Computers in Human Behavior 99 (2019) 76–85

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Full length article

Use of a non-human robot audience to induce stress reactivity in human participants

Julie M. Turner-Cobb[a,d,*], Mashal Asif[a], James E. Turner[b], Chris Bevan[c], Danae Stanton Fraser[c]

[a] The STress, Endocrine and Lifecourse LAboRatory (STELLAR), Department of Psychology, University of Bath, UK
[b] Department for Health, University of Bath, UK
[c] The CREATE Laboratory, Department of Psychology, University of Bath, UK
[d] Bournemouth University, Department of Psychology, Faculty of Science and Technology, UK

ARTICLE INFO

Keywords:
Stress testing
Salivary cortisol
Heart rate
Social evaluative threat
Robot audience
Coping

ABSTRACT

This study examined whether a non-human robot audience can elicit a stress response in human participants. A 90-min experimental laboratory session based on the Trier Social Stress Test (TSST) using a pre-recorded robot audience, was presented as a live on-screen simulation. Nineteen participants (female = 16) aged 21–57 years (M = 29.74) underwent a 10-min mock interview and mathematics task in front of the robot audience. Salivary cortisol was assessed at 10-min before and immediately prior to the start of the stress test, and +10-, +30- and +40-min after the start of the test. Heart rate was assessed 20 min before, at 5 min into and 40-min after the test. Perceived stress and trait coping responses were provided at entry and participants were interviewed post task about their subjective experience. Significant increases in salivary cortisol and heart rate were observed over time with no significant interactions by participant subjective report. Coping responses including active coping and planning showed significant relationships with cortisol and heart rate reactivity and recovery. Until now, a non-human robot audience has not been used in a social stress testing paradigm. This methodology offers an innovative application with potential for further in-depth evaluation of stress reactivity and adaptation.

1. Introduction

It is uncontested that the Trier Social Stress Test (TSST) (Kirschbaum, Pirke, & Hellhammer, 1993) offers a well validated and established acute stress experimental paradigm. A recent review cites it as the "gold standard in human experimental stress research" having been applied across a range of settings, different populations and age groups (Allen et al., 2017, p. 115). Various adaptations of this test include those that have changed the setting or other relevant parameters including the TSST-G designed for group testing (von Dawans, Kirschbaum, & Heinrichs, 2011), the inclusion of a placebo (Het, Rohleder, Schoofs, Kirschbaum, & Wolf, 2009) or friendly version (Het et al., 2009) as a control comparison and a growing number of virtual reality (VR) versions (Jonsson et al., 2010; Kotlyar et al., 2008; Shiban et al., 2016; Wallergard, Jonsson, Osterberg, Johansson, & Karlson, 2011; Kothgassner et al., 2016). Throughout these adaptations the audience has typically involved a panel or audience of two or three adults, whether in-person (termed "in vivo" by some authors e.g. Shiban et al. (2016) as in the original TSST (Kirschbaum et al., 1993) or delivered via VR. Some exceptions have also experimented with video delivery using a pre-recorded audience such as in the Leiden public speaking task (Westenberg et al., 2009).

Supplementary video related to this article can be found at https://doi.org/10.1016/j.chb.2019.05.019

The presence or absence of an in-person audience and the degree of human authenticity differentiates within social stress testing paradigms, highlighting the importance of a key stressor characteristic, that of social evaluative threat (SET). Described by Dickerson and Kemeny (2004) in the context of their social self-preservation theory, the self-preservation system is attuned to detect esteem or status threats to the social self, initiating psychological and physiological responses to protect against these experiences. VR versions of the TSST and comparable social stress tests are particularly important in evaluating the role and nuances of SET. They offer significant ecological validity in identifying the components of, and potential to ameliorate, stress reactivity responses to social component stress situations. In VR or pre-recorded audience simulations, sympathetic and hypothalamic-pituitary-adrenal (HPA) axis reactivity in response to the stressor is observed but generally to a lesser degree than seen with the in-person environment, particularly with regard to cortisol reactivity (e.g. (Shiban et al., 2016;

* Corresponding author. Department of Psychology, Faculty of Science & Technology, Talbot Campus, Bournemouth University, Poole, Dorset, BH12 5BB, UK.
E-mail address: jturnercobb@bournemouth.ac.uk (J.M. Turner-Cobb).

https://doi.org/10.1016/j.chb.2019.05.019
Received 9 October 2018; Received in revised form 27 January 2019; Accepted 12 May 2019
Available online 18 May 2019

Montero-Lopez et al., 2016). Importantly, the level of reactivity appears dependent on the perceived authenticity of the VR environment with greater increases in salivary cortisol, closer to in-person responses, when participants are subjected to more immersive environments (e.g. Jonsson et al., 2010; Kothgassner et al., 2016). Attempts to attune the VR TSST salivary cortisol responses to those seen via in-person stress testing, through inclusion of a competitive element have not proved successful (Shiban et al., 2016) indicating the subtlety of in-person stress inducing parameters to activate neuroendocrine responses and the difficulty in translating these across virtual environments. Explanations for the ability of specific VR adaptations to induce comparable cortisol reactivity appear to lie in the degree to which the parameters of the stressor paradigm are meaningful and deliver sufficient socially threatening triggers to activate the social self-preservation system. In a recent virtual adaptation of the TSST paradigm, Kothgassner et al. (2016) directly compared stress reactivity between participants performing in a lecture theatre with a live in-person audience, a virtual lecture theatre with virtual audience and a control group consisting of a real lecture theatre but no audience. Regardless of whether the audience was virtual or real, both elicited similar and significant patterns of stress reactivity and recovery across cardiovascular, cortisol and self-reported state anxiety, compared to the control group (Kothgassner et al., 2016) although still less than half the cortisol increase reported by Jonsson et al. (2010). Kothgassner et al. (2016) and Jonsson et al. (2010) conclude that SET has relevance within VR settings just as within in-person, physically present or in vivo settings and discuss their findings of stress reactivity within the virtual environment in relation to support for what has been termed the media equation. This equation maintains that "*media equal real life*" and is based on evidence across a range of interpersonal behaviours, which suggests that "individuals' interactions with computers, television, and new media are *fundamentally social and natural,* just like actions in real life" (Reeves & Nass, 2002, p. 5). The theories of social self-preservation/SET and the media equation are not disparate but both stem from an evolutionary perspective. Prior to the relatively recent emergence of technology, the social human brain has been programmed throughout evolution to respond automatically to encounters with the real-world environment "making acceptance of what only seems to be real … automatic" (Reeves & Nass, 2002, p. 12). Similarly, an anthropomorphic response may be engendered in which robots or computers may be capable of inducing stress responses in humans who naturally interpret the situation as a social-evaluative threat in which preservation of self-identity, social standing or self-esteem is essential to survival.

In all the TSST adaptations, whether delivered in-person or via VR, one crucial element that has not been manipulated is the composition of the audience, with the exception of one paradigm shifting study, the Bath Experimental Stress Test for Children (BEST-C) (Cheetham & Turner-Cobb, 2016). In this study, designed to more meaningfully elicit stress in children aged 7–11 years of age, researchers sought to match participant characteristics with the audience. This pre-recorded audience was composed of two children, age-matched to participants who performed a live standard TSST public speaking and mathematics task (Cheetham & Turner-Cobb, 2016).

