# Cultural Social Signal Interplay with an Expressive Robot

# Cultural Social Signal Interplay with an Expressive Robot

Peter E. McKenna
p.mckenna@hw.ac.uk
Heriot-Watt University
Edinburgh, UK

Ayan Ghosh
ayan.ghosh@sheffield.ac.uk
University of Sheffield
Sheffield, UK

Ruth Aylett
r.s.aylett@hw.ac.uk
Heriot-Watt University
Edinburgh, UK

Frank Broz
f.broz@hw.ac.uk
Heriot-Watt University
Edinburgh, UK

Gnanathusharan Rajendran
t.rajendran@hw.ac.uk
Heriot-Watt University
Edinburgh, UK

## ABSTRACT

Social robots are being developed as a form of social skills training for individual's with an autism spectrum condition (ASC). Effective training will therefore require the social signals produced by a robot to be contingent with people's knowledge and expectations of social cognition and behaviour. Designing recognisable facial expressions is an important part of this challenge; ensuring interactions are more believable and motivating. This design process requires - amongst other factors - consideration of how culture and native language affects social signal processing. In this experiment participants offered a full-bodied robot (named 'Alyx') food items to which Alyx reacted autonomously, producing either an approving or disapproving expression. Participant's responded to these expressions (i.e. the robots social signals) by indicating whether Alyx liked or disliked the food. Task performance was examined both quantitatively (response time and accuracy) and qualitatively (participant's reactionary expressions). The results revealed significant cultural differences, as non-native English speakers were less accurate at interpreting expressions, but also a similar response trend between these groups. Qualitative analysis supported the notion that Alyx's expressions were not universally understood. These findings are discussed in the context of social skills training.

## CCS CONCEPTS

• **Human-centered computing** → **HCI design and evaluation methods**; **Laboratory experiments**; **Scenario-based design**; *User studies*; • **Computer systems organization** → Robotic autonomy;

## KEYWORDS

Human-Robot interaction, social signal processing, autism spectrum disorder, emotion universality


*IVA '18, November 5–8, 2018, Sydney, NSW, Australia*

ACM ISBN 978-1-4503-6013-5/18/11…$15.00
https://doi.org/10.1145/3267851.3267905

**ACM Reference Format:**
Peter E. McKenna, Ayan Ghosh, Ruth Aylett, Frank Broz, and Gnanathusharan Rajendran. 2018. Cultural Social Signal Interplay with an Expressive Robot. In *IVA '18: International Conference on Intelligent Virtual Agents (IVA '18), November 5–8, 2018, Sydney, NSW, Australia.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3267851.3267905

## 1 INTRODUCTION

Designing robot expressive behaviours presents a variety of challenges for engineers, computer scientists, and psychologists. Here, we discuss some of the challenges in designing robot expressive behaviour from a computer science and psychology point of view, as we endeavour to design a robot training buddy capable of delivering social skills training. Specifically, the proposed system will deliver elements of behavioural skills training (BST; a learning theory based training) to improve the social signal recognition of adults with an autism-spectrum condition (ASC). We focus on these adults as is estimated that only 16% in the UK are currently in full-time employment [26] and because of the widely held consensus that improving these adults social skills will positively impact their employability [20]. A potential route to social skills improvement is through the modification of social signals.

Social signals are described as the "...expression of one's attitudes towards social situations and interplay, and they are manifest through a multiplicity of non-verbal behavioural cues, including facial expression, body postures and gestures, and vocal outbursts like laughter" [38], p.1743. In other words, social signals are the subtle verbal and non-verbal signals that are exchanged by conversation partners. Social signal processing is an emerging field in psychology, with initial work proposing a model for autistic social signal processing based on small, detectable differences in multi-modal social-emotional signals (e.g. body postures that convey a person's mood) and communication pragmatics (e.g. timing, turn-taking) [5].

For robot based BST, the challenge is to demonstrate that these signals can be modified through repeated and structured human-robot interaction. Successful modification of a person's social signals will have both low- and high-level benefits. At the low-level, their signal detection will be fine tuned to home in on information that is contextually relevant. For example, successful employment programmes for adults with an ASC often adopt training schedules that train low-level signals, such as recognition of typical greeting and farewell behaviours. Training low-level signal process can

then lead to higher-level benefits, like successful social integration [11, 13].

