction. We also analysed the differences in coachee's perceptions within sessions (over time) and between conditions (robot form); however, we did not find any statistical differences across the repeated measures. We applied the *framework method* to analyse qualitative data [49] collected from semi-structured post-interaction interviews, conducted after the final session. The *framework method* consists of five key stages: 1) familiarization with the data, 2) identifying a thematic framework, 3) indexing, 4) charting, and 5) mapping and interpretation. We construct our framework by drawing on the three research questions we established a priori (see Section 3) while allowing for other emergent observations in the data. During the charting stage, we explicitly compare the two groups (QT and M) in the three research questions.

5 FINDINGS

This section reports the main findings of the data analysis, via quantitative and qualitative results to address our research questions. Our analysis highlighted that the form has a major impact on the perception of the robots as well-being coaches, robot personality, and coachee-coach alliance.

Perception of the Robots as Well-being Coaches (RQ1). We conducted Mann–Whitney U tests to compare the coachees' perceptions of the two robotic platforms (QT and M). We found that the coachees perceived the two robotic coaches significantly differently after the four weeks for the RoSAS *warmth* sub-scale ($U = 42.50$, $z = -2.11$, corrected $p^{*}<0.05$), while no significant differences were found for the RoSAS *competence* and *discomfort* sub-scales. The RoSAS *warmth* sub-scale was significantly higher for coachees who

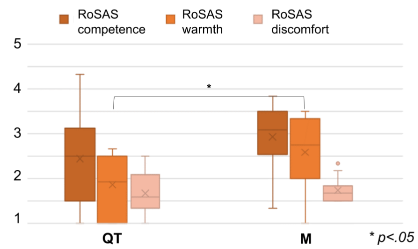


Figure 2: RoSAS sub-scales (competence, warmth, and discomfort) in the QTrobot (QT) and Misty robot (M) conditions.

interacted with the Misty robot (M, *Mdn* = 2.75) than the perceived warmth of coachees who interacted with the QTrobot (QT, *Mdn* = 1.91), as depicted in Figure 2. The RoSAS sub-scale *competence* for QT was *Mdn* = 1.91 and for M was *Mdn* = 1.91, the *discomfort* sub-scale for QT was *Mdn* = 1.91,and for M was *Mdn* = 1.91. We performed Mann–Whitney U tests to compare NARS scores between the two conditions before and after the study. We did not find any differences in the NARS scores before the study - i.e., they were evenly distributed between QT and M. We observed a difference ($U = 44.50$, $z = -2.006$, corrected p^{*}=0.044) in *negative attitude* towards the robots after the study. The NARS score for M (*Mdn* = 34.5) was higher than for QT (*Mdn* = 30). Interestingly, coachees reported that they perceived the robotic coaches interacting differently across the conditions (the robots' interactive behavior were designed to be identical), see also Table 1 of the Supplementary Material. On average the coachees found the robot gestured appropriately (QT: M=2.86, SD= 1.56 and M: M = 3.00, SD=0.85), and coachees in M scored the robot more positively than in QT in terms of listening (QT: M=2.14, SD= 1.23 and M: M = 3.08, SD=0.90), caring about what they said (QT: M=1.86, SD= 1.03 and M: M = 1.92, SD=0.67), and naturalness (QT: M=2.21, SD= 1.12 and M: M = 2.92, SD=1.08). Coachees also reported that they perceived that the robot acknowledged them when they spoke in M (QT: M=2.64, SD= 1.28 and M: M = 3.25, SD=1.06), and that it adapted more to what they said and did in M than in QT (QT: M=1.79, SD= 1.12 and M: M = 2.58, SD=0.90). Most of the coachees did not believe that the robot was able to understand what they said (QT: M=1.71, SD= 0.83 and M: M = 2.08, SD=0.79) or how they felt during the interaction (QT: M=1.50, SD= 0.85 and M: M = 1.75, SD=0.87), but they still had a more positive perception in M than in QT.

Interview results supported the quantitative findings. In general, coachees in QT viewed the robot more negatively in terms of coaching interactive behavior. For example, P3 (QT) described that the robot would "need to be a conversation partner first", before attempting well-being coaching. P7 (QT) noted that while the exercises were useful, the robot "didn't add any value", and P8 (QT) noted that the robot was "not showing any care". Fewer coachees in M seemed to have these experiences, with P23 (M) explicitly stating that the robot made them feel "a lot more engaged" in comparison to hypothetically doing the exercise on their own. While negative attitudes were also present in M, coachees in this condition seemed to be more lenient toward the robot when talking about its lack of responsiveness and a "scripted" impression. P24 (M) noted "it's

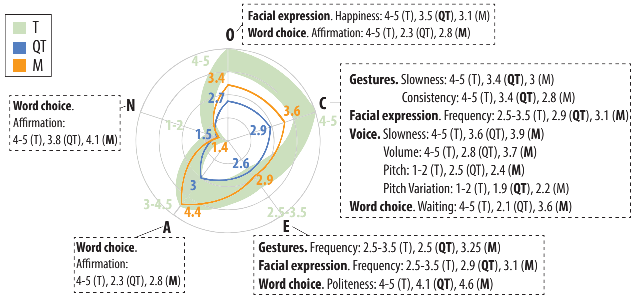


Figure 3: The robot's target personality (T) vs. its perceived personality for the QTrobot (QT) and Misty robot (M) conditions for Openness (O), Conscientiousness (C), Extroversion (E), Agreeableness (A), and Neuroticism (N) traits. Bold highlights the score closer to the target personality.

