PHILOSOPHICAL TRANSACTIONS B

royalsocietypublishing.org/journal/rstb

Live human–robot interactive public demonstrations with automatic emotion and personality prediction

Hatice Gunes[1], Oya Celiktutan[2] and Evangelos Sariyanidi[3]

[1]Department of Computer Science and Technology, University of Cambridge, Cambridge CB3 0FD, UK
[2]Centre for Robotics Research, Department of Informatics, King's College London, London WC2R 2LS, UK
[3]Center for Autism Research, Philadelphia, PA, USA

HG, 0000-0003-2407-3012; OC, 0000-0002-7213-6359

Research

Cite this article: Gunes H, Celiktutan O, Sariyanidi E. 2019 Live human–robot interactive public demonstrations with automatic emotion and personality prediction. *Phil. Trans. R. Soc. B* **374**: 20180026. http://dx.doi.org/10.1098/rstb.2018.0026

Accepted: 14 January 2019

One contribution of 17 to a theme issue 'From social brains to social robots: applying neurocognitive insights to human–robot interaction'.

Subject Areas:
behaviour, cognition

Keywords:
real-time human–robot interaction, facial action units, facial expressions, affect, personality, public demonstration

Author for correspondence:
Hatice Gunes
e-mail: hatice.gunes@cl.cam.ac.uk

Communication with humans is a multi-faceted phenomenon where the emotions, personality and non-verbal behaviours, as well as the verbal behaviours, play a significant role, and human–robot interaction (HRI) technologies should respect this complexity to achieve efficient and seamless communication. In this paper, we describe the design and execution of five public demonstrations made with two HRI systems that aimed at automatically sensing and analysing human participants' non-verbal behaviour and predicting their facial action units, facial expressions and personality in real time while they interacted with a small humanoid robot. We describe an overview of the challenges faced together with the lessons learned from those demonstrations in order to better inform the science and engineering fields to design and build better robots with more purposeful interaction capabilities.

This article is part of the theme issue 'From social brains to social robots: applying neurocognitive insights to human–robot interaction'.

1. Introduction

A social robot is defined as a (semi-)autonomous robot that is able to communicate with humans or other robots, and engage in social interactions by following social behaviours and norms [1]. Arguably, the most important aspect of a social robot is its perception capability—if a robot is able to accurately understand its surrounding world and its people, it can be made to communicate in an appropriate and social manner.

Social robots are a disruptive technology that have an enormous potential to transform multiple domains. However, the public, largely misled by sci-fi movies and driven mainly by fear or fascination, has skewed opinions and unrealistic expectations of these robots. Therefore, there is a genuine need for scientists working in the fields of robotics and artificial intelligence to demonstrate their work and engage the public.

The field of social robotics is exponentially growing and evolving, motivated by a wide range of promising applications in public settings [2] including assisting people in hospitals, schools, shopping malls [3] and homes [4]. User profiling and behavioural adaptation is key in deploying social robots in such dynamic environments. Rossi *et al.* [5] divided user profiling strategies into three categories, namely, physical, cognitive and social, which are essential to enhance user's satisfaction and robot acceptance. Physical profiling is concerned with processing human sensory inputs and recognizing actions and activities, whereas cognitive and social profiling requires a higher level of interpretation of human behaviours, namely recognizing their intentions, inferring their mental states, personalities and emotions. In these applications, the success of social robots will depend on how proper use profiling can be achieved, and how effective behavioural adaptation can be made

THE ROYAL SOCIETY PUBLISHING

based on the users' profiles, and to what degree individuals come to trust the robot that assists them.

Within the scope of cognitive and social user profiling, affective and social signals play a prominent role. Humans exchange information and convey their thoughts and feelings through gaze, facial expressions, body language and tone of voice along with spoken words, and infer 60–65% of the meaning of the communicated messages from these non-verbal behaviours [6]. These non-verbal behaviours carry significant information regarding higher-level social phenomena such as emotions, personality and engagement. Recognizing and interpreting these signals comes naturally for humans. The main thrust of an effective human–robot interaction platform should be to empower robots with similar skills.

In this paper, we describe a number of public demonstrations we conducted for automatically sensing and analysing human participants' non-verbal behaviours in real time while the participant interacted with a small humanoid robot in two different contexts: (i) predicting their perceived personality via the MAPTRAITS-HRI system and (ii) predicting their facial action units (AUs) and facial expressions in the context of an interactive game via the TeachMeEQ system.

2. Background and related work

Researchers addressed the effect of various phenomena during human–robot interaction (HRI), such as cognitive biases [7], erroneous behaviour by the robot [8] or the social gaze [9], in order to better understand and improve HRI. In particular, a number of studies focused on predicting the personality and emotions of humans during HRI as we summarize in the following sections.

(a) Personality prediction in human–robot interaction

Incorporating human personality analysis to adapt a robot's behaviour for engaging a person in an activity is becoming an important component for social robots [10–12]. One prominent work by Rahbar *et al.* [13] focused on the prediction of the extroversion trait only, when a participant was interacting with the humanoid iCub [14], a robot shaped like a 4 year-old child. They extracted both individual features and interpersonal features. The individual features were associated with the participant's amount of movement. The interpersonal features modelled synchrony and dominance between the movements of iCub and the participant, as well as proxemics features (i.e. the distance between iCub and the participant). They achieved the best recognition results by fusing individual and interpersonal features.