Furthermore, some stress reactivity studies have also compared participant subjective reports of the stress experience in panel stress testing with objectives stress measures. Inconsistent objective-subjective ratings have found in adults (e.g. Coppens et al., 2018) compared to consistent ratings in children (Cheetham & Turner-Cobb, 2016). Such a comparison is important to consider since it provides some indication of the level of conscious awareness of the corresponding physiological stress response and of the degree of threat associated with social evaluative situations.

A further test of the theories of social self-preservation and the media equation would be to manipulate the stress testing audience by replacing the in-person or VR humans with a non-human or robot audience. This maintains the audience element of the stress test, consistent with previous in-person and VR applications of TSST style testing but removes the human element retained by VR adaptations and normally associated with generating SET. A robot audience enables an assessment of whether the non-human (media) have the capacity to generate the social conditions that elicit a stress response. In other words, it directly maps social stress testing to a non-human robot audience.

Furthermore, on-going chronic stress and individual differences in psychological resources such as coping responses have been found to influence reactivity patterns in response to acute stressors as evaluated using TSST style testing paradigms (e.g. Hohne et al., 2014; Janson & Rohleder, 2017; Rohleder, 2014; Villada, Hidalgo, Almela, & Salvador, 2016). Coping strategies used under conditions of social threat, particularly when that threat may appear uncontrollable, provide indication of both positive and negative ways in which responses may be diminished or heightened. The ecological validity of acute stress paradigms such as the TSST is demonstrated in the ability of certain coping responses to serve as brief interventions to reduce the stress experience or to act as "stress inoculation tools" (Abelson et al., 2014), p.68; (Bendezu, Sarah, & Martha, 2016).

1.1. Aims and hypotheses

The aim was to conduct a proof of concept study to assess whether a non-human robot audience used within a TSST setting could induce a psychosocial stress response in human participants. Primarily, we hypothesized that the robot audience would be capable of inducing a stress response in participants as evidenced by cortisol reactivity and heart rate and that this would be indicative of SET activation. Secondly, we hypothesized that there would be a degree of incongruity between objective measures of stress responsivity and subjective verbal reports of the stress test experience. We also hypothesized that underlying perceived stress levels and trait coping responses would be associated with cortisol and heart rate responses to the robot stress testing audience.

2. Materials and methods

2.1. Participants and recruitment

A random sample of twenty participants were recruited via poster advertisements and flyers distributed on University campus, online noticeboards and advertisements on social media platforms. Participants were required to be in good health, above the age of 18 years, and in local proximity to visit the laboratory for testing; exclusion criteria included taking oral steroid medication or being registered as having special educational needs (SEN). One participant was excluded as an extreme outlier on cortisol measurements, returning the final sample for inferential analysis as nineteen participants (17 women, 3 men; M *age* = 29.74 years, SD = 11.25 years, age range: 21–57 years composing 73.7% aged 21–27 years). Three quarters of the sample were white Caucasian and the remaining were Asian; the majority (90%) had a first degree or postgraduate qualification.

2.2. Measures

2.2.1. Questionnaires: demographics, perceived stress, and coping

Demographic information including age, sex, ethnicity, and highest level of educational attainment were obtained.

Participants completed two self-report questionnaires: i) The Perceived stress scale (PSS-10), an abbreviated version of the PSS (Cohen, Kamarck, & Mermelstein, 1983) containing 10 items that measure the degree to which nonspecific situations in an individual's life are appraised as stressful; they indicated how frequently had they found their lives to be unpredictable, uncontrollable, and overloaded in

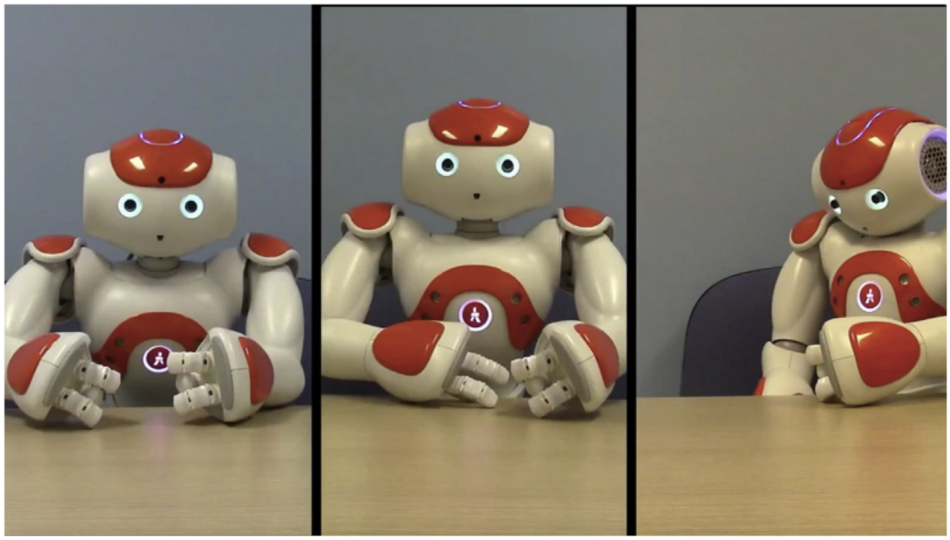

Fig. 1. Non-human robot stress testing audience.

the last month using a 5-point response scale with higher scores indicating greater perceived stress. The PSS has adequate internal and test-retest reliability and is correlated with a range of self-report and behavioral criteria and several health-related factors (Cohen et al., 1983); ii) The Brief COPE (Carver, 1997) containing 28 items, concerned with the ways in which individuals manage stress, measuring the extent and frequency of the use of fourteen coping strategies. These coping responses have been found to be reliable and predictive of potential physiological effects (Carver, Scheier, & Weintraub, 1989) and the measure has been used across a multitude of different populations and coping contexts.

2.2.2. The stress test

The stress test was based on the well-established Trier Social Stress Test (TSST) that involves a 10-min verbal presentation followed by a mental arithmetic task in front of a live audience (Kirschbaum et al., 1993). In the present study, the live audience was manipulated to consist of non-human robot images delivered via a life-sized video screen (see Fig. 1). The video was based on the mannerisms expressed in the BEST-C adaptation of the TSST (Cheetham & Turner-Cobb, 2016) and operationalized the usual techniques encompassed by a standard TSST panel such as simultaneous focused attention or inattention towards the participant (Kirschbaum et al., 1993).

2.2.2.1. The non-human robot audience. The robot used was a Softbanks Robotics 'Nao' humanoid robot (version 4). The Nao is a 58 cm tall humanoid robot with 25 degrees of freedom. In the video, the robot was positioned such that it appeared to be seated at a desk, resting on its forearms. The scene was framed to disguise the true size of the robot. To create the film, the robot was instrumented with remote controls to allow for the direct real-time puppetry of the hands, arms, eyes, waist and head. Using a video recording of live actors simulating a stress testing panel as a reference, two researchers puppeteered the robot to mimic the movements of three individual panel members as closely as possible for a 10-min period. In practice, this involved matching the head position and movements (for example, looking at the floor in disinterest, making eye contact intensely or breaking direct eye contact), leaning forward and back, and performing general fidgeting with the hands (tapping of digits). The robots were filmed against a neutral backdrop using a camcorder (resolution 1920x1080p). As the study involved the ruse that the robots were 'Skyping in' (i.e. video teleconferencing), no stage lighting was used, in order to give the appearance of a live film being transmitted over a webcam. The final video recording was constructed by compositing the performance of three separate puppetry sessions into a single film, with the robots 'appearing' to be taking part in a group Skype call. Background sound was provided by overlaying the soundtrack of the live-acted reference video (essentially ambient background noise, such as clearing of throats, coughing, muttering and paper shuffling).