a calming presence, it doesn't need to be super reflective to what I say", and P15 (M) stated that "it doesn't really matter as much what Misty is doing, but the exercise that she's giving me". In QT, coachees mainly mentioned the robot being helpful by creating an environment where they could learn PP exercises, and take their time to think about the positive aspects in life. In M, coachees noted similar things and some (P23, P26) also explicitly mentioned feeling better after the exercises. In QT and M, coachees wished the robot would ask them follow-up questions related to certain answers, and that they could ask the robot for clarification about exercises (they had tried this, but the robot was not able to fulfil their requests).

Robot Personality (RQ2). The robot personalities were designed to be the same and express a robotic well-being coach personality (as defined by behavioural variables, in Figure 3 and in Table 3 of the Supplementary Material) in both conditions. The robot's target personality was represented by the following values for each trait (see Figure 3): Openness ($4-5$), Conscientiousness ($4-5$), Extroversion ($2.5-3.5$), Agreeableness ($3-4.5$), and Neuroticism ($1-2$). However, our results showed interesting differences between the two conditions. We conducted Mann–Whitney U tests to compare the coachees' perceptions of the robot's personality for the two conditions (QT and M). Our results show a statistically significant difference ($U = 35.00$, $z = -2.495$, corrected p^{*}<0.05) in terms of appropriateness of perceived robot personality. Coachees who interacted with the Misty robot (M, Mdn = 3.92) perceived the robotic coach's personality to be significantly more appropriate than those who interacted with the QTrobot (QT, Mdn = 2.64). Additionally, our findings show that the perceived personality of the robot is in line with the target personality (identified in the design phase of the study as reported in Section 4.1.4), and it is impacted by the robot form. The perceived personality traits for the QTrobot (QT) are Openness (M=2.71, SD=1.32), Conscientiousness (M=2.93, SD=1.20), Extroversion (M=2.64, SD=1.21), Agreeableness (M=3, SD=1.46), and Neuroticism (M=1.50, SD=1.28). These differ from the ones perceived for the Misty robot (M) — Openness (M=3.42, SD=0.79), Conscientiousness (M=3.58, SD=1.08), Extroversion (M=2.92, SD=0.99), Agreeableness (M= 4.42, SD=0.51), and Neuroticism (M=1.42, SD=0.99).

The framework method's analysis supports the aforementioned findings. In QT, coachees viewed the robot's personality more negatively, with many describing it as not having a personality (P01, P03, P04, P05, P06, P07, P08, P09). Some did describe it neutrally or positively, as "friendly" (P02, P05, P06), "professional" (P02), and "non-judgemental" (P10, P11). In M, coachees had a more detailed view of the robot's personality, and coachees perceived the robot's personality to be much more in line with the design traits. The robot was described as "empathetic" (P15), "caring" (P22, P23), "more introvert than extrovert" (P15), "warm" (P15, P16, P26), "calm" or "calming" (P16, P18, P22, P24, P26), "fair" (P17), "motherly" (P19, P21), "understanding" (P21), and "relaxed" (P26). P16 remarked that the personality was intrinsically related to the robot's voice. P18 described the robot as "slightly unrelenting at times" with regards to leading the session with questions. P23 made a similar observation, calling the robot "diligent" and "well-prepared", "clearly having a plan for the sessions", and "wanting to stick with what it had planned". In terms of negative and neutral assessments, P20 and P17 remarked that they did not think the robot had a personality. P20 also described the robot's personality as "robotic", and P24 said its responding was "not very attentive".

As reported in Section 4.1.4, we designed the robot personality to be reflected by its behaviour, according to recommendations from literature and professional coaches. Both robotic coaches were designed to have the same behaviour. We see that the overall perception of OCEAN personality traits was more aligned with the original design in M, as shown in Figure 3. For Extroversion, Agreeableness and Neuroticism, both QT and M fell within the target values. For Openness and Conscientiousness, M was closer to the target values. Coachees also perceived the behaviours in M to be more in line with the targets, as more behaviours in M were closer to the target values than in QT. The differences between conditions indicate that while the behaviours were successful in expressing the personality traits as designed, the form of the robot also influenced this perception. The most "successfully" designed personality trait (and its behavioural indicators) was Extroversion, where both robots fell within all target values (for gestures, facial expressions and word choices). Openness was the least successful, with neither condition achieving the target in trait or behaviours. There was no clear success or failure across designed behaviour dimensions (gestures, facial expressions, voice and word choices). The qualitative data supports that coachees preferred the behaviours in M. The robot's voice was the same in both conditions, but was viewed more positively in M. In QT, coachees viewed the robot's voice more neutrally, describing it as "natural [...] but not really natural" (P09), "a bit robotic" (P12), and "neutral" (P14). In M, coachees viewed the voice much more positively, citing it as "warm" (P16), "calming" (P16, P22, P24), and "comforting" (P25). The difference in perceived movements was not clear between QT and M. In QT, coachees noted the robot "moving or nodding at inappropriate times" (P01, P06, P07, P11, P13), which could "interrupt them" (P06) or "be intrusive" (P07). In M, movements were perceived both positively and negatively. P15 thought the nodding "reinforced active listening", although they noted that the movements were initially "off-putting". Similarly, P18 said that the "head tilting was good as a form of acknowledgement". The robot's facial expressions, especially its eye movements, were viewed mostly positively in both conditions.