Research has shown that humans tend to be attracted to characters who have either matching personality traits (similarity rule) or non-matching personality traits (complementarity rule) [15]. Salam *et al.* [11] investigated the impact of the participants' personalities on their engagement states in a setting where two participants interacted with a Nao robot.[1] Similar to Rahbar *et al.* [13], they extracted two sets of features, namely, individual and interpersonal features. Individual features described the individual behaviours of each participant, e.g. body activity. Interpersonal features characterized the interpersonal behaviours of the participants with respect to each other and the robot. These included the total amount of group movement, the relative body orientation of the participants with respect to the robot, etc. They first predicted the personality of each participant, and then combined the personality predictions with the individual and interpersonal features to recognize whether the participants were engaged or not. The best results were achieved using individual features together with personality predictions.

Motivated by applications such as childcare and education, a recent work by Abe *et al.* [16] focused on predicting children's extroversion and agreeableness during interactions with a social robot. To this effect, they observed their distance from the robot, their facial expressions and the duration of their eye contact during these interactions, which yielded an accuracy over chance.

Despite its importance, research on automatic personality analysis in the context of social robotics is scarce. To the best of our knowledge, there is no system that is integrated onto a robot, and performs real time analysis of personality in the course of human–robot interactions. One of the challenges is that, although modelling the dynamics of expressions and emotions has been extensively studied in the literature, how to model personality in a time-continuous manner has been an open problem. Most of the previous approaches make inference about personality from a post analysis of short behavioural episodes, ranging from 10 to 14 s to several minutes. During our demonstrations, we therefore used the MAPTRAITS system, which we specifically designed for predicting personality in real-time, in the course of interactions.

(b) Emotion recognition in human–robot interaction

Emotion recognition methods used by social robots were extensively surveyed by Yan *et al.* [1] and McColl *et al.* [17]. Here, we only considered the prominent works that performed the recognition task by automatically extracting features from visual cues, and integrated the developed method onto a robotic platform.

The categorical model of emotion has been the most widely adopted approach in the literature. Cid *et al.* [18] developed an emotion recognition system by extracting features based on the facial action coding system (FACS) [19], and implemented it on a robotic head, Muecas [20], for an imitation task. For emotion recognition, they first applied a preprocessing step to the face image taken by Muecas to normalize the illumination and remove the noise, and highlight the facial features. From the processed face, a set of edge-based features were extracted and modelled to detect a total of 11 AUs. The detected AUs were used to represent the four basic emotions of happiness, sadness, fear and anger, according to a rule-based approach, and were mapped on the Muecas robot to display the inferred emotion in real time. Boucenna *et al.* [21] used similar visual features for enabling the robot to learn facial expressions of emotion from interactions with humans, through an online learning algorithm. The Muecas robot was able to learn all the emotions successfully, except for sadness. This was due to the large intra-class variability for sadness, namely, each person expressed sadness in a different manner.

Leo *et al.* [22] developed an automatic emotion recognition system to measure the facial emotion imitation capability of children with Autism Spectrum Disorders (ASD). R25 is a small cartoon-character-like robot by Robokind[2] that was first made to display a facial expression, and then the child was instructed to imitate the displayed facial expression while being analysed through R25's camera located in its right eye. The emotion recognition method was based on a generic pipeline including four components: face detection, face registration, appearance-based face representation and classification. This method was tested on three children with ASD, and achieved good emotion recognition performance especially for happiness and sadness.

Robust facial expression recognition is technically challenging, especially if there is no control over illumination conditions and the age range of the intended participants is large; the latter is particularly problematic if young children are considered to be included, as facial expression datasets generally contain only adult participants. The technical challenges are compounded by the fact that expression recognition needs to be carried with real-time processing speed on a standard computer. With TeachMeEQ, we intended to do multiple live demonstrations in different locations and include children as well as adult participants. This required us to build a robust facial expression recognition pipeline. Three technical improvements have been critical to achieve this and to reach high accuracy: (i) using (neutral) features based on an initial calibration stage, (ii) using illumination-normalized spatio-temporal Gabor features and (iii) combining appearance and shape features. Those improvements and our facial expression recognition pipeline are discussed in more detail in §3bi.

3. Public demonstration platforms

In this section, we introduce two platforms that were publicly demonstrated throughout 2016 and 2017, namely, the MAPTRAITS-HRI system and the TeachMeEQ system. First, the MAPTRAITS-HRI system was demonstrated live in 2016 in the context of two public demonstrations, namely in a Research Showcase setting and an International Conference setting. In the subsequent year, the TeachMeEQ system was demonstrated live in the context of science communication at the Cambridge Science Festival 2017, the Wellcome Collections' Friday Late Spectacular—Body Language Event, and the Humans and Robots in Public Space Showcase. In the rest of this paper, we first review emotion and personality prediction in the context of human–robot interaction, and then describe the design and the execution of the demonstrations, and provide an overview of the challenges faced, together with the lessons learned.

(a) The MAPTRAITS-HRI system for automatic personality prediction

(i) The MAPTRAITS system

The MAPTRAITS system is a multimodal framework that performs automatic personality prediction according to the Big Five personality model in real time [23], which comprises the trait set of extroversion, agreeableness, conscientiousness, neuroticism and openness, and is the standard and widely used approach in the area of personality computing [24].