2.2.2.2. The stress task. The 5-min speech task involved participants discussing their skills, strengths and relevant work experience in order to be selected for their ideal job; the 5-min mental arithmetic component comprised serial subtraction starting at 1022 (in multiples of 13). The researcher who conducted the initial pre-test assessments showed the participant into the stress testing room and introduced the experiment. They then sat far away from the participant with their back toward them and did not interact with them during the stress test except to provide minimal prompts for the tasks. Participants were informed immediately prior to entering the stress testing room that the audience would consist of robots who would appear on a large screen in front of them. They were advised that the observers were in a separate room along the corridor for the purposes of providing the best observation point to assess participant performance. There was a visible webcam mounted on top of the computer screen to give the impression of an authentic Skype call and a video camera mounted on a tripod. Participants were informed that the stress task would be recorded for the purposes of subsequent analysis.

2.2.3. Heart rate measurement

Participants were fitted with the wrist-worn Mio Link heart rate monitor on their non-dominant arm, approximately one-inch above the wrist bone at the start of the laboratory session. The Mio Link (manufactured by Mio Global mioglobal.com) is one of a number of current generation fitness devices which have proven effective in scientific testing and have been reported as showing acceptable standards particularly in healthy populations (e.g. Gilgen-Ammann et al., 2018; Sartor et al., 2018; Thiebaud et al., 2018). The device uses optical light sensor technology to detect blood flow in the capillaries and continuous heart rate is transmitted in beats per minute (BPM). Mio devices can override the contact and heart rate spike issues of chest strap monitors which work electromagnetically; this wrist-worn device was chosen as an alternative to a more intrusive chest-strap device, providing accuracy and ease of fitting. The Mio device was connected to an Android tablet via Bluetooth 4.0, which allowed streaming of heart rate data throughout the experimental session using the "Wahoo Fitness: Workout Tracker" app. Heart rate recordings were not visible to the participant during testing. Heart rate was recorded continuously throughout, at 1 s intervals, from the point at which the Mio Link was fitted on the wrist of the participant at the start of the session prior until the end of the testing session (−20 to +50 min relative to the start of the stress task). Three time points were extracted from this 70-min period, each averaged across 1-min intervals representing 'pre' (sample 1, baseline), 'peak' (sample 2, 5 min after start of the stress task) and 'post' (sample 3, 40 min after the start of the stress task) in order to observe change in heart rate at specific points over time.

2.2.4. Salivary cortisol measurement

Participants were instructed to abstain from eating, drinking and smoking for 1 h prior to their arrival at the laboratory. All data collection took place in the afternoons between the hours of 14:30 and 18:00 to minimise the impact of diurnal variability in cortisol. Five saliva samples were collected from each participant using a Salivette® device (Sarstedt): two pre-test saliva samples were collected 20 and 30 min after arriving at the laboratory (baseline and anticipatory assessments respectively). Subsequent samples were collected 10 min after completion of the stress tasks (peak response); and two further samples in the follow-up phase at 30- and 40-min post-test (recovery 1 and 2). Samples were stored at 4 °C for up to 48 h prior to being centrifuged at 2500 × *g* for 10 min to remove particulate matter. Saliva supernatant was aliquoted into micro-centrifuge tubes and subsequently frozen at −20 °C. All samples were assayed in-house at the laboratories of the Department for Health without the need for transportation. Samples were thawed in one batch and assayed for cortisol on a single day using three commercially available enzyme immunoassay kits (Salimetrics salivary cortisol ELISA kit 1–3002). Samples from each time point for a given participant were assayed on the same plate in duplicate. Briefly, 25 μL of samples, controls and standards were added to a 96-well plate followed by 25 μL of the anti-cortisol antibody conjugated to a horse radish peroxidase enzyme (1:1600 dilution in assay buffer). The plate was sealed, mixed for 5 min at 500 rpm and incubated for 55 min at room temperature. After washing the plate, 200 μL of 3,3′,5,5′-Tetramethylbenzidine (TMB) was added. The plate was sealed, mixed for 5 min at 500 rpm, and incubated for 25 min at room temperature in the dark. To stop the reaction, 50 μL of sulfuric acid was added to each well and absorbance was read on a plate reader (SPECTROstart Nano; BMG Labtech) within 5 min at 450 nm (correction at 490 nm). Cortisol concentration was calculated using 4-parameter curve fitting. Assay sensitivity was 0.193 nmol/L. Intra-assay precision was 3.89% and inter-assay precision was 12.04%.

2.2.5. Post-test manipulation check: subjective stress reports

At the end of the recovery2 period, participants were given a brief (10-min) interview about their experience of the testing, reporting how stressful they found it and about the nature of the experience. Only one other study has conducted an extended manipulation check in this way and used participant self-reports to examine objective stress reactivity by subjective grouping (Cheetham & Turner-Cobb, 2016) but the method enables a comparison of objective-subjective congruity. In the current study we were particularly interested in the subjective reports of exposure to the non-human robot audience. The interviewer probed questions about the activities they were asked to perform and the experience of having robots as the panel audience. Interviews were audio recorded and transcribed then the description of the stress experience was used to code the subjective experience into one of three participant groups: 1) did not find the test stressful; 2) found the test stressful but not the robots; 3) found the test and the robot audience stressful.

2.3. Procedure

On entry to the laboratory, participants were led into the waiting room where they were provided with study information and consented (approximately 10 min). The Mio Link heart-rate monitoring device was then attached. Next, participants were given 10 min alone to complete the questionnaires with the researcher returning at the end of this period to collect the first saliva sample. Following this initial settling in period (20 min) participants were allowed 10 min to prepare for a role play interview in which they were to be applying for their 'dream' job and provided with pencil and paper to assist with making notes. Participants were made aware that they would not be permitted to take their notes into the testing room. Participants were informed that they would need to speak for the entire 5 min and should prepare accordingly. They were told that after their 5-min speech they would engage in a mental arithmetic task in which they would be asked to count backwards. The second saliva sample was collected at the end of this preparation period, after which the researcher articulated the following speech just prior to entering the second room for stress testing:

> "I will now be taking you through to the stress lab for the interview and testing, however I do need to make you aware that the interview panel will be operating via avatars that will appear on the screen. The panel are in a separate room along the corridor and this gives them the best observation point to assess your performance".

In addition, they were told that the researcher would be present in the room for the purpose of providing instructions but unable to communicate or help in any other way. The participant was then taken to the second room and instructed to stand behind the microphone, with functioning speaker, situated in front of the projector screen. A Cannon video camera (XM2 Mini DV Camcorder) was positioned in a tripod stand on top of a table, to the side of the participant, with the recording light on to give participants the impression of being recorded. The researcher activated the robot video and sat at the computer facing away from the participant and instructed them to start their 5-min speech. If participants stopped talking before the time was up, the researcher would wait a few seconds before urging the participant to continue until instructed to stop. The researcher also informed participants when 3 min had passed, that they had 2 min remaining. If participants spent too long or were silent then the researcher would urge them to provide an answer as quickly as possible. After 5 min, the researcher stopped the participant and initiated the serial subtraction task for the duration of 5 min. Participants were then escorted to a debrief room and informed that the stress testing was finished. They then spent a further 40 min in the debrief room with the researcher entering at intervals to take three further saliva samples corresponding with +20, +40, and +50 min from the start of the stress test. After completion of saliva sampling, participants engaged in a brief (5–10 min) structured interview about their experience during the stress testing; this functioned as an elaborated manipulation check to assess the subjective stress experience of participants. Interviews were audio recorded and transcribed. On completion, participants were verbally debriefed and provided with a hard copy printed debrief sheet. Fig. 2 illustrates the protocol and testing points for outcome variables. Time spent in the laboratory was a total of 90 min. Full ethical approval for the study was given by the University of Bath, Department of Psychology Ethics Committee (reference: 16–115) and the work was conducted in accordance with the Declaration of Helsinki and the British Psychological Society's code of ethics and conduct.