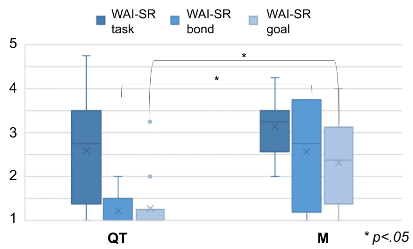


Figure 4: WAI-SR sub-scales (task, bond, and goal) in the QTrobot (QT) and Misty robot (M) conditions

Coachee-Coach Alliance (RQ3). To analyse the data gathered from the questionnaires, we conducted Wilcoxon signed-rank tests to compare the well-being measures (RPWS, satisfaction scale, mood and readiness to change scale) pre- and post-study and the WAI-SR scale to measure the coachee-coach alliance. As our study population was screened to exclude higher levels of anxiety and depression, we did not find statistically significant differences between the well-being measures before and after the study. Thus, we performed Mann–Whitney U tests to compare the coachee-coach working alliance between the two robotic form conditions (QT and M) to address our research question (RQ3). We found that the coachees develop significantly different working alliance with the two robots over the four weeks for the WAI-SR sub-scales *bond* (U = 31.000, $z = -2.700$, corrected p^*<0.05) and *goal* (U = 29.000, $z = -2.803$, corrected p^*<0.05), while no significant difference was found for the WAI-SR *task* sub-scale. The WAI-SR sub-scales *bond* and *goal* were significantly higher for participants who interacted with the Misty robot (M, Mdn_{bond} = 2.56, Mdn_{goal} = 2.31) than the scores collected from coachees who interacted with the QTrobot (QT, Mdn_{bond} = 1.21, Mdn_{goal} = 1.27), as depicted in Figure 4. The WAI-SR *task* sub-scale scored for QT as Mdn = 2.75, while for M as Mdn = 3.25. We observed similar results from the additional questionnaire data about the coachees' connection with the robot (see also Table 2 of the Supplementary Material). On average the participants found the robot helpful (QT: M=2.92, SD= 1.38 and M: M = 3.66, SD=0.65) and useful (QT: M=3, SD=1.41 and M: M = 3.5, SD=1), and for all the items, participants in M scored the robot more positively than in QT in terms of feeling connected to the robot (QT: M=1.78, SD=0.97 and M: M = 2.91, SD=0.99), and comfort in talking with the robot (QT: M=2.21, SD= 1.18 and M: M = 3.41, SD=0.90). Participants also perceived both robots more as a "stranger" (QT: M=2.07, SD=1.26 and M: M = 3.08, SD=0.99) than a "friend" (QT: M=1.42, SD=0.64 and M: M = 2, SD=0.95), and participants in M felt that talking with the robot was similar to talking with a well-being coach (QT: M=2.21, SD=1.36 and M: M = 2.58, SD=1.16). Again, interview results supported the quantitative findings. Participants in M described more instances of feeling connected with the robot, e.g., P3 (M) said "'attached' is maybe the wrong word but [...] getting used to the voice and the little gestures", P26 (M) noted that "there's a little emotional connection going", and P15 stating "I do feel an affinity with her". In QT, only P14 (QT) mentioned connection: "[the correct timing] built up my connection with the robot, and then it went and destroyed all its hard work by [talking] in the wrong places". In M, coachees used diminutive language more often when describing the robot.