The MAPTRAITS-HRI system has been trained [25] using human data recorded with the SEMAINE system [26]. The MAPTRAITS-HRI dataset was created as a subset of the audio-visual recordings of the SEMAINE corpus [27]. It consists of 30 clips of 10 subjects interacting with three SEMAINE agents. Annotations for these data were obtained by asking the independent raters to provide their impressions continuously in time along the dimensions of agreeableness, openness, neuroticism, conscientiousness and extroversion. The temporal variability of personality impressions is examined by developing time-continuous assessment. Rather than obtaining a single rating for the whole clip, raters continuously recorded their annotations for the aforementioned dimensions as the clip of the target subject played. For feature extraction, we took into account a multitude of visual features including face appearance, face geometric and body features. We then used the long/short term memory neural networks for time-series regression to model the temporal relationships between the continuously generated annotations and extracted features. In table 1, we presented our best results for each feature type in terms of coefficient of determination (R^2) and mean square error (MSE), where body features were the most useful feature in general, yielding larger R^2 and smaller MSE as compared to face features.

We extended the MAPTRAITS system [28] to work in a human–robot interaction setting (MAPTRAITS-HRI system, henceforth).

(ii) The real-time demonstrator

The robotic system we used is the humanoid robot Nao developed by Aldebaran Robotics[1] with NaoQi v. 2.1, head v. 4.0 and body v. 25 operating on it.

In the MAPTRAITS-HRI system demonstration, one human participant is sitting facing the Nao robot. The participant is wearing a headset for voice analysis and a video camera is used for head gesture, facial and bodily expression analysis. The robot is speaking and showing both verbal and non-verbal behaviour. A computer screen next to the robot is displaying graphically the current system detection of the participants' face and the prediction of their perceived personality, estimated based on the participants' observable behaviour. The robot sustains the conversation by being an active speaker and listener using verbal utterances and head and hand gestures.

We adopted a Wizard of Oz (WoZ) interaction set-up to design the initial stages of the MAPTRAITS-HRI system for which the sensing component has been fully implemented, but not the robotic component. WoZ refers to a human operator, unknown to the participant, remotely controlling the robot [29]. WoZ is widely used in HRI studies particularly when 'the robot's hardware and design has been completed but the robot's sensory, motor or cognitive abilities are still limited' [29].

The non-verbal behaviour of the robot is adopted as it has been implemented by default. More specifically, the Wizard does not control the non-verbal aspects of the robot behaviour such as blinking, head movement, hand and arm gesturing, and body posture. The natural language processing aspect of the robot was controlled via a graphical user interface specifically designed with a pre-scripted structure and flow of the envisaged interaction. The Wizard was required to listen to the participant answer and select between two possible options either in the form of feedback (e.g. 'That is exciting!' or 'I see . . .') or asking the next question. Prior to the demonstrations, the Wizard learned how to use the interface to control the WoZ setup. In summary, by employing WoZ we aimed to learn the limitations of the automatic sensing system and if this impedes the interaction.

Table 1. The best prediction results per personality trait are highlighted in italics. AG, agreeableness; CO, conscientiousness; EX, extroversion; NE, neuroticism; OP, openness.

		AG	CO	EX	NE	OP
face	R^2	0.22	*0.41*	0.07	0.20	0.23
appearance	*MSE*	0.53	*0.38*	0.62	0.57	0.75
face	R^2	*0.23*	0.33	0.07	0.20	0.13
geometric	*MSE*	*0.47*	0.44	0.55	0.45	0.71
body	R^2	0.19	0.39	*0.10*	*0.26*	*0.25*
	MSE	0.60	0.39	*0.69*	*0.45*	*0.59*

Our goal was to compensate for these issues in the final autonomous version of the system.

In previous research, we analysed human-robot interactions with an extroverted and an introverted robot and found that people enjoyed interacting more with the robot that exhibited an extroverted personality [10,12]. Therefore, for the live MAPTRAITS-HRI demonstrator, we used the extroverted robot personality from our previous study [12], which displayed hand gestures and talked relatively fast and loud.

The conversation consisted of five parts: (1) *Greeting*. The robot initiates the interaction by greeting the person and making *small talk*, and asks the person about her name and occupation, and how the day has been. The robot provides verbal and nonverbal feedback via simple comments (e.g. 'That sounds amazing!/Exciting!'). (2) *Task*. The robot asks the person to bring her face closer so that he can learn her face. The robot provides verbal feedback (e.g. 'I have now learned your face/I could not learn your face, let's try again.'). (3) *Emotions*. The robot asks the person a personal question, i.e. 'Is there something you would like to change in your life?; Can you tell me about the best memory you have or the best event you have experienced in your life?/ Can you tell me about an unpleasant or sad memory you have had in your life?' The robot provides verbal and gestural feedback (e.g. 'How nice!'/'I understand.'). (4) *Opinion*. The robot asks the person about their feelings and knowledge of robots (e.g. 'What are your feelings toward robots? Do you like them?/Have you watched Wall-e? Do you like it?'). (5) *Performance*. The robot offers to dance (i.e. 'Would you like me to dance for you?'). Once the participant agrees, the robot performs Tai-Chi moves.

(iii) Observations

Even though we had the personality predictions printed in real time on the computer screen, participants did not pay much attention to that and simply focused on the conversation. In the conference demo, we had an additional screen that showed the overall personality prediction of the participant right after the conversation, and this was the only moment when participants were indeed interested in seeing their personality prediction results. The system listed personality predictions in terms of three dimensions: agreeableness, extroversion and neuroticism. When people were assessing their personality predictions, they did not focus on the precise meaning of these dimensions, they were rather interested in the connotations. For instance, people thought that both agreeableness and extroversion have a positive connotation, and they simply wanted to see if they score high in these *positive* personality dimensions, or if they score high in the *negatively perceived* neuroticism dimension, without being interested in the actual meaning of the personality traits.