2.4. Statistical analyses

Statistical analyses utilized IBM SPSS statistics version 23. Data screening scrutinizing z-scores and box-plots identified one participant as an extreme outlier based on cortisol variables. Their data were excluded from further statistical analysis, resulting in a total sample size of 19 participants with complete data. Distributions were assessed and logarithmic (log 10) or square root transformations applied where necessary (log10 for heart rate data and square root for delta peak cortisol). All analyses reported are parametric, with two-tailed comparisons using a p value of less than 0.05 as the significance criterion and Greenhouse-Geisser correction for degrees of freedom in ANOVA where

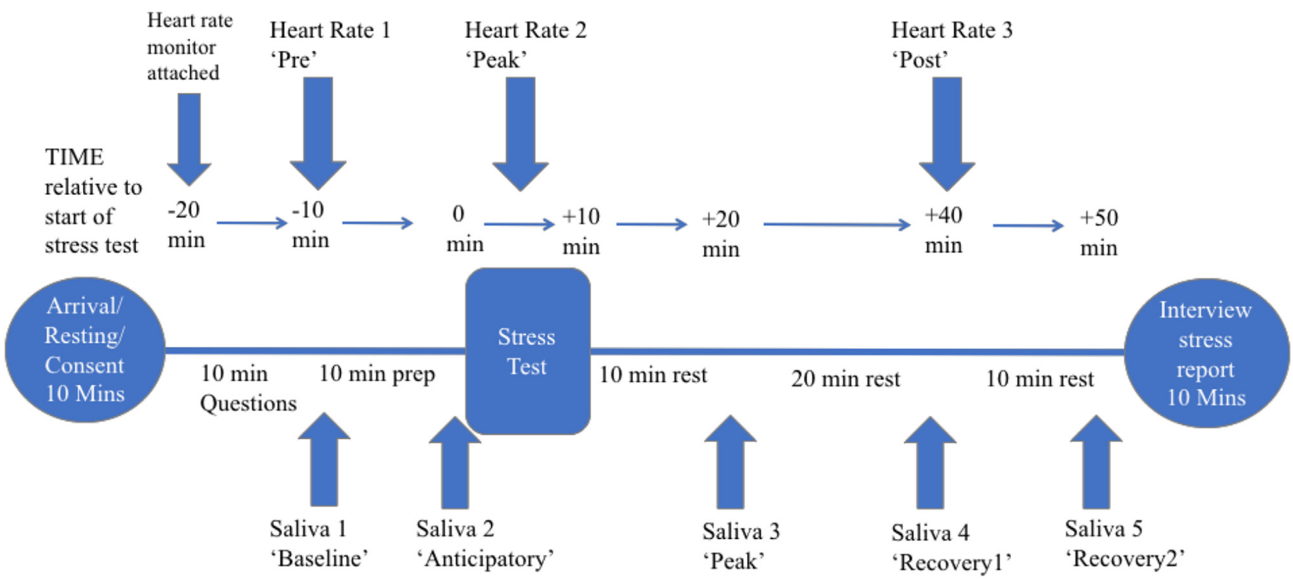

Fig. 2. Stress testing protocol showing testing points for outcome variables.

sphericity could not be assumed. Effect size is reported using partial eta squared (partial η^2) for ANOVA statistics or r values for correlation analysis (2-tailed).

Inferential statistics were conducted using mixed ANOVA procedures to examine stress reactivity over time as determined by the outcome variables of cortisol and heart rate, using TIME (five assessments for cortisol; three for heart rate) as the repeated assessment. Subjective stress experience scores obtained qualitatively in interview were coded as one of three groups and were entered into the mixed ANOVA as the between group factor. Post hoc paired sample *t*-test analyses for within participant comparisons across time-points (TIME) were interpreted using Bonferroni correction for multiple testing. Area under the curve (AUCg and AUCi) calculations were derived as cortisol indices as a measure of total output and overall reactivity response respectively (Pruessner, Kirschbaum, Meinlschmid, & Hellhammer, 2003). Peak value minus baseline value (delta peak) were calculated separately for cortisol and heart rate as a measure of peak reactivity to the stressor; and peak to post stressor cortisol and heart rate calculated as a measure of recovery by calculating delta recovery as post minus peak. These three calculations were used in order to conduct correlation analyses that assessed associations between these cortisol and heart rate indices and the self-report questionnaire scores for perceived stress and coping subscales.

3. Results

3.1. Salivary cortisol responses

A significant main effect of TIME was observed for salivary cortisol across the five points of assessment, $F\,(3,41) = 3.306$, $p = .035$, partial $\eta^2 = 0.171$ (see Fig. 3). Pairwise comparisons (Bonferroni corrected $p < .01$) indicated a non-significant decrease from baseline to anticipation sample, a significant increase from anticipation to peak ($p = .008$), a non-significant decline from peak to recovery1, and a significant decline from recovery1 to recovery2 ($p = .007$). Due to gender differences within the sample we conducted a post hoc follow up analysis with women only (n = 16); effect of TIME remained significant at the same level ($p = .049$).

3.1.1. Salivary cortisol responses by subjective stress experience

The subjective stress reports (three groupings) included as a between participant comparison in the mixed ANOVA revealed no significant main effect of GROUP for cortisol reactivity and no significant interaction of TIME x GROUP. Means for the subjective groupings across TIME are illustrated in Fig. 4 for descriptive purposes. They reveal a pattern of subjective-objective congruity for the "no stress" group; the "task only stressful" group reveal the expected pattern of response and adaptation to the stress task and those reporting both the "tasks and robots" as stressful show an extended stress response beyond peak and into recovery1 phase.

3.2. Heart rate responses

A significant main effect of TIME was observed for heart rate across the three sampling points, $F\ (2,28) = 22.13$, $p < .001$, partial $\eta^2 = 0.58$ (see Fig. 5). Pairwise comparisons (Bonferroni correction $p < .016$) indicated a non significant increase from pre to peak ($p = .024$) and a significant decline from peak to post stress test ($p < .001$). Due to gender differences within the sample we conducted a post hoc follow up analysis with women only (n = 16); effect of TIME remained significant at the same level ($p < .001$).

3.2.1. Heart rate responses: assessed by subjective stress experience

The subjective stress reports (three groupings) included as a between participant comparison in the mixed ANOVA revealed no significant main effect of GROUP for cortisol reactivity and no significant interaction of TIME x GROUP (group means are illustrated in Fig. 6).

Means for the subjective groupings across TIME are illustrated in Fig. 6 for descriptive purposes. They reveal an indicative pattern of subjective-objective incongruity for the "no stress" with higher heart rate elicited in those rating the "task and robot" as stressful than those only rating the "task only" as stressful.