6 DISCUSSION & CONCLUSIONS

Our results show that, given the two robotic platforms investigated, form impacts how the coachees perceived the robotic well-being coach. We found that overall the Misty robot was perceived more positively than the QTrobot (**RQ1**). These results could be explained by the *form function attribution bias* [22]. Since Misty is more toy-like, smaller, and less humanoid than QTrobot (which has a human-like upper and lower body shape), people may expect less when interacting with the robot—i.e., the Misty's form better matched the skills and behaviours it exhibited while administering the well-being exercises. Coachees more often described getting used to the Misty robot and its idiosyncrasies, while coachees described becoming disappointed with the QT's level of function. This indicates that the robotic coach's form was a major contributor to people's expectations and perceptions when delivering the well-being exercises. Interestingly, our results showed that people also perceived the robotic coach's behaviours (i.e., voice and gestures) and its personality to be different across the two robot platforms investigated. The behaviours and the displayed personality were designed to be exactly the same for both robots. Coachees perceived Misty's voice, gestures, and personality more positively, while coachees were more critical of the QTrobot and many noted that it did not have a personality (**RQ2**). This indicates that coachees were more lenient when judging Misty as a mental well-being robotic coach, possibly due to its smaller, more "toy-like" form, which does not have as many humanoid features as QTrobot. Another explanation could be that Misty's form (i.e., less humanoid) better matched the designed behaviours and personality. These results suggest that participants may perceive a robotic coach's behaviour differently due to a change in the robotic form. Researchers should be aware of interrelations in the perception of a robotic well-being coach - i.e., the robot's appearance or selected voice could influence the perception of its social behaviours. Our results show that coachees developed a stronger relationship with Misty robot than with the QTrobot delivering well-being exercises, confirming that the robot form has an impact on how coachees perceived the robotic coach, even in terms of coachee-coach alliance (**RQ3**). However, our results also show that coachee mental well-being did not improve throughout the four-week study period. This could be due to two main reasons. First, we screened the coachees - i.e., no significant difference was found before and after the study because their level of mental well-being was already high, and did not require further improvement. Second, the PP practice was quite short (4 weeks). This amount of time may not be sufficient to observe a significant improvement in mental well-being. Previous studies [39] on the use of mental well-being apps like Calm and Headspace showed that healthy participants had an improvement in their well-being after 8 weeks of continuous practice. Well-being practices can have benefits that differ between individuals and could take longer to be effective according to each individual's needs. Additionally, as in mindfulness practice, mental well-being practices are meant for everybody and can also be helpful for people who do not experience mental health problems [23]. Given the fact that robotic coaches are not meant to substitute professionals [5] and could be used by healthy populations, we argue that the efficacy of robotic mental well-being coaches with healthy populations can be measured—alongside pre- and post-assessments of coachees' mental well-being—by the capabilities of the robot to build a connection with coachees. Our argument is supported by the literature on coaching by humans [13, 48]. For example, Qina'au and Masuda [48] showed that patients who have strong rapport with their coach are better able to manage stress.

Deploying Robotic Well-being Coaches In the Workplace. From the qualitative data gathered (interviews and focus groups), we observed that the robotic coaches can be beneficial in the workplace acting as a *strong visual reminder* for doing the exercises, as noted by several coachees, and as stated in a previous study that investigated user requirements for robotic well-being coaches [3]. P10 (QT) stated that walking past the robot in the office would probably help them do more exercises. P17 (M) also said "a robot [could] remind you to keep doing [the exercises] yourself". P12 (QT) said that the exercises they learned during the study would be "even more memorable by virtue of the fact that the first time I did them was with this robot". Despite these promising findings, there are many open challenges that need to be addressed. Firstly, people's *unrealistic expectations* may be one of the barriers to adoption of robotic coaches in workplace settings. Focus group data provided to us by the company shows that coachees' expectations of the robot capabilities do not match reality, possibly distorted by how robots are portrayed in the media, reinforced by sales videos. Coachees stated that they had expected more from these robots "because of demos from cutting edge teams", and "Alexa and Google Assistant were what drove their expectations". It also emerged that the coachees expected the robotic coach to adapt and personalise to what they had said – i.e., "to change responses based on what the human says" and "to have more personalisation (between people and in time), e.g., referencing across the sessions". Coachees had reported to have very little prior experience with robots (average=1.3 on a 5-point Likert scale) before joining our study. This might be one of the main reasons their expectations did not match the actual skills and capabilities of the robotic coach(es) delivering well-being exercises. The *expectation-reality mismatch* is known to create priming before interacting with robots [35]. Another barrier was related to the feelings of embarrassment for using the robotic coach. Coachees in both groups noted that they might be *embarrassed to use the robot* in the future, if their colleagues saw them using it. P10 (QT) noted that using the robot would "need to be seen as a normal thing to do", and that they would prefer to use it in a closed room. Some coachees (P10, P14 [QT], P26 [M]) described "being seen going to the robot room" outside of the scope of this study as potentially having negative social consequences. Embarrassment has previously been shown to affect the choice to undertake counselling [24, 30]. Therefore, future research should examine how the social framing of the robotic coach (e.g., keeping it private vs. public) will impact coachees' motivations for using it in the workplace.

7 SUMMARY & FUTURE WORK

In this paper, we presented the *first study* that investigated the use of two different forms of robotic well-being coaches in the workplace. We conducted a longitudinal between-subjects study that involved 26 employees (coachees) who interacted with either a QT or Misty robot with a coach personality that led well-being practice sessions. We then investigated and provided results on: (1) the coachees' perceptions of two robotic coaches' *forms*; (2) the coachees' perceptions of the robotic coaches' personalities; and (3) the perceptions of the coachee-coach alliance after 4 weeks of well-being practice.

Our work has several limitations. First, most of the literature conducted questionnaire-based studies where they evaluated robot form via static images. In this study, we refer to the term *robot form* as an overarching concept that includes all aspects of a robotic platform (e.g., size, degrees of freedom in movement, motor noise, etc.). These platform characteristics could have confounded the study results. Second, we designed a coach personality for the robots grounding our design in the literature and suggestions from two well-being professionals. We were unable to find and use any specific and purposeful tool to generate a robot personality. However, we found the well-being professionals' suggestions were sufficient, as our results showed (i.e., the designed personality matched the perceived personality of the robotic coach). Future work should investigate other methods and tools for designing robotic coach personalities. Third, our study design included the creation of a specific questionnaire—which has not been validated—to measure the perceived robot personality, as no standardised test to measure robot personality is currently available within the HRI community. Future work could design and validate new standardised measures and tests for assessing robot personalities. Our sample size could be seen as small, however, this work was a long-term in-the-wild mixed methods study, and the data acquired and analysed (1040 minutes of recordings and 650 minutes of interviews) is not small, and is in line with past works in HRI [47]. Finally, the interactions were simplistic and pre-scripted, and not adaptive. This choice was deliberate, and was based on the human well-being coaches' advice to adhere to the well-being practice (human coaches themselves follow a specific structure and script).