(b) The TeachMeEQ system for automatic expression prediction

(i) The TeachMeEQ system

Our goal with the TeachMeEQ system was to elicit facial expressions with simple instructions from Nao. For this purpose, we decided to ask participants to display three out of the six basic expressions, namely, happiness, surprise and sadness, as in-house experiments showed that those were the best for generating cohesive expressions across participants. Indeed, multiple studies showed that those are the expressions that are the easiest to recognize for humans [30–32]; challenging the notion of universality for the six basic expressions, those studies showed that the recognition rates for expressions such as fear or disgust can be as low as in the range of 40–50%.

As a result, we decided to consider only the expressions of happiness, surprise and sadness. We opted for an FACS-based recognition of those expressions to address the possibility that participants display those expressions only partially; for example, a participant instructed to display the expression of surprise may display it only with eyebrow raising (AU 1 + 2), overlooking the lip part (AU 25) and/or the jaw drop (AU 26) movements. We focused on a total of seven AUs, namely, inner brow raiser (AU1), outer brow raiser (AU2), brow lower (AU4), cheek raiser (AU6), lip corner puller (AU12), lips parted (AU25) and jaw drop (AU26). For the automatic AU detection, Sariyanidi *et al.* [33] highlighted the importance of two practices: (i) combining shape and appearance features, which yields better performance because they carry complementary information, and (ii) using differential features that describe information with respect to a reference image (i.e. the neutral face in the case of emotion recognition). The main advantage of the differential features is to place higher emphasis on the facial action by reducing person-specific appearance cues. We therefore extracted four types of features, namely, shape, appearance, differential-appearance and differential-shape features. The details of this system are described by Ondras *et al.* [34].

During the on-the-fly tests, we ensured that we had the neutral face of human subjects by programming the robot to ask the participant to stand still and make a neutral face in front of the camera prior to beginning the interaction session. We trained four binary support-vector machine classifiers, each in conjunction with one of the above-mentioned feature types, per AU. The final AU detection decisions are obtained by fusing the outputs of the four individual classifiers. Specifically, we adopt the *consensus fusion* approach, where an AU is detected based on the condition that all four classifiers are in full agreement. Prior to using in live demonstrations, we evaluated the performance of this AU detection system offline using the MMI dataset [35] via five-fold cross-validation. Table 2 presents AU detection results of the four individual features as well as their combination via consensus fusion approach in terms of 2AFC metric [36], where higher AFC scores indicate a better recognition

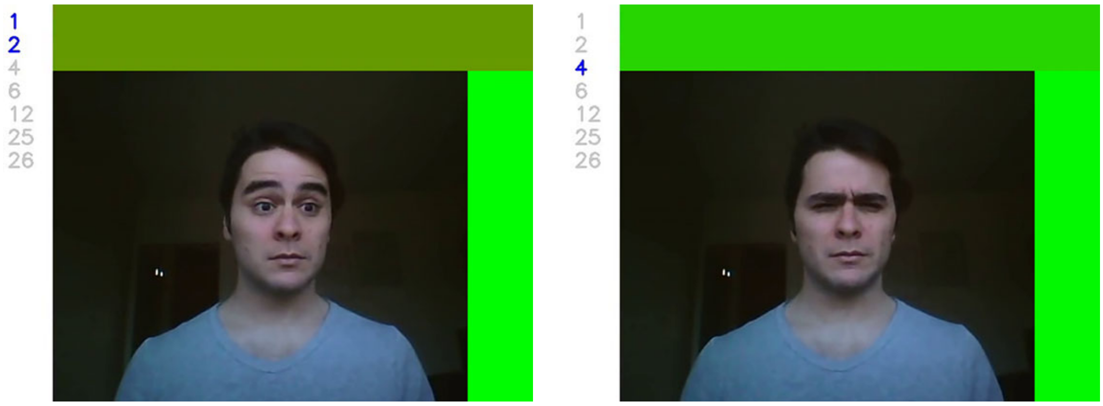

Figure 1. Illustration of AU detection results. Vertical and horizontal bars indicate the head rotation—green colour is associated with frontal/nearly frontal head pose. The detected AUs in each face image are highlighted in blue: AU1 and AU2, and AU4.

Table 2. AU detection performance in terms of 2AFC score. Italic text indicates the best (i.e. highest) score.

2AFC	AU1	AU2	AU4	AU6	AU12	AU25	AU26
shape	0.74	0.53	0.67	0.61	0.79	0.73	0.53
appearance	0.74	0.73	0.65	0.78	0.82	0.78	0.67
δ-shape	0.78	0.67	0.71	0.74	0.78	0.82	0.64
δ-appearance	0.90	*0.92*	*0.87*	0.82	0.92	*0.89*	0.78
fusion	*0.91*	0.89	0.78	*0.87*	*0.93*	0.86	*0.79*

performance. Looking at the AFC scores, the best-performing individual feature is the δ-appearance feature, and the consensus fusion achieves a higher AFC score than the δ-appearance feature for 4 (AU1, AU6, AU12, AU26) out of 7 AUs. We further used the best performing trained models in the real-time demonstration.