3.3. Relationships between stress response, perceived stress and coping

3.3.1. Salivary cortisol

In examination of relationships between salivary cortisol indices of AUC and delta peak with questionnaire assessed perceived stress and coping responses, AUCg was not significantly associated with perceived stress or with any of the coping responses with the exception of self-blame ($r = 0.56$, $p = .013$). AUCi and delta peak cortisol were not significantly associated with perceived stress and a relationship observed only for the coping response of planning ($r = .50$, $p = .031$; $r = 0.59$, $p = .033$). These effects suggest no effects of pre-existing

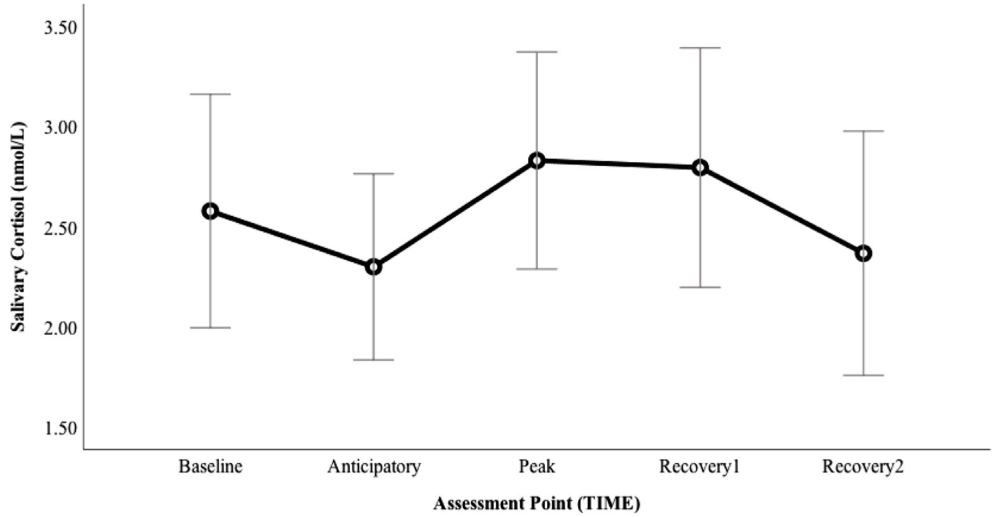

Fig. 3. Salivary cortisol (mean) assessed across five time points pre to post stress testing.

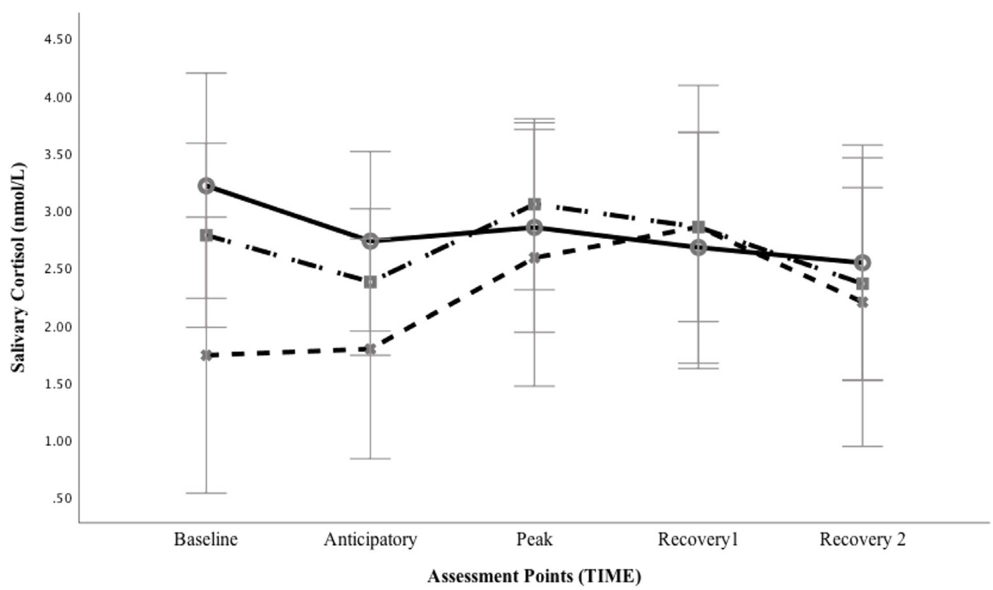

Fig. 4. Salivary cortisol responses by subjective stress experience. Groupings indicated as: No stress **____**(n = 6)/Task only stressful **-.-.-.-.**(n = 9)/Task and robot stressful **- - - -**(n = 4).

perceived stress levels and the influence of some specific trait coping responses in response to this non-human stress panel on cortisol stress reactivity.

3.3.2. Heart rate

Perceived stress scores were not significantly associated with heart rate over time when assessed using a within participant ANOVA, indicating no effect of perceived stress on heart rate reactivity or recovery. Delta peak heart rate was negatively correlated with the coping strategies of active coping (r = -0.52, p = .022) and positive reframing (r = -0.49, p = .035). This suggests that these trait coping strategies are effective in inducing less of an increase in heart rate in response to the stress test. Delta recovery heart rate was positively correlated with active coping (r = 0.48, p = .039) and planning (r = 0.49, p = .038) indicating that these trait coping responses as negatively associated with heart rate recovery following the stress test (slower recovery).

4. Discussion

4.1. Stress reactivity using a non-human robot audience

In this proof of concept study, we evaluated the feasibility of a TSST style acute stress paradigm with a manipulation that involved using a pre-recorded non-human robot audience in place of a human audience. Initial support was found for our primary hypothesis, that the non-human robot audience would be capable of inducing a stress response in participants via SET as evidenced by cortisol and heart rate reactivity. However, the magnitude of effect was considerably smaller than that seen in studies using traditional TSST protocols. We also aimed to examine the subjective verbal reports of the stress test experience as compared with the objective physiological assessments. Within the subjective groupings we note indications of congruity with objective assessments for cortisol but incongruity for heart rate reactivity. We also found no evidence for a relationship between underlying perceived stress levels on cortisol or heart rate response to the stressor but some specific trait coping responses were significantly correlated with cortisol and heart rate stress reactivity. We discuss these

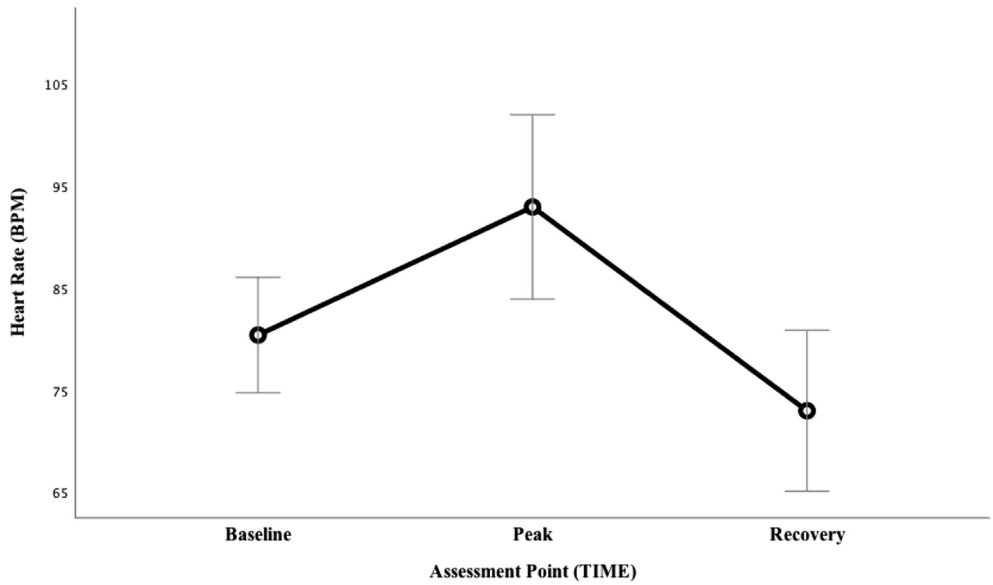

Fig. 5. Heart rate response (BPM) assessed from mean of 1-min extraction across three timepoints pre to post stress testing.

findings in the context of preliminary indicative effects worthy of further systematic investigation.