Despite these limitations, it is our genuine hope that the findings and insights from this study will contribute to developing effective, autonomous, and engaging well-being coaches in the workplace. Our future work will certainly focus on these aspects, including the development of more sophisticated and adaptive robotic mental well-being coaches.

ACKNOWLEDGMENTS

We thank Cambridge Consultants Inc. and their employees for participating in this study. **Funding:** M. Spitale and H. Gunes are supported by the EPSRC/UKRI under grant ref. EP/R030782/1 (ARoEQ) and EP/R511675/1. M. Axelsson is funded by the Osk. Huttunen foundation and the EPSRC under grant EP/T517847/1. **Data access:** Raw data related to this publication cannot be openly released due to anonymity and privacy issues.

REFERENCES

[1] Nida Abbasi, Micol Spitale, Joanna Anderson, Tamsin Ford, Peter Jones, and Hatice Gunes. 2022. Can Robots Help in the Evaluation of Mental Wellbeing in Children? An Empirical Study. In *2022 31st IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*. IEEE.

[2] Sean Andrist, Bilge Mutlu, and Adriana Tapus. 2015. Look like me: matching robot personality via gaze to increase motivation. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 3603–3612.

[3] Minja Axelsson, Indu P Bodala, and Hatice Gunes. 2021. Participatory Design of a Robotic Mental Well-being Coach. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*. IEEE, 1081–1088.

[4] Minja Axelsson, Nikhil Churamani, Atahan Caldir, and Hatice Gunes. 2022. Participant Perceptions of a Robotic Coach Conducting Positive Psychology Exercises: A Systematic Analysis. *arXiv preprint arXiv:2209.03827* (2022).

[5] Minja Axelsson, Micol Spitale, and Hatice Gunes. 2022. Robots as Mental Well-being Coaches: Design and Ethical Recommendations. *arXiv preprint arXiv:2208.14874* (2022).

[6] Jenay M Beer and et al. 2014. Toward a framework for levels of robot autonomy in human-robot interaction. *Journal of human-robot interaction* 3, 2 (2014), 74.

[7] Elisabetta Bevacqua, Etienne De Sevin, Sylwia Julia Hyniewska, and Catherine Pelachaud. 2012. A listener model: introducing personality traits. *Journal on Multimodal User Interfaces* 6, 1 (2012), 27–38.

[8] Indu P Bodala, Nikhil Churamani, and Hatice Gunes. 2021. Teleoperated robot coaching for mindfulness training: A longitudinal study. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*. IEEE, 939–944.

[9] Elizabeth Broadbent, Vinayak Kumar, Xingyan Li, John Sollers 3rd, Rebecca Q Stafford, Bruce A MacDonald, and Daniel M Wegner. 2013. Robots with display screens: a robot with a more humanlike face display is perceived to have more mind and a better personality. *PloS one* 8, 8 (2013), e72589.

[10] Kate B Carey, Stephen A Maisto, Michael P Carey, and Daniel M Purnine. 2001. Measuring readiness-to-change substance misuse among psychiatric outpatients: I. Reliability and validity of self-report measures. *Journal of Studies on Alcohol* 62, 1 (2001), 79–88.

[11] 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 2017 ACM/IEEE International Conference on human-robot interaction*. 254–262.

[12] Nikhil Churamani, Minja Axelsson, Atahan Caldir, 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.

[13] Erik De Haan and Judie Gannon. 2017. The coaching relationship. *The SAGE handbook of coaching* (2017), 195–217.

[14] David DeVault, Ron Artstein, Grace Benn, Teresa Dey, Ed Fast, Alesia Gainer, Kallirroi Georgila, Jon Gratch, Arno Hartholt, Margaux Lhommet, et al. 2014. SimSensei Kiosk: A virtual human interviewer for healthcare decision support. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*. 1061–1068.

[15] Connor Esterwood and Lionel P Robert. 2020. Personality in healthcare human robot interaction (h-hri) a literature review and brief critique. In *Proceedings of the 8th International Conference on Human-Agent Interaction*. 87–95.

[16] Juan Fasola and Maja J Matarić. 2013. A socially assistive robot exercise coach for the elderly. *Journal of Human-Robot Interaction* 2, 2 (2013), 3–32.

[17] Samuel T Gladding and Promila Batra. 2007. *Counseling: A comprehensive profession*. Pearson Education India.

[18] Lewis R Goldberg. 1993. The structure of phenotypic personality traits. *American psychologist* 48, 1 (1993), 26.

[19] L Suzzy Green, Lindsay G Oades, and Anthony M 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.

[20] Tammy Gregersen, Peter D MacIntyre, Kate Hein Finegan, Kyle Talbot, and Shelby Claman. 2014. Examining emotional intelligence within the context of positive psychology interventions. (2014).