Before developing this robot-based training, we first studied the social signal interplay between typically developing participant's and a robot when the robot expressed a change in state: through approval and disapproval facial expressions. In the task participants interacted with a full-bodied FLASH-MKII robot with an EMYS head (https://emys.co/). Cultural background was included in this research given the likelihood that our future target ASC audience will have a diverse heritage. We assessed their interpretation of the robot's expressions and examined their voluntary facial responses to the robot's change in expression during a novel, task based interaction.

## 1.1 Designing and testing robot expressive behaviour

Robot facial expressions clearly depend on the embodiment of the robot: not all robots do have anything approximating to a face and must use other expressive modes [10], while at the other extreme some robots are designed with faces intended to be as close to a human one as feasible from an engineering point of view [1], though this approach is known to run the risk of 'uncanny valley' effects [33].

Early work in designing such faces is covered in [10] and since then many approaches have been tried, whether humanoid to varying degrees of naturalism [1], [31], intentionally 'robotic' or machine-like, animal-like (iCat [37], SAM [17]), or more abstract in design (MiRAE [2]).

There is no consensus as yet on which of these approaches works better with human interaction partners with an ASD [32] and many variants have been used, though almost always with children. There is almost no work with adults with an ASC, though relatedly, adults with Asperger's show a preference for text chat role play over face-to-face communication [30] suggesting that robots could leverage these adults affinity for technology-based interaction.

The EMYS robot head used in this study falls into a cartoon-like more machine-oriented category [18], as we follow the argument that a simplified set of expressive features helps to counter the known issues of sensory over-stimulation [16] for those with an ASC (e.g. see [24]). We were also careful to experiment using a methodology appreciative of social signal processing.

Evaluation of expressive facial behaviour in both graphical and robotic agents has often been related both directly to emotion recognition and less directly to perceptions of 'life-likeness'. Emotion recognition approaches have often been based on Eckman's six 'primitive emotions' [8], influenced by his finding that a small set of static facial expressions labelled as emotions were recognized across cultures. In fact, most human-robot interaction research of expression (or emotion) recognition achieve human-like face configurations by mapping the robot's degrees of freedom (DOF) with an analogous human facial muscle (or action units; AUs); a task that requires some aggregation given the numerical and positional difference between human and robot AUs.

Two recent examples [2] [17] have taken this approach. In one case [2] required adult participants to directly categorize a robot with schematic expressions using a fixed set of labels, and the other [17] asked neuro-typical children to match a cartoon-like monkey robot face with human photographs.

This approach is not suitable for the work reported for two reasons. The first is that dramatic and stereotypical emotional expressions are very rarely found in adult workplaces; the context for our proposed training. The second is that we do not seek context-independent recognition: a social signal is not merely or even necessarily an emotional expression, but is a contextual response. Thus, testing Alyx's facial expressions out of context is unlikely to be useful.

The second approach commonly taken is more holistic and looks at post-hoc subjective assessments of social agent attributes. A participant interacts with an agent and is then asked to assess it, whether by questionnaire, interview, focus group, or some combination. In a graphical character example [3], participants watched the character giving a presentation, and then were asked about attributes such as trustworthiness, like-ability and degree of expressiveness, as well as being tested for recall of the presentation.

While this approach does take contextual factors into account, its post-hoc nature is also unsuitable for the current work, in which identification of specific social signals in specific contexts is required. An approach that does meet these requirements can be found in work investigating negotiation between a human participant and a graphic character [7]. Here the character uses facial expressions indicative of happiness and anger in appropriate negotiation contexts and this is shown to have similar impact on the behaviour of the human participant as in human-human negotiation, thus demonstrating the recognition of in-context social signals.

So, in the present experiment, participants responses were part of an ongoing interaction with Alyx, thereby giving context to their choices and behaviour. That is, both their expression recognition and facial responses were a direct response to the social signals elicited by the robot.