Findings from this study provide some initial evidence for the ability of a non-human robot stress audience to elicit social evaluative threat in human adult participants. Despite a number of increasingly authentic and convincing VR adaptations of the TSST, this study is the first to use a non-human robot audience in an acute stress paradigm. In doing so it maintains the conventional TSST style panel set-up whilst shifting the focus of attention from a comparison of in-person versus VR TSST environment to a physical environment with a manipulated pre-recorded audience of non-human avatars. This provided the potential to examine in more depth the theories of social evaluative threat (Dickerson & Kemeny, 2004) and the media equation (Reeves & Nass, 2002), as evaluated together by Kothgassner et al. (2016). We were only in part able to address the degree to which participants experience social evaluation and judgement and assess whether social threat is perceived and elicited as such when from a non-human evaluator. Overall, the experience of SET as assessed by the mounting of a stress response within this robot paradigm provides some evidence that participants ascribe human evaluative properties to non-human 'judges'. It indicates that albeit to a relative small degree in this sample, when taken at face value the robot audience appeared able to induce the familiar stressor characteristics of social evaluation, lack of control, unpredictability and time pressure (Dickerson & Kemeny, 2004; Miller, Chen, & Zhou, 2007). This finding is in line with other TSST studies including a number of VR studies (see Allen et al., 2017) in which reactivity was observed but to a lesser degree than non-VR or fully immersive VR environments (Kotlyar et al., 2008; Wallergard et al., 2011; Montero-Lopez et al., 2016; Shiban et al., 2016; Kothgassner et al., 2016). However, it could be argued that the magnitude of cortisol response is comparatively low and may be more akin to protocols designed to elicit stress via a mentally challenging task without social evaluation, questioning whether SET was effectively elicited from the robot presence or simply an effect of the task itself. Further work is needed to address the application of a robot panel in more depth, using relevant control and comparison groups to systematically disentangle the relative contribution of SET within this context. The complexity of incongruity between subjective and objective ratings as described below also needs consideration in further

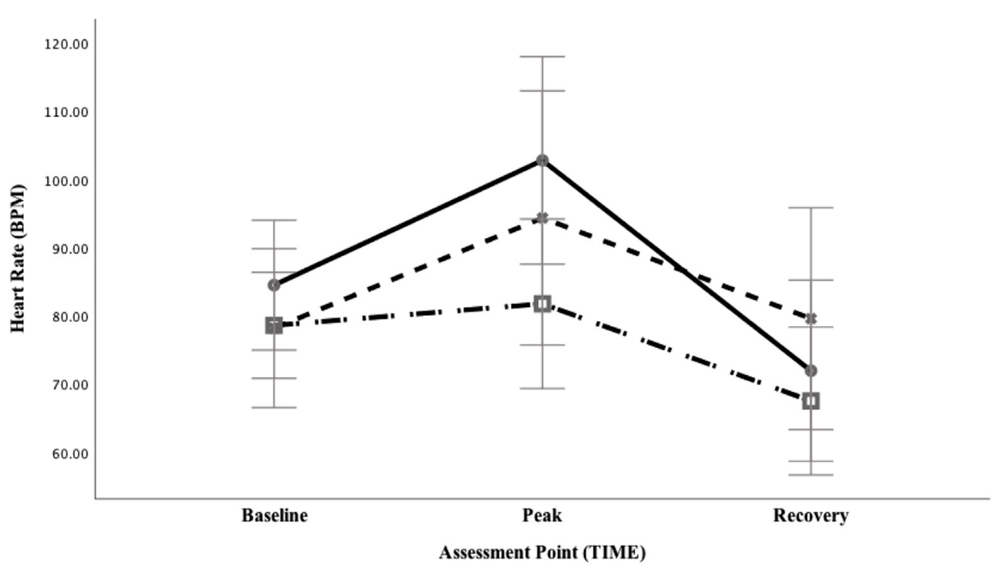

Fig. 6. Heart rate response (BPM) assessed from mean of 1-min extraction by subjective stress experience. Groupings indicated as: No stress ____(n = 6)/Task only stressful -.-.-.-.(n = 9)/Task and robot stressful - - - -(n = 4).

evaluating these effects.

4.2. Objective assessment and subjective evaluation of the stress experience

Evaluation of subjective responses from the extended manipulation interview enabled a comparison with the objective assessment of heart rate and salivary cortisol. Whilst there were no significant differences in reactivity between the groups, descriptive comparison of subjective ratings indicates congruity with objective salivary cortisol assessment and incongruity with heart rate reactivity. For cortisol, participants who reported experiencing no stress during the task despite being higher on the first two cortisol samples, show a corresponding flat response to the task. The task only stressful group revealed the expected cortisol peak followed by recovery pattern, whilst the task and robot stressful group show a delayed peak in cortisol. For heart rate reactivity, it was the subjective "no stress" group who showed the greatest response whilst those reporting the robots as stressful revealed a more pronounced reaction than the task only stressful group. The only other study to have conducted a similar extended manipulation check in this way (Cheetham & Turner-Cobb, 2016) was with a child cohort of 7–11 year-olds, in which cortisol responses were consistent with objective measures of cortisol. Other stress reactivity studies have compared subjective questionnaire or visual analogue assessments with objective measures, usually in adult populations and frequently found incongruity between verbal reports of the stress experience and measures such as heart rate and salivary cortisol (e.g. Coppens et al., 2018). In the current study the subjective assessment was not simply about whether participant reports of stress were consistent with objective assessments but were focused around their experience of the stress inducing nature of the non-human robot audience. It is possible that those reporting only the tasks as stressful, were either unaware of the effect of the robot component to activate a stress response or were aware but not willing to divulge that they found the robots stressful. However, a full evaluation of these effects was not possible with this current proof of concept study given the small numbers in individual groups and lack of a control group. It is not surprising that a dispassionate robot entity could generate a stress response given that it is widely acknowledged that during human interaction with robotic entities individuals tend to ascribe human characteristics (Reeves & Nass, 2002). Furthermore, dispassionate involvement with participants is precisely the intended style of the traditional in-person TSST paradigm, which is designed to maximize participant SET and in this way is not dissimilar to the robot audience.