[21] Kristina Gyllensten and Stephen Palmer. 2007. The coaching relationship: An interpretative phenomenological analysis. *International Coaching Psychology Review* 2, 2 (2007), 168–177.

[22] Kerstin S Haring, Katsumi Watanabe, Mari Velonaki, Chad C Tossell, and Victor Finomore. 2018. FFAB—The form function attribution bias in human–robot interaction. *IEEE Transactions on Cognitive and Developmental Systems* 10, 4 (2018), 843–851.

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

[24] Miriam Heyman, Jeff Dill, and Robert Douglas. 2018. *The Ruderman white paper on mental health and suicide of first responders*. Vol. 41. Ruderman Family Foundation Boston, MA.

[25] 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.

[26] Michiel Joosse, Manja Lohse, Jorge Gallego Perez, and Vanessa Evers. 2013. What you do is who you are: The role of task context in perceived social robot personality. In *2013 IEEE International Conference on Robotics and Automation*. IEEE, 2134–2139.

[27] Malte Jung and Pamela Hinds. 2018. Robots in the wild: A time for more robust theories of human-robot interaction. , 5 pages.

[28] Casey Kennington, Daniele Moro, Lucas Marchand, Jake Carns, and David McNeill. 2020. rrSDS: Towards a robot-ready spoken dialogue system. In *Proceedings of the 21th annual meeting of the special interest group on discourse and dialogue*. 132–135.

[29] Thomas Kiderle, Hannes Ritschel, Kathrin Janowski, Silvan Mertes, Florian Lingenfelser, and Elisabeth André. 2021. Socially-aware personality adaptation. In *2021 9th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*. IEEE, 1–8.

[30] Robert King, Matthew Bambling, Chris Lloyd, Rio Gomurra, Stacy Smith, Wendy Reid, and Karly Wegner. 2006. Online counselling: The motives and experiences of young people who choose the Internet instead of face to face or telephone counselling. *Counselling and Psychotherapy Research* 6, 3 (2006), 169–174.

[31] Christian U Krägeloh, Jaishankar Bharatharaj, Senthil Kumar Sasthan Kutty, Praveen Regunathan Nirmala, and Loulin Huang. 2019. Questionnaires to measure acceptability of social robots: a critical review. *Robotics* 8, 4 (2019), 88.

[32] Kwan Min Lee, Wei Peng, Seung-A Jin, and Chang Yan. 2006. Can robots manifest personality?: An empirical test of personality recognition, social responses, and social presence in human–robot interaction. *Journal of communication* 56, 4 (2006), 754–772.

[33] X Alvin Li, Maria Florendo, E Luke Miller, Hiroshi Ishiguro, and P Ayse Saygin. 2015. Robot form and motion influences social attention. In *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 43–50.

[34] Bernd Löwe, Kurt Kroenke, Wolfgang Herzog, and Kerstin Gräfe. 2004. Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). *Journal of affective disorders* 81, 1 (2004), 61–66.

[35] Bertram F Malle. 2020. Trust And The Discrepancy Between Expectations And Actual Capabilities. *Human-robot interaction: Control, analysis, and design* (2020), 1.

[36] Thomas Munder, Fabian Wilmers, Rainer Leonhart, Hans Wolfgang Linster, and Jürgen Barth. 2010. Working Alliance Inventory-Short Revised (WAI-SR): psychometric properties in outpatients and inpatients. *Clinical Psychology & Psychotherapy: An International Journal of Theory & Practice* 17, 3 (2010), 231–239.

[37] Tatsuya Nomura, Tomohiro Suzuki, Takayuki Kanda, and Kensuke Kato. 2006. Measurement of negative attitudes toward robots. *Interaction Studies* 7, 3 (2006), 437–454.

[38] 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.

[39] Alison O'Daffer, Susannah F Colt, Akash R Wasil, Nancy Lau, et al. 2022. Efficacy and Conflicts of Interest in Randomized Controlled Trials Evaluating Headspace and Calm Apps: Systematic Review. *JMIR Mental Health* 9, 9 (2022), e40924.

[40] World Health Organization. 2022. *Mental health at work*. https://www.who.int/news-room/fact-sheets/detail/mental-health-at-work

[41] Anastasia K Ostrowski, Cynthia Breazeal, and Hae Won Park. 2022. Mixed-Method Long-Term Robot Usage: Older Adults' Lived Experience of Social Robots. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*. 33–42.

[42] Corrado Pacelli, Tharushi Kinkini De Silva Pallimulla Hewa Geeganage, Micol Spitale, Eleonora Beccaluva, and Franca Garzotto. 2022. " How Would You Communicate With a Robot?" People with Neourodevelopmental Disorder's Perspective. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*. 968–972.

[43] Maike Paetzel, Giulia Perugia, and Ginevra Castellano. 2020. The persistence of first impressions: The effect of repeated interactions on the perception of a social robot. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. 73–82.

[44] Maike Paetzel-Prüsmann, Giulia Perugia, and Ginevra Castellano. 2021. The influence of robot personality on the development of uncanny feelings. *Computers in Human Behavior* 120 (2021), 106756.