An additional consideration to this experiment was an individual's cultural background and native language. As will be discussed below, a person's culture shapes their interpretation of social signals, as shown by studies of both human and robot facial recognition. As such, ASC could be considered as a distinct 'culture', as studies of the conditions often highlight a unique processing style relative to typical development.

## 1.2 Cultural, native language, and expression recognition

A culture is a set of morals, values, and thought processes held by members of a group that are passed down and maintained by cultural predecessors [19]. Cultural distinctness therefore is a consequence of these values development and longevity, leading to specific 'value orientations'. Traditionally these orientations have been classified into two types: *individualist*, where emphasis on an individual's autonomy and self-directed achievement; *collectivist*, where individual's are viewed as part of an integrated system that thrives on cooperation [12, 27].

There is a significant research effort questioning how these value orientations affect the interpretation and use emotion. For example, Matsumoto and colleagues contend that individualist cultures use

emotion to guide behaviour, and establish independence, whereas in-group norms drive behaviour in collectivist cultures [19]. As such, countries with an Eastern culture are often viewed as collectivist, as they favour emotional moderation and avoid the use of intense emotional responses [21, 36].

Studies of agent and robot faces offer a unique window to this debate, as the parameters of facial expressions can be manipulated to identify the psychological processes that underpin perceivers expression recognition. Recent work using dynamically generated agent faces supports uniqueness over universality [15], showing that Westerners attended to the region of the mouth more so when observing expressions than their Eastern Asian counterparts, who focused more on the eye region [40]. Further support for these unique emotion processing styles was found when Asian and Western participants assessed the expressive behaviour of a full-bodied robot. In [35], Asian participants perceived the eyebrows of the robot as a salient feature in their decision making, more so than their Western counterparts. Further, an in-group preference was evident, as both groups rated expressions typical to their culture more favourably.

Despite a growing body of literature on the cultural uniqueness of emotion, a *universality* hypothesis continues to be debated subject owing to the original arguments put forward by Darwin - that humans and animal share primal expressions [6] - and modern research techniques demonstrating that emotions are understood equally well between different cultures [41]. The issue is complex and clearly requires empirical investigation to resolve. Here, we contribute indirectly to this literature by investigating how native language - a cultural component - affected expressive behaviour interpretation.

Studies of native language and emotion have focussed on the interpretation of words of multi- or bi-linguals native or non-native languages, as well as the recognition of emotion expressions. This will be an important factor in the language used by agents for therapy, especially those intended to elicit an emotional response as part of training. Seemingly, native words in advertising slogans are rate as more emotional than non-native words [28] by bi-linguals, and a similar preference has been shown for emotional words and phrases [4]. These differences has been attributed to the language spoken at home and school during development, as children will associate emotional experiences with their dominant language [4]. However, recent work using an emotion based face-word Stroop task found equivalent performance trends between emotional faces paired with native and non-native word descriptors [9]. This suggests that the cognitive processes underlying non-native language processing are related to the faculties responsible for expression recognition. Thus, Alyx's use of spoken English be benign in the context of emotion recognition.

So, work on culture and native language suggest that lens through which emotion is perceived is mediated by cultural background, but that there may be some expressive behaviour that is also universally recognisable. Thus, studies of cultural and native language influences on emotion are somewhat equivocal

## 1.3 Expressive behaviour mimicry

In an interaction, the facial movements of both interlocutors helps to establish common ground, for example, by allowing each other to anticipate the goal of the conversation, and each other's intentions [39].

Support for shared understanding between faces has been shown by studies that directly manipulate participant's face musculature. Expression stimuli are rated more positively when smiling, and the opposite when frowning [14]. Also, prohibiting participant's facial movements entirely reduces their expression recognition accuracy [34].

Though studies of humans propensity to spontaneously imitate the expressive behaviour of robots is lacking, research with children shows that, even from a young age there is a tendency to copy and imitate a robot's behaviours [25].

Thus, mimicry was also of interest here, as Alyx's social signals could conceivably match that of participants: that each liked/disliked the food on offer and would share this evaluation with their facial expression. Evidence of mimicry therefore would demonstrate a shared experience between Alyx and participants and the potential for robots to improve certain elements of social cognition (i.e. theory of mind).