(ii) The real-time demonstrator

We performed the real-time implementation using C++ and integrated it onto the Nao robot. The computational power of the Nao robot did not allow us to run the AU detection algorithm in real time. For this reason, we used a pair of external cameras plugged into a laptop (Intel Core i6, 16 GB RAM), and ran the AU detection algorithm on the laptop. These cameras were attached to Nao's head using custom three-dimensional printed glasses. Example AU detections from the robot's point of view are shown in figure 1. Vertical and horizontal bars indicate the head pose, and green colour is associated with frontal/nearly frontal head poses that yield more reliable AU detection. The detected AUs are highlighted in blue on the left-hand side of each frame (e.g. AU1 and AU2 in figure 1*a*).

For the live demonstrations, the Nao robot was programmed to stand on a table while the participant sat facing the robot, participant's eye level matching the robot's eye level. The robot was programmed to interact with the participant autonomously by asking questions to the participant and sensing their audio-visual response (whether they say yes/no and what facial or hand gesture they display) through the cameras and the headphones mounted on his head.

The first part of the live demo focused on analysing the facial gestures of the participant. The Nao robot emulates a child who is only 4 years old, and it does not have all the facial features a human has such as eyebrows, lips, nostrils, etc. These make the robot less capable of expressing itself in an emotionally and socially intelligent way. Therefore, the motivation for the first part of the TeachMeEQ system is that the Nao robot needs to work on improving its emotional intelligence (EQ). The game starts with the robot asking the participant to teach him how to express himself emotionally by displaying facial gestures. The participant could choose to display any facial gesture such as pulling lip corners up (smile), pulling eyebrows up (surprise), dropping the mouth/chin (surprise), lowering the eyebrows (frown), etc. As illustrated in figure 2, Nao attempted to recognize each AU displayed by the participant, and inferred the expressed emotion based on a rule-based approach, and then asked the participant for feedback in the form of pressing the button on his left/right toes for yes/no.

Sample images from the Body Language Spectacular that took place in the Wellcome Collection, London, on the 4 November 2016,[3] are shown in figure 2. The images illustrate the moment that the participants from the public displayed different facial gestures.

(iii) Observations

The majority of the participants tended to sit far away from the robot and tended to back away when the robot prepared itself to stand up to start the game. The participants seemed not to listen to the robot fully and carefully, instead they had their own mental models of, for example, which button to press and when. More specifically, they would be more focused on how the robot looked and behaved and therefore not listen to his instructions and end up asking the

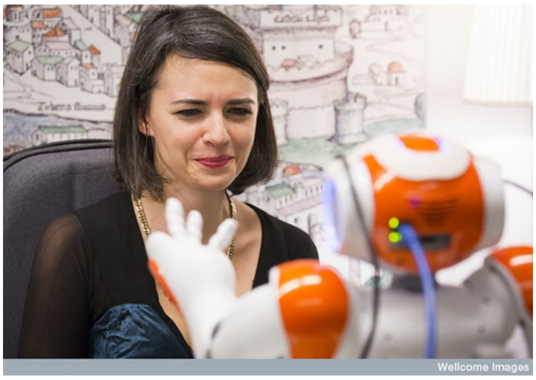

Figure 2. Images from the Body Language Spectacular at the Wellcome Collection (copyright: Wellcome Images).

experimenter questions such as 'what did he say?' and 'what do I press?' Additionally, they would not wait for the robot to finish its instructions, and would go ahead and press a button immediately before waiting for the robot to tell them which one to press. But when the robot asked them to touch its head to continue the game, the majority of the participants were reluctant to touch the robot's head, and they would first look at the experimenter for confirmation that this was indeed the request and it would be OK to touch the head of the robot.

When the robot asked the participants to display a neutral face participants were mostly displaying a smiling face assuming that this was their neutral face. The participants also tended to hold the expression they displayed for the robot for a very long time without the robot giving them instructions about this. We hypothesize that people assumed that the robot needed extra time to detect and recognize their facial gestures and expressions.

4. Challenges and lessons learned

Methods of evaluation for human studies in HRI are listed as (1) self-assessments, (2) interviews, (3) behavioural measures, (4) psychophysiology measures and (5) task performance metrics, with self-assessment and behavioural measures being the most common ones [37]. In our formal study [10], we took into account these methods and asked the participants to fill in a pre-study questionnaire to record demographics information and self-reported personality and a post-study questionnaire to evaluate their interaction experience [10]. We also recorded their physiological signals and audio-visual behaviour.

For the live demonstrations, we wanted to follow these methodologies and asked the participants to fill in questionnaires. However, we quickly realized that this was not going to work in a public demonstration due to the following reasons: (i) people were not interested in filling out formal documents, they wanted to have an experience with the technology; and (ii) there were other demos in the area that they wanted to try out and had only a limited amount of time. Although around 72 one-to-one interactions with the robot took place during our live public demonstrations, we are unable to report detailed statistics. In what follows, we summarize the challenges faced during the public demonstrations under novelty effect, assumptions and misconceptions, side effects, system effects, effects of the self-reported feedback and effects of WoZ, based on qualitative observations.

(a) Novelty effect

Participants are 'too' excited about being able to interact with a 'real' robot, which for some people was probably a first time experience. To overcome this issue in a 'formal' experimental setting, Kidd and Breazeal suggest that the participant should be introduced to the experiment and should be familiarized with the robot prior to the actual experiment taking place [38]. At the end of the interaction, the participant should ideally be interviewed and debriefed about the aims of the experiment. Following such a protocol is reported to reduce novelty effects [38]. But how do we define and follow such a protocol for public demonstrations? This remains an open question to be explored and will potentially enable better ways for gathering public-setting data.