[45] Michael Quinn Patton. 1999. Enhancing the quality and credibility of qualitative analysis. *Health services research* 34, 5 Pt 2 (1999), 1189.

[46] William Pavot and Ed Diener. 2008. The satisfaction with life scale and the emerging construct of life satisfaction. *The journal of positive psychology* 3, 2 (2008), 137–152.

[47] Giulia Perugia, Alessandra Rossi, and Silvia Rossi. 2021. Gender revealed: Evaluating the genderedness of furhat's predefined faces. In *International Conference on Social Robotics*. Springer, 36–47.

[48] Joanne Qina'au and Akihiko Masuda. 2020. Cultural considerations in the context of establishing rapport: A contextual behavioral view on common factors. In *Handbook of Cultural Factors in Behavioral Health*. Springer, 75–92.

[49] Jane Ritchie, Liz Spencer, Alan Bryman, and Robert G Burgess. 1994. Analysing qualitative data.

[50] Lionel Robert. 2018. Personality in the human robot interaction literature: A review and brief critique. In *Robert, LP (2018). Personality in the Human Robot Interaction Literature: A Review and Brief Critique, Proceedings of the 24th Americas Conference on Information Systems, Aug*. 16–18.

[51] Kristin E Schaefer, Tracy L Sanders, Ryan E Yordon, Deborah R Billings, and Peter A Hancock. 2012. Classification of robot form: Factors predicting perceived trustworthiness. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 56. SAGE Publications Sage CA: Los Angeles, CA, 1548–1552.

[52] Martin EP Seligman. 2007. Coaching and positive psychology. *Australian Psychologist* 42, 4 (2007), 266–267.

[53] Leah B Shapira and Myriam Mongrain. 2010. The benefits of self-compassion and optimism exercises for individuals vulnerable to depression. *The Journal of Positive Psychology* 5, 5 (2010), 377–389.

[54] Jennifer L Smith and Agnieszka A Hanni. 2019. Effects of a savoring intervention on resilience and well-being of older adults. *Journal of Applied Gerontology* 38, 1 (2019), 137–152.

[55] Sinan Sonlu, Uğur Güdükbay, and Funda Durupinar. 2021. A conversational agent framework with multi-modal personality expression. *ACM Transactions on Graphics (TOG)* 40, 1 (2021), 1–16.

[56] Micol Spitale, Chris Birmingham, R Michael Swan, and Maja J Matarić. 2021. Composing harmoni: An open-source tool for human and robot modular open interaction. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 3322–3329.

[57] Micol Spitale, Sarah Okamoto, Mahima Gupta, Hao Xi, and Maja J Matarić. 2022. Socially Assistive Robots as Storytellers That Elicit Empathy. *ACM Transactions on Human-Robot Interaction* (2022).

[58] Robert L Spitzer, Kurt Kroenke, Janet BW Williams, and Bernd Löwe. 2006. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of internal medicine* 166, 10 (2006), 1092–1097.

[59] Adriana Tapus, Cristian Ţăpuş, and Maja J Matarić. 2008. User—robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy. *Intelligent Service Robotics* 1, 2 (2008), 169–183.

[60] Adriana Tapus, Cristian Tapus, and Maja J Mataric. 2009. The use of socially assistive robots in the design of intelligent cognitive therapies for people with dementia. In *2009 IEEE international conference on rehabilitation robotics*. IEEE, 924–929.

[61] Ewa Topolewska, Ewa Skimina, WŁODZIMIERZ Strus, Jan Cieciuch, and Tomasz Rowiński. 2014. The short IPIP-BFM-20 questionnaire for measuring the Big Five. *Roczniki Psychologiczne* 17, 2 (2014), 385–402.

[62] Dirk Van Dierendonck. 2004. The construct validity of Ryff's Scales of Psychological Well-being and its extension with spiritual well-being. *Personality and individual differences* 36, 3 (2004), 629–643.

[63] Marieke van Otterdijk, Heqiu Song, Konstantinos Tsiakas, Ilka van Zeijl, and Emilia Barakova. 2022. Nonverbal Cues Expressing Robot Personality-A Movement Analysts Perspective.

[64] Sue Wheeler. 2000. What makes a good counsellor? An analysis of ways in which counsellor trainers construe good and bad counselling trainees. *Counselling Psychology Quarterly* 13, 1 (2000), 65–83.

[65] Steve Whittaker, Yvonne Rogers, Elena Petrovskaya, and Hongbin Zhuang. 2021. Designing Personas for expressive robots: personality in the new breed of moving, speaking, and colorful social home robots. *ACM Transactions on Human-Robot Interaction (THRI)* 10, 1 (2021), 1–25.

Supplementary Materials: Robotic Mental Well-being Coaches for the Workplace: An In-the-Wild Study on Form

Micol Spitale[a*], Minja Axelsson[a], Hatice Gunes[a]
[a]Department of Computer Science and Technology, University of Cambridge, Cambridge, United Kingdom
[*]Corresponding author: ms2871@cam.ac.uk

December 2022

1 Post-study Interview Questions

In the final session, each coachee was also interviewed using the following additional questions:

1. How would you describe the robot's personality?
2. What do you think was good about the robot's appearance, voice and movements?
3. Do you think the robot understood what you said?
4. Do you think the robot understood how you felt?
5. Do you think the robot adapted to what you said and did?