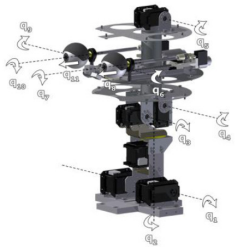
## 1.4 Hypotheses

The present study examined how native language affected the interpretation of an expressive robot's social signals (i.e. facial expressions). We recognise that native language is an indirect classifier of culture - English a first language in several different cultures - but hope to add to the discussion by offering insight from a language perspective: insight that will be especially useful for speaking robots. Our robot spoke English, so our hypothesis were as follows:

*H1*: Native English participants will identify the robot's expressions more accurately and faster because of an increased emotional salience of the robot's language.

*H2*: However, both groups will also present a similar response trend, in terms of their accuracy and speed, given the universality of emotion.

*H3*: As Eastern cultures tend to be more collectivist and less expressive, we expected Native English speakers to show more mimicry and facial reactions than non-native English speakers.



| Joint | Joint name |
|---|---|
| q1 | Neck pitch (rotation) |
| q2 | Head yaw (rotation) |
| q3 | Head pitch (rotation) |
| q4 | Lower disc (rotation) |
| q5 | Upper disc (rotation) |
| q6 | Left eyelid (rotation) |
| q7 | Left eyebrow (rotation) |
| q8 | Left eye (translation) |
| q9 | Right eyelid (rotation) |
| q10 | Right eyebrow (rotation) |
| q11 | Right eye (translation) |

**Figure 1: EMYS head 11 degrees of freedom (DOF).**

# 2 METHOD

## 2.1 Facial expressions of Alyx

The EMYS robot head used in this work has minimalist facial features with only 11 degrees of freedom (DOFs) as shown in Figure 1. This contrasts with the much higher number of degrees of freedom on the human face.

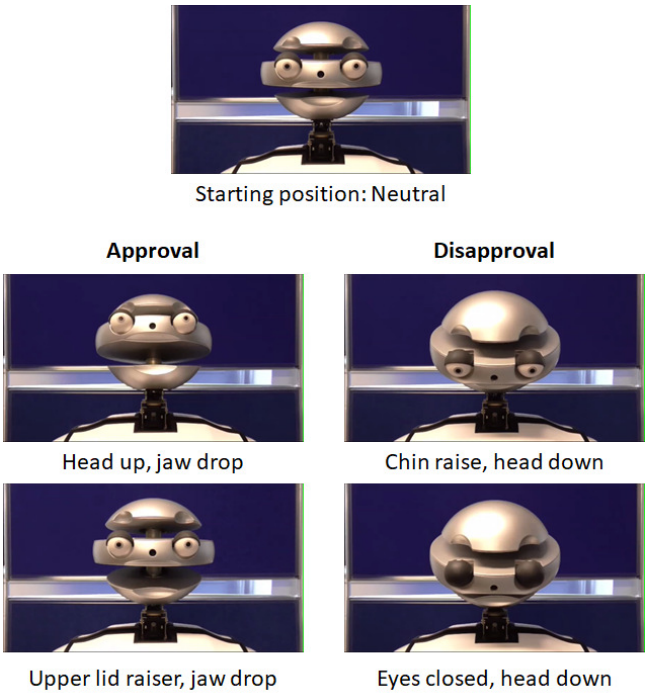



**Figure 2: Alyx's dynamic expressive behaviour. Approval expressions: Head up, jaw drop; Upper lid raiser, jaw drop. Disapproval expressions: Chin raise, head down; Eye's closed, head down.**

Expressions were designed bottom-up by first linking EMYS's DOFs to the single facial movements defined by the Facial Action Coding System (FACS) [8], then matching them against affective states in the dimensional model Pleasure-Arousal-Dominance (PAD) [23]; for full details see [22]. An *in the wild* evaluation of these expressions found that two approval expressions (head up, jaw drop and upper lid raiser, jaw drop) were viewed positively at above chance level [22]. Here, adopted the framework of [22] using the least ambiguous expressions to empirically test other dimensions of participant interpretation (see Figure 2).