(b) Assumptions and misconceptions

A humanoid robot that has 'a head with eyes suggests that the robot has advanced sensory abilities e.g. vision' [29, section 38.9]. During the MAPTRAITS-HRI demo participants therefore assumed that the robot sees them through its eyes which is not the case with Nao—its cameras are positioned around the mouth and the forehead. To mitigate this issue in the TeachMeEQ demo, we placed a three-dimensional printed headset with two cameras where the eyes are located. This correctly corresponded to the users' assumptions regarding the location of the eyes and ultimately provided better image quality to aid the automatic analysis process.

(c) Side effects

Participants are likely to experience and be affected by side effects of various events that occur during a public demonstration. These include effects attributed to the public nature of the environment–social desirability effect, which is due to participants answering in a way that they perceive as socially acceptable in the given situation, and high levels of variability due to noise/chatter, lighting and crowds, other people walking into the setting, other people asking questions, commenting or interrupting the interaction. The latter is also known as the Hawthorne effect [39], a phenomenon that relates to participants displaying certain behaviour because they know that they are being observed, in our case either by the robot or by others around them. Investigating the value of side effects, and exploring whether it should be mitigated in some ways, remain interesting research questions to be explored for HRI in public settings.

(d) System effects

Participants are likely to experience effects attributed to the system, e.g. the automatic analyses system crashing, the sensors halting, robot breaking or behaving in ways that were not anticipated. On the occasions that the system does not crash, there might be errors in the analysis of the human non-verbal behaviour and prediction of the personality or facial expressions due to the system not getting a clean input image. The ideal solution to such issues would however depend on the interaction type and the intended context. In our demonstrations, we observed that participants did not care about system effects as long as the speech and the behaviour of the robot was not affected.

(e) Effects of the self-reported feedback

There are no standardized methods for evaluating HRI in public settings. As pointed out by Bethel & Murphy [37] using a single evaluation measure is not sufficient to interpret accurately the responses of participants to a robot with which they are interacting. Self-assessments have problems with validity and corroboration—e.g. participants might report differently from how they are actually thinking or feeling [37]. It is indeed a challenge to attribute the participants' responses to their true behaviours. We expected that the system effects described above would directly affect the self-reported feedback. However, participants refrained from making negative comments about the system output. Rather, they used vague comments, which might be due to the social desirability effect.

(f) Effects of WoZ

The main critisim for WoZ is the ethical problems referred to as *Turing Deceptions* by Miller [40]. More specifically, it is not nice to deceive humans and make them believe that they are interacting with a fully autonomous robot when in reality they are interacting with a human that is hiding behind the robot. This is indeed a challenge for designing and executing public demonstrations that rely in one way or another on WoZ set-ups. During the MAPTRAITS-HRI demonstrations, we observed that the majority of the participants believed that the robot was fully autonomous. Despite differences in demographics, it was clear that *the public* expects a humanoid robot interacting with them to be fully autonomous in its perception, control and output. This was also confirmed during our discussions with the curator, Ben Russell, of the London Science Museum's 2017 exhibition that explored the 500-year story of humanoid robots.[4] Indeed, we observed this trend during the live TeachMeEQ public demos which were designed and executed as autonomous human–robot interactive demonstrations.

5. Conclusion

The availability of commercial robotic platforms and developments in collaborative academic research show we have achieved a lot, but the cognitive and social capabilities of the current humanoid robots are still very limited. There is a genuine need for scientists working in the fields of robotics and artificial intelligence to demonstrate their work and engage the public. As emphasized on the EPSRC's website,[5] this is important for two reasons: (i) to demystify the human-like robots and to help the general public become technology literate by creating a better understanding of the abilities and the potential of these robots and (ii) to acknowledge the public's concerns and get to know their views that can help steer how we develop human-like robots in the best interests of society.

In this paper, we presented the design and implementation of a number of live public demonstrations we have conducted in the period of 2015–2017 with the two proposed systems, namely the MAPTRAITS-HRI system and the TeachMeEQ system, in the context of science communication. These demonstrations aimed at automatically sensing and analysing human participants' non-verbal behaviour and predicting their personality, facial AUs and expressions in real time while they interacted with a humanoid robot (Nao).

HRI is known to have lower repeatability [1], and tools and metrics developed in human–computer interaction do not directly transfer to HRI. Public demonstrations provide insights into various aspect of human–robot interactions that may not be obvious or emerge during formal HRI studies. As the area of social robotics and HRI is growing, public demonstrations have the potential to provide insights about the robot/system effectiveness in public settings and reactions of the people. As indicated by our Challenges and Lessons Learned section, live public demonstrations enable us to better understand humans and inform the science and engineering fields to design and build better robots with more purposeful interaction capabilities.

Data accessibility. This article has no additional data.

Competing interests. We declare we have no competing interests.

Funding. This work was funded by the EPSRC under its IDEAS Factory Sandpits call on Digital Personhood (grant no. EP/L00416X/1).

Acknowledgements. This work was undertaken while O. Celiktutan and E. Sariyanidi were with the University of Cambridge.

Endnotes

[1]https://www.softbankrobotics.com/emea/en/nao.
[2]http://robokind.com/.
[3]https://web.archive.org/web/20161223135744/https://wellcome-collection.org/bodylanguage.
[4]https://www.sciencemuseum.org.uk/what-was-on/robots.
[5]https://epsrc.ukri.org/blog/robotics/.