2 Quantitative and Qualitative Results: Further Details

Item: The robot..	**QT** ($M\pm SD$)	**M** ($M\pm SD$)
1) .. was listening carefully	2.14±1.23	**3.08**±0.90
2) .. behaviour was natural	2.21±1.12	**2.92**±1.08
3) .. gestured appropriately	2.86±1.56	**3.00**±0.85
4) .. acknowledged me when I spoke	2.64±1.28	**3.25**±1.06
5) .. cared about what I said	1.86±1.03	**1.92**±0.67
6) .. did not care about what I said*	**3.86**±1.35	3.67±0.89
7) .. was not listening to me*	**2.71**±1.44	2.25±1.06
8) .. understood what I was saying	1.71±0.83	**2.08**±0.79
9) .. understood how I felt	1.50±0.85	**1.75**±0.87
10) .. adapted to what I said and did	1.79±1.12	**2.58**±0.90

Table 1: Items of the specific questionnaire about the coachee's perception of robot behavior. '*' indicates an inverse item (lower score preferred); bold highlights the highest score.

Item: Talking with the robot..	**QT** $M \pm SD$	**M** $M \pm SD$
1) .. I felt a close connection to the robot	1.78±0.97	**2.91**±0.99
2) .. was similar to talking with a friend	1.42±0.64	**2**±0.95
3) .. was similar to talking with a stranger	2.07±1.26	**3.08**± 0.99
4) .. felt like talking with a coach	2.21±1.36	**2.58**± 1.16
5) .. was comforting	2.21±1.18	**3.41**±0.90
6) .. I felt the robot was helpful	2.92±1.38	**3.66**±0.65
7) .. I felt the robot was useful	3±1.41	**3.5**± 1

Table 2: Items of the specific questionnaire about the coachee's connection with the robot. Bold highlights the highest score.

	O (high) Target: 4-5 **QT: 2.7, M: 3.4**	**C (high) Target: 4-5** **QT: 2.9, M: 3.4**	**E (med) Target: 2.5-3.5** **QT: 2.6, M: 2.9**	**A (med-high) Target: 3-4.5** **QT: 3 , M: 4.4**	**N (low) Target: 1-2** **QT: 1.5, M: 1.4**
Gestures		Slow and consistent movement [6] Target slowness: 4-5, QT: 3.4, M: 3 Target consistency: 4-5, QT: 3.4, M: 2.8	Medium frequency [2, 1] Target frequency: 2.5-3.5, QT: 2.5, M: 3.25		
Facial expressions	Pleasant and happy [4] Target happiness: 4-5, QT: 3.5, M: 3.1	Medium frequency [3] Target frequency: 2.5-3.5, QT: 2.9, M: 3.1	Medium frequency [2] Target frequency: 2.5-3.5, QT: 2.9, M: 3.1		
Voice		Slow pace [coach], high volume [5] low pitch [5], low pitch variation [6] Target slowness: 4-5, QT: 3.6, M: 3.9 Target volume: 4-5, QT: 2.8, M: 3.7 Target pitch: 1-2, QT: 2.5, M: 2.4 Target pitch variation: 1-2, QT: 1.9, M: 2.2			
Word choices	Affirmation, more words [4] Target affirmation: 4-5, QT: 2.3, M: 2.8	Decisiveness [4], waiting after user speaks [6] Target waiting: 4-5, QT: 2.1, M: 3.6	Medium level of polite language [4] Target politeness: 4-5, QT: 4.1, M: 4.6	Medium to high level of affirmation [4] Target affirmation: 4-5, QT: 2.3, M: 2.8	Highly decisive [4] Target decisiveness: 4-5, QT: 3.8, M: 4.1

Table 3: Robot personality design choices for Openness (O), Conscientiousness (C), Extroversion (E), Agreeableness (A), and Neuroticism (N). Averaged values that fell within the target lower-upper bounds are in green. The value of the condition that is closer to the aimed target value is highlighted in cyan.

References

[1] Elisabetta Bevacqua et al. "A listener model: introducing personality traits". In: *Journal on Multimodal User Interfaces* 6.1 (2012), pp. 27–38.

[2] David DeVault et al. "SimSensei Kiosk: A virtual human interviewer for healthcare decision support". In: *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems.* 2014, pp. 1061–1068.

[3] Margaret McRorie et al. "Evaluation of four designed virtual agent personalities". In: *IEEE Transactions on Affective Computing* 3.3 (2011), pp. 311–322.

[4] Sinan Sonlu, Uğur Güdükbay, and Funda Durupinar. "A conversational agent framework with multimodal personality expression". In: *ACM Transactions on Graphics (TOG)* 40.1 (2021), pp. 1–16.

[5] Jürgen Trouvain et al. "Modelling personality features by changing prosody in synthetic speech". In: (2006).

[6] Steve Whittaker et al. "Designing Personas for expressive robots: personality in the new breed of moving, speaking, and colorful social home robots". In: *ACM Transactions on Human-Robot Interaction (THRI)* 10.1 (2021), pp. 1–25.