4.3. Perceived stress and trait coping responses

In examining the psychosocial self-report measures, underlying perceived stress levels were not correlated with cortisol total output, overall reactivity or peak response, or with changes in heart rate. These effects indicate a lack of involvement of pre-existing perceived stress levels on cortisol and heart rate stress reactivity. Janson and Rohleder (2017) also report acute stress effects as independent of perceived stress using the same scale as reported here. Whilst underlying chronic stressor levels are important potential indicators of reactivity in acute stress settings, specific life stressors such as caregiving, early life adversity or low socioeconomic status known to effect physiological functioning (Rohleder, 2014) were not the focus of the current study. For the fourteen trait coping responses assessed, positive associations were found between self-blame and cortisol output and between planning and cortisol reactivity to the stressor; whilst active coping and positive reframing were associated with lower heart rate reactivity to the stressor and use of active coping and planning were independently related to slower heart rate recovery. It is possible that these effects were influenced by multiple testing and with a Bonferroni correction applied, the findings do not reach the required ($p < .004$) level of significance. However, in the context of this proof of concept study these findings highlight the relevance of specific trait coping responses and are indicative of potential differences that require further focus. For example, Planning as a trait coping style was suggested as being comparatively poor in regard to coping with the acute stressor for both cortisol reactivity during the stress test and slower return to heart rate baseline in recovery. Alternatively, Active coping was a potentially useful response during the stress test but indicated slower return to baseline in recovery. A number of previous studies have reported the beneficial effects of certain types of coping responses dependent on the type of stressor, its characteristics and the restrictions it imposes. For example, Janson and Rohleder (2017) acknowledge the beneficial effect of distraction coping in stressful conditions where control of the situation is not possible and the negative effects of responses such as self-blame in inescapable socially evaluative situations. Similarly, whilst planning may be seen as adaptive coping in many situations, in one where plans cannot be acted upon to diffuse the stress, a poorer outcome may result (Connor-Smith, Compas, Wadsworth, Thomsen, & Saltzman, 2000). The trait coping response of positive reframing is in keeping with the shift and persist strategy (Chen, Miller, Lachman, Gruenewald, & Seeman, 2012) that indicates the need to cognitively alter perception of the situation that cannot be altered or controlled by behaviour. This type of cognitive alteration of the situation is seen in TSST intervention studies, such as shifting to the use of compassionate goals in order to cope (Abelson et al., 2014). Furthermore, distinction between the phases of reactivity to an acute stressor and post stress recovery pattern has been indicated as important in previous work (Kudielka, Buske-Kirschbaum, Hellhammer, & Kirschbaum, 2004).

4.4. Limitations and future research

The strength of this study is the valuable extension of the TSST acute stressor paradigm to a non-human robot stress testing audience. This pushes the boundaries beyond recent successful VR technologies (e.g. Kothgassner et al., 2016; Shiban et al., 2016) that applied the TSST and allows for theoretical testing, the implications of which have potential for development of interventions through audience manipulation. As noted, this was a pilot feasibility study and as such there are a number of significant limitations which restrict the generalizability and interpretation of findings. Firstly, we acknowledge that the study was underpowered for subgroup analyses due to a small overall sample size of 19 participants, although noting that effect sizes were medium to large. In comparison however, other work such as that by Jonsson et al. (2010) had only 10 participants in their landmark fully immersive VR study with significant findings and implications. The scope for inclusion of psychosocial measures to assess coping styles in this stress testing context is identified in this work but a larger sample size is essential to avoid issues associated with multiple comparisons. Similarly, a focus on specific relevant coping styles as suggested here or with a larger sample through factor analysis of coping types is needed. Secondly, the lack of a control or comparison group is a further shortcoming of the present study. In order to fully examine the effect of a robot audience in future work it would be crucial to include a number of comparison conditions. For example, an in-person audience, a VR audience, and a non-human robot audience present in the room in addition to a pre-recorded video, as well as a exposure simply to the robot audience without the stress testing paradigm applied. Thirdly, whilst the present focus was on testing the robot panel via video, it was necessary within the current study to retain the use of a researcher in the stress testing room to provide verbal instructions to participants about the tasks. Although the researcher maintained a very low profile, we identify this as a potential confound and significant limitation; the mere presence of the researcher in the room may have impacted on participant stress reactivity. Since the robots were not communicating directly with the participant, this may have detracted from the ruse that they were in real time. Future studies are needed that eliminate human presence entirely in the stress testing room, which could be achieved through robotic voice simulation with appropriate software and equipment. This would eliminate the potential contamination effect and help to create a more believable scenario. Furthermore, the particular robot used was reproduced in triplicate via video with different mannerisms. Pre-recorded in this way the audience may not have been as convincing or stress inducing as a robot audience that was present in the room with differently programmed movements and mannerisms occurring in real time. The Nao robot also has features that may be deemed appealing or at least relatively non-threatening, such as facial appearance and short stature. Further work is needed that accounts for the appearance characteristics of the robot audience and is able to manipulate these characteristics across different testing conditions in order to examine the effects of specific attributes. Fourthly, we acknowledge a number of limitations with regard to salivary cortisol sampling in this study which present potential confounding factors. Whilst we attempted to control for time of day by sampling in the afternoon, we did not measure waking time or atypical sleep/wake patterns that could have influenced cortisol diurnal variability. Additional potential confounding factors either not assessed or controlled for in this study include menstrual cycle phase, age and sex; these need consideration in future work. The age range in this study was relatively broad, including adults from 21 to 57 years. The study was underpowered to assess differences across this age group but it is feasible that stress reactivity to a robotic panel may have differential effects on older and younger adults. This highlights the importance of panel manipulation in accordance with participant characteristics including age. Sex differences also require further examination; in the current study only three of the sample were male and as such meaningful comparisons could not be made.

Further exploration of participant psychosocial stress and coping characteristics are called for and possible interventions to reduce stress responses developed once these factors have been assessed in greater detail.

5. Conclusions

This proof of concept study enabled a demonstration of the application of social stress testing using a non-human robot audience. It is the first study to manipulate the audience in this way, using robots in place of humans, whether in-person or VR. The concept and findings apply the theory of SET and the media equation to address the experience of participants under acute stress conditions, when the vehicle for inducing stress is non-human. In the context of much discussed dispassionate technological advances, there is further ecological validity in the issues raised by extending the social stress testing paradigm to a non-human robot audience. These concern the relationship between human-robot interaction in stress induction and the potential to facilitate future work that enhances coping and adaptation in uncontrollable acute stress situations.

Funding

The Nao robot used in this project was funded by the Engineering and Physical Sciences Research Council (EPSRC), UK under its IDEAS Factory Sandpits call on Digital Personhood, grant ref: EP/L00416X/1.

Declarations of interest

None.

References

Abelson, J. L., Erickson, T. M., Mayer, S. E., Crocker, J., Briggs, H., Lopez-Duran, N. L., et al. (2014). Brief cognitive intervention can modulate neuroendocrine stress responses to the trier social stress test: Buffering effects of a compassionate goal orientation. *Psychoneuroendocrinology, 44*, 60–70. https://doi.org/10.1016/j.psyneuen.2014.02.016.

Allen, A. P., Kennedy, P. J., Dockray, S., Cryan, J. F., Dinan, T. G., & Clarke, G. (2017). The trier social stress test: Principles and practice. *Neurobiol Stress, 6*, 113–126. https://doi.org/10.1016/j.ynstr.2016.11.001.

Bendezu, J. J., Sarah, E. D., & Martha, E. W. (2016). What constitutes effective coping and efficient physiologic regulation following psychosocial stress depends on involuntary stress responses. *Psychoneuroendocrinology, 73*, 42–50. https://doi.org/10.1016/j.psyneuen.2016.07.005.

Carver, C. S. (1997). You want to measure coping but your protocol's too long: Consider the brief COPE. *International Journal of Behavioral Medicine, 4*(1), 92–100. https://doi.org/10.1207/s15327558ijbm0401_6.

Carver, C. S., Scheier, M. F., & Weintraub, J. K. (1989). Assessing coping strategies: A theoretically based approach. *Journal of personality and social psychology, 56*(2), 267–283.

Cheetham, T. J., & Turner-Cobb, J. M. (2016). Panel manipulation in social stress testing: The Bath experimental stress test for children (BEST-C). *Psychoneuroendocrinology, 63*, 78–85. https://doi.org/10.1016/j.psyneuen.2015.09.013.