References

1. Yan H, Ang MH, Poo AN. 2014 A survey on perception methods for human-robot interaction in social robots. *Int. J. Soc. Robot.* **6**, 85–119. (doi:10.1007/s12369-013-0199-6)
2. Foster ME, Alami R, Gestranius O, Niemelä O, Lemon M, Odobez J-M, Pandey AK. 2016 The MuMMER project: engaging human-robot interaction in real-world public spaces. In *8th Int. Conf. on Social Robotics, Kansas City, MO, 1–3 November 2016* (eds A Agah, J-J Cabibihan, AM Howard, MA Salichs, H Hepp), pp. 753–763. Berlin, Germany: Springer.
3. Sabelli AM, Kanda T. 2016 Robovie as a mascot: a qualitative study for long-term presence of robots in a shopping mall. *Int. J. Soc. Robot.* **8**, 211–221. (doi:10.1007/s12369-015-0332-9)
4. Agrigoroaie R, Ferland F, Tapus A. 2016 The enrichme project: lessons learnt from a first interaction with the elderly. In *8th Int. Conf. on Social Robotics, Kansas City, MO, 1–3 November 2016* (eds A Agah, J-J Cabibihan, AM Howard, MA Salichs, H Hepp), pp. 735–745. Berlin, Germany: Springer.
5. Rossi S, Ferland F, Tapus A. 2017 User profiling and behavioral adaptation for HRI: a survey. *Pattern*