## 2.2 Participants

Fifty-seven university staff and students (Mean age = 25.84, SD = 8.60) participated, including 24 females and 33 males. The sample had a mixed demographic; 72% White, 19% Asian, 3.5% mixed, 3.5% Arab, and 2% African. Of the sample, 34 were native English speakers and 22 non-Native English speakers. Participants were recruited via lectures, university advertising boards, posters, and an internal news article. Participation criteria included being over the age of 18 years, having no diagnosed psychiatric condition, and normal or corrected vision. Participants provided written consent and were entered into a voucher prize draw upon completion. One participant did not understand the requirements of the experiment and so their data was omitted from further analysis (data loss 1.75%). The experiment was given ethical approval by the Heriot-Watt University School of Social Sciences: ethical code 2017-516.

## 2.3 Materials

*Robot setup.*

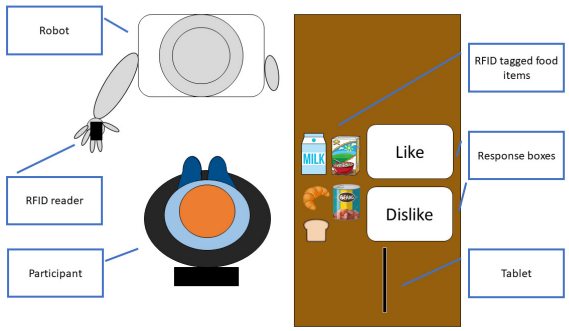



**Figure 3: Expression recognition task setup. RFID reader strapped to robots right hand. RFID tagged items on table.**

Participants were tested individually at Heriot-Watt University's Robotarium HRI lab. Two large poster screens sectioned off the testing area, occluding potential visual distractions. Participants sat on a adjustable office chair facing the robot. A scripted interaction by the robot was written in URBI. Execution and collection of data was performed using GOSTAI software. Other than the head and the right arm the robot was static for the experiment. An RFID reader was attached to the robots right hand with elastic. A desk was positioned opposite participants, where the plastic food items and two Tupperware boxes marked 'Like' and 'Dislike' were placed. Each food item had an RFID tag attached (see Figure 3).

## 2.4 Design

The study adopted a mixed-design, whereby participants viewed all four of the robot expressions but were split according to native language. The order expression was counterbalanced to reduce the risk of ordering effects. Analysis variables included native language (English, non-native English), expression (Head up, jaw drop; Upper lid raiser, jaw drop; Chin raise, head down; Head down, eyes closed) and Likert scale responses to the post-interaction questionnaire items.

Recognition accuracy (Correct, Incorrect) was based on our categories of approval and disapproval: a participant providing a correct answer would place the food item in the 'Like' box for expressions *Head up, jaw drop* and *Upper lid raiser, jaw drop*, and to place the item in the 'Dislike' box for the expressions *Chin raise, head down* and *Eyes closed, head down*. All other participant responses were marked as incorrect.

## 2.5 Procedure

Participants faced the robot and adjusted the seat to be at the robot's eye-level. The experimenter explained that they were to offer food items to the robot and that Alyx would respond to their offerings with a facial expression. The full experimental procedure was explained meticulously by the experimenter as pilot work demonstrated that verbal instructions from the robot alone did not adequately inform participants of the tasks requirements.

Participants were asked to place each food item near the robots right hand with the RFID tag facing the reader, one at a time when the robot offered its hand. A beep indicated that the RFID tag had been registered by the reader. After the beep, the robot glanced at its hand, turned to face the participant, lowered its right hand, and produced an expression (duration 2 sec per expression). Importantly, participants were told to keep hold of the object and to maintain attention on the robot's facial expression after it had lowered its hand; pilot work found that occasionally participants interpreted Alyx's glance toward it's hand as an expression, or look away from the robot during expression generation. Participants responded to expressions by placing the food item in the 'Like' or 'Dislike' box on the table next to them. That is, responses were made in the context of food offering - participant's choice of box was based on the social signals elicited by Alyx, showing that it liked or disliked what was on offer. After giving these instructions Alyx then repeated each step in English with synchronous movements of the mouth.

Once the task was complete participants then filled out a short post-interaction evaluation on a tablet. Each item in the evaluation was presented along a 5-point Likert scale: 1) Alyx was... (Friendly ... Unfriendly); 2) Alyx was pleased with me (No, I disagree ... Yes, I agree); 3) I liked Alyx (Not at all ... Very much); 4) Alyx's voice was (Hard to understand ... Easy to understand); 5) I think I did (Very badly ... Very well); 6) I liked interacting with Alyx (No, I disagree ... Yes, I agree). For ease, these items are henceforth referred to as 1) Friendliness, 2) Perceived positiveness, 3) Likeability, 4) Performance rating, 5) Voice clarity, and 6) Interaction rating.

# 3 RESULTS

## 3.1 Modelling performance from native language group

Data were analysed using binary logistic regression with group membership of native English ($n$ = 34; coded as 1) and non-native English ($n$ = 22; coded as 0) entered as the outcome variable. Stepwise analysis testing the AIC between model iterations revealed that response accuracy, response time, as well as the questionnaire items pertaining to Friendliness, Likeability, Voice clarity, and Interaction rating were the strongest predictors of group membership. There was no significant interaction between them. Results of the regression are presented in Table 2.

As a first step, we tested the model's predictive power against the null variance model. The results of this analyses was significant, $\chi^2$ (6) = 59.837, $p < 0.001$, Nagelkerke $R^2 = 0.324$ (a moderate effect size). Overall, 85.58% of participants were correctly classified, with better classification in the non-native English group (85.07%) compared to the native English group (58.82%).

**Table 1: Response Accuracy (% Correct) According to Robot Expression and Native language.**

| EMYS AUs | Native English | Non-native English | Overall |
|---|---|---|---|
| *Approval* | | | |
| Head up, jaw drop | 76.47% | 68.18% | 71.93% |
| Upper lid raiser, jaw drop | 94.11% | 72.72% | 85.96% |
| *Disapproval* | | | |
| Chin raise, head down | 88.24% | 63.63% | 78.94% |
| Eyes closed, head down | 94.12% | 86.36% | 89.47% |
| Overall | 88.24% | 72.72% | 81.58% |

**Table 2: Exponent B and significance values of optimal model predictors.**

| | *B* | *S.E.* | *Low CI* | *Upp CI* | *Wald* |
|---|---|---|---|---|---|
| Intercept | 5.45 | 1.373 | 0.410 | 94.061 | 1.235 |
| Response accuracy | 0.378 | 0.436 | 0.157 | 0.877 | -2.227* |
| Response time | 0.922 | 0.041 | 0.849 | 1.001 | -1.955. |
| Friendliness | 2.912 | 0.258 | 1.797 | 4.954 | 4.144*** |
| Likeability | 0.342 | 0.345 | 0.075 | 0.292 | -5.382*** |
| Voice clarity | 1.174 | 0.239 | 0.724 | 1.867 | 0.672 |
| Interaction rating | 1.652 | 0.231 | 1.058 | 2.637 | 2.163*** |

$Signif.codes$ : '.' $p < 0.1$. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

The results showed a number of group differences between native and non-native English participants. Firstly, participants in the Native English group were significantly less likely to provide incorrect responses, $B$ = 0.378, $p < 0.05$. This group also rated the robot as significantly less likeable to non-native participants, $B$ = 0.342, $p < 0.001$. Lastly, there was a near significant effect of response time, whereby native English speakers spent less time interpreting expressive behaviour, $B$ = 0.922, $p$ = 0.051.

From the non-native speakers results it was evident they rated Aylx as significantly more friendly than the native speakers, $B$ = 2.912, $p < 0.001$, and also rated the overall interaction more favourably, $B$ = 1.652, $p < 0.001$. See Figure 4 for means and confidence intervals between groups on each questionnaire item.

## 3.2 Group performance similarity

As well as testing the differences between participants interpretations of Alyx's expressive behaviour we also examined the similarities in responses between the native and non-native English speakers. To do so, a Kendall's correlation was run between the means for each groups accuracy, response time, Friendliness, Likeability, Perceived Positiveness, Voice