Chen, E., Miller, G. E., Lachman, M. E., Gruenewald, T. L., & Seeman, T. E. (2012). Protective factors for adults from low-childhood socioeconomic circumstances: The benefits of shift-and-persist for allostatic load. *Psychosomatic Medicine, 74*(2), 178–186. https://doi.org/10.1097/PSY.0b013e31824206fd.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior, 24*(4), 385–396.

Connor-Smith, J. K., Compas, B. E., Wadsworth, M. E., Thomsen, A. H., & Saltzman, H. (2000). Responses to stress in adolescence: Measurement of coping and involuntary stress responses. *Journal of Consulting and Clinical Psychology, 68*(6), 976–992.

Coppens, E., Kempke, S., Van Wambeke, P., Claes, S., Morlion, B., Luyten, P., et al. (2018). Cortisol and subjective stress responses to acute psychosocial stress in fibromyalgia patients and control participants. *Psychosomatic Medicine, 80*(3), 317–326. https://doi.org/10.1097/PSY.0000000000000551.

von Dawans, B., Kirschbaum, C., & Heinrichs, M. (2011). The trier social stress test for groups (TSST-G): A new research tool for controlled simultaneous social stress exposure in a group format. *Psychoneuroendocrinology, 36*(4), 514–522. https://doi.org/10.1016/j.psyneuen.2010.08.004.

Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin, 130*(3), 355–391. https://doi.org/10.1037/0033-2909.130.3.355.

Gilgen-Ammann, R., Buller, M. J., Bitterle, J. L., Delves, S. K., Veenstra, B. J., Roos, L., et al. (2018). Evaluation of pulse rate measurement with a wrist worn device during different tasks and physical activity. *Current Issues in Sport Science, 3*, 1–9. https://doi.org/10.15203/CISS_2018.011.

Het, S., Rohleder, N., Schoofs, D., Kirschbaum, C., & Wolf, O. T. (2009). Neuroendocrine and psychometric evaluation of a placebo version of the 'trier social stress test. *Psychoneuroendocrinology, 34*(7), 1075–1086. https://doi.org/10.1016/j.psyneuen.2009.02.008.

Hohne, N., Poidinger, M., Merz, F., Pfister, H., Bruckl, T., Zimmermann, P., & Ising, M. (2014). Increased HPA axis response to psychosocial stress in remitted depression: The influence of coping style. *Biological Psychology, 103*, 267–275. https://doi.org/10.1016/j.biopsycho.2014.09.008.

Janson, J., & Rohleder, N. (2017). Distraction coping predicts better cortisol recovery after acute psychosocial stress. *Biological Psychology, 128*, 117–124. https://doi.org/10.1016/j.biopsycho.2017.07.014.

Jonsson, P., Wallergard, M., Osterberg, K., Hansen, A. M., Johansson, G., & Karlson, B. (2010). Cardiovascular and cortisol reactivity and habituation to a virtual reality version of the trier social stress test: A pilot study. *Psychoneuroendocrinology, 35*(9), 1397–1403. https://doi.org/10.1016/j.psyneuen.2010.04.003.

Kirschbaum, C., Pirke, K. M., & Hellhammer, D. H. (1993). The 'Trier Social Stress Test'–a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology, 28*(1–2), 76–81. https://doi.org/10.1159/000119004.

Kothgassner, O. D., Felnhofer, A., Hlavacs, H., Beutl, L., Palme, R., Kryspin-Exner, I., et al. (2016). Salivary cortisol and cardiovascular reactivity to a public speaking task in a virtual and real-life environment. *Computers in Human Behavior, 62*, 124–135.

Kotlyar, M., Donahue, C., Thuras, P., Kushner, M. G., O'Gorman, N., Smith, E. A., et al. (2008). Physiological response to a speech stressor presented in a virtual reality environment. *Psychophysiology, 45*(6), 1034–1037. https://doi.org/10.1111/j.1469-8986.2008.00690.x.

Kudielka, B. M., Buske-Kirschbaum, A., Hellhammer, D. H., & Kirschbaum, C. (2004). Differential heart rate reactivity and recovery after psychosocial stress (TSST) in healthy children, younger adults, and elderly adults: The impact of age and gender. *International Journal of Behavioral Medicine, 11*(2), 116–121. https://doi.org/10.1207/s15327558ijbm1102_8.

Miller, G. E., Chen, E., & Zhou, E. S. (2007). If it goes up, must it come down? Chronic stress and the hypothalamic-pituitary-adrenocortical axis in humans. *Psychological Bulletin, 133*(1), 25–45. https://doi.org/10.1037/0033-2909.133.1.25.

Montero-Lopez, E., Santos-Ruis, A., Garcia-Rios, M. C., Rodriguez-Blazquez, R., Perez-Garcia, M., & Peralta-Ramirez, M. I. (2016). A virtual reality approach to the trier social stress test: Contrasting two distinct protocols. *Behavior Research Methods, 48*, 223–232. https://doi.org/10.3758/s13428-015-0565-4.

Pruessner, J. C., Kirschbaum, C., Meinlschmid, G., & Hellhammer, D. H. (2003). Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. *Psychoneuroendocrinology, 28*(7), 916–931.

Reeves, B., & Nass, C. (2002). *The media equation: How People treat computers, television, and new media like real People and places.* Palo Alto, CA: CSLI Publications.

Rohleder, N. (2014). Stimulation of systemic low-grade inflammation by psychosocial stress. *Psychosomatic Medicine, 76*(3), 181–189. https://doi.org/10.1097/PSY.0000000000000049.

Sartor, F., Gelissen, J., van Dinther, R., Roovers, D., Papini, G. B., & Coppola, G. (2018).

Wrist-worn optical and chest strap heart rate comparison in a heterogeneous sample of healthy individuals and in coronary artery disease patients. *BMC Sports Science, Medicine and Rehabilitation, 10*, 10. https://doi.org/10.1186/s13102-018-0098-0.
Shiban, Y., Diemer, J., Brandl, S., Zack, R., Muhlberger, A., & Wust, S. (2016). Trier Social Stress Test in vivo and in virtual reality: Dissociation of response domains. *International Journal of Psychophysiology, 110*, 47–55. https://doi.org/10.1016/j.ijpsycho.2016.10.008.
Thiebaud, R. S., Funk, M. D., Patton, J. C., Massey, B. L., Shay, T. E., Schmidt, M. G., et al. (2018). Validity of wrist-worn consumer products to measure heart rate and energy expenditure. *Digital Health, 4*, 1–7. https://doi.org/10.1177/2055207618770322.
Villada, C., Hidalgo, V., Almela, M., & Salvador, A. (2016). Individual differences in the psychobiological response to psychosocial stress (trier social stress test): The relevance of trait anxiety and coping styles. *Stress and Health, 32*(2), 90–99. https://doi.org/10.1002/smi.2582.
Wallergard, M., Jonsson, P., Osterberg, K., Johansson, G., & Karlson, B. (2011). A virtual reality version of the trier social stress test: A pilot study. *Presence-Teleop Virt, 20*, 325–336.
Westenberg, P. M., Bokhorst, C. L., Miers, A. C., Sumter, S. R., Kallen, V. L., van Pelt, J., et al. (2009). A prepared speech in front of a pre-recorded audience: Subjective, physiological, and neuroendocrine responses to the leiden public speaking task. *Biological Psychology, 82*(2), 116–124. https://doi.org/10.1016/j.biopsycho.2009.06.005.