Recognit. Lett. **99**, 3–12. (doi:10.1016/j.patrec.2017.06.002)
6. Burgoon JK, Guerrero LK, Floyd K. 2009 *Nonverbal communication*. Boston, MA: Allyn and Bacon.
7. Biswas M, Murray J. 2016 The effects of cognitive biases in long-term human-robot interactions: case studies using three cognitive biases on Marc the humanoid robot. In *8th Int. Conf. on Social Robotics, Kansas City, MO, 1–3 November 2016* (eds A Agah, J-J Cabibihan, AM Howard, MA Salichs, H Hepp), pp. 148–158. Berlin, Germany: Springer.
8. Mirnig N, Stollnberger G, Miksch M, Stadler S, Giuliani M, Tscheligi M. 2017 To err is robot: how humans assess and act toward an erroneous social robot. *Front. Robot. AI* **4**, 21. (doi:10.3389/frobt.2017.00021)
9. Suvei S-D, Vroon J, Sanchéz VVS, Bodenhagen L, Englebienne G, Krüger N, Evers V. 2018 'I would like to get close to you': making robot personal space invasion less intrusive with a social gaze cue. In *Int. Conf. on Universal Access in Human-Computer Interaction, Las Vegas, NV, 14–20 July* (eds M Antona, C Stephanidis), pp. 366–385. Berlin, Germany: Springer.
10. Celiktutan O, Gunes H. 2015 Computational analysis of human-robot interactions through first-person vision: personality and interaction experience. In *Proc. of 24th IEEE RO-MAN, Kobe, Japan, 31 August–4 September*, pp. 815–820. IEEE.
11. Salam H, Celiktutan O, Hupont I, Gunes H, Chetouani M. 2017 Fully automatic analysis of engagement and its relationship to personality in human-robot interactions. *IEEE Access* **5**, 705–721. (doi:10.1109/ACCESS.2016.2614525)
12. Celiktutan O, Skordos E, Gunes H. 2018 Multimodal human-human-robot interactions (MHHRI) dataset for studying personality and engagement. *IEEE Trans. Affect. Comput.* 1–1. (doi:10.1109/taffc.2017.2737019)
13. Rahbar F, Anzalone SM, Varni G, Zibetti E, Ivaldi S, Chetouani M. 2015 Predicting extraversion from non-verbal features during a face-to-face human-robot interaction. In *Social robotics* (eds A Tapus, E André, JC Martin, F Ferland, M Ammi), pp. 543–553. Lecture Notes in Computer Science, vol. 9388. Berlin, Germany: Springer.
14. Natale L, Nori F, Metta G, Fumagalli M, Ivaldi S, Pattacini U, Randazzo M, Schmitz A, Sandini G. 2013 The icub platform: a tool for studying intrinsically motivated learning. In *Intrinsically motivated learning in natural and artificial systems* (eds G Baldassarre, M Mirolli), pp. 433–458. Berlin, Germany: Springer.
15. Buisine S, Martin JC. 2009 The influence of user's personality and gender on the processing of virtual agents' multimodal behavior. *Adv. Psychol. Res.* **65**, 1–14.
16. Abe K, Hamada Y, Nagai T, Shiomi M, Omori T. 2017 Estimation of child personality for child-robot interaction. In *26th IEEE Int. Symp. on Robot and Human Interactive Communication* (*RO-MAN*)*, Lisbon, Portugal, 28 August–1 September 2017*, pp. 910–915. New York, NY: IEEE.
17. McColl D, Hatakeyama A, Hong N, Nejat G, Benhabib B. 2016 A survey of autonomous human affect detection methods for social robots engaged in natural HRI. *J. Intell. Robot. Syst.* **82**, 101–133. (doi:10.1007/s10846-015-0259-2)
18. Cid F, Prado JA, Bustos P, Nunez P. 2013 A real time and robust facial expression recognition and imitation approach for affective human-robot interaction using gabor filtering. In *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, Tokyo, Japan, 3–7 November, 2013*, pp. 2188–2193. IEEE. (doi:10.1109/IROS.2013.6696662)
19. Ekman P, Friesen WV. 1978 *Facial action coding system: a technique for the measurement of facial movement*. Palo Alto, CA: Consulting Psychologists Press.
20. Cid F, Moreno J, Bustos P, Núnez P. 2014 Muecas: a multi-sensor robotic head for affective human robot interaction and imitation. *Sensors* **14**, 7711–7737. (doi:10.3390/s140507711)
21. Boucenna S, Gaussier P, Andry P, Hafemeister L. 2014 A robot learns the facial expressions recognition and face/non-face discrimination through an imitation game. *Int. J. Soc. Robot.* **6**, 633–652. (doi:10.1007/s12369-014-0245-z)
22. Leo M, Carcagn MD, Distante P, Coco C, Bernava M, Pioggia G, Palestra G. 2015 Automatic emotion recognition in robot-children interaction for ASD treatment. In *IEEE Int. Conf. on Computer Vision Workshop* (*ICCVW*)*, Santiago, Chile, 7–13 December 2015*, pp. 537–545. IEEE. (doi:10.1109/ICCVW.2015.76)
23. Celiktutan O, Gunes H. 2017 Automatic prediction of impressions in time and across varying context: personality, attractiveness and likeability. *IEEE Trans. Affect. Comput.* **8**, 29–42. (doi:10.1109/TAFFC.2015.2513401)
24. Vinciarelli A, Mohammadi G. 2014 A survey of personality computing. *Affect. Comput. IEEE Trans.* **5**, 273–291. (doi:10.1109/TAFFC.2014.2330816)
25. Celiktutan O, Eyben F, Sariyanidi E, Gunes H, Schuller B. 2014 Maptraits 2014: The first audio/visual mapping personality traits challenge. An introduction: perceived personality and social dimensions. In *Proc. 16th Int. Conf. on Multimodal Interaction, Istanbul, Turkey, 12–16 November 2014*, pp. 529–530. New York, NY: ACM. (doi:10.1145/2663204.2668317)
26. Schröder M *et al.* 2012 Building autonomous sensitive artificial listeners. *IEEE Trans. Affect. Comput.* **3**, 165–183. (doi:10.1109/T-AFFC.2011.34)
27. McKown G, Valstar M, Cowie R, Pantic M, Schroder M. 2012 The semaine database: annotated multimodal records of emotionally colored conversations between a person and a limited agent. *IEEE Trans. on Affective Computing* **3**, 5–17. (doi:10.1109/T-AFFC.2011.20)
28. Celiktutan O, Sariyanidi E, Gunes H. 2015 Let me tell you about your personality! Real-time personality prediction from nonverbal behavioural cues. In *11th IEEE Int. Conf. and Workshops on Automatic Face and Gesture Recognition (FG), Ljubljana, Slovenia, 4–8 May 2015*, pp. 1–1. IEEE. (doi:10.1109/FG.2015.7163171)
29. Dautenhahn K. 2014 Human-robot interaction. In *The encyclopedia of human-computer interaction* (eds M Soegaard, DR Friis). Aarhus, Denmark: Interaction Design Foundation.
30. Du S, Martinez AM. 2011 The resolution of facial expressions of emotion. *J. Vis.* **11**, 24–24. (doi:10.1167/11.13.24)
31. Jack RE, Garrod OG, Yu H, Caldara R, Schyns PG. 2012 Facial expressions of emotion are not culturally universal. *Proc. Natl Acad. Sci. USA* **109**, 7241–7244. (doi:10.1073/pnas.1200155109)
32. Nelson NL, Russell JA. 2013 Universality revisited. *Emot. Rev.* **5**, 8–15. (doi:10.1177/1754073912457227)
33. Sariyanidi E, Gunes H, Cavallaro A. 2015 Automatic analysis of facial affect: a survey of registration, representation, and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **37**, 1113–1133. (doi:10.1109/TPAMI.2014.2366127)
34. Ondras J, Celiktutan O, Sariyanidi E, Gunes H. 2017 Automatic replication of teleoperator head movements and facial expressions on a humanoid robot. In *Robot and Human Interactive Communication* (*RO-MAN*)*. 2017 26th IEEE Int. Symp., Lisbon, Portugal, 28 August–1 September*, pp. 745–750. New York, NY: IEEE.
35. Pantic M, Valstar M, Rademaker R, Maat L. 2005 Web-based database for facial expression analysis. In *Proc. IEEE Int. Conf. Multimedia and Expo, Amsterdam, The Netherlands, 6–9 July*, pp. 317–321. New York, NY: IEEE.
36. Littlewort G, Whitehill J, Wu T, Fasel I, Frank M, Movellan J, Bartlett M. 2011 The computer expression recognition toolbox (CERT). In *Proc. IEEE Int. Conf. Automatic Face & Gesture Recognition Workshops, Santa Barbara, CA, 21–25 March*, pp. 298–305. New York, NY: IEEE.
37. Bethel CL, Murphy RR. 2010 Review of human studies methods in HRI and recommendations. *Int. J. Soc. Robot.* **2**, 347–359. (doi:10.1007/s12369-010-0064-9)
38. Kidd CD, Breazeal C. 2005 Human-robot interaction experiments: lessons learned. In *Robot Companions: Hard Problems and Open Challenges in Robot-Human Interaction Symp., Hatfield, UK, 12–15 April*, pp. 141–142. UK: AISB.
39. McCarney R, Warner J, Iliffe S, Van Haselen R, Griffin M, Fisher P. 2007 The Hawthorne effect: a randomised, controlled trial. *BMC Med. Res. Methodol.* **7**, 30. (doi:10.1186/1471-2288-7-30)
40. Miller KW. 2010 It is not nice to fool humans. *IT Professional* **12**, 51–52. (doi:10.1109/MITP.2010.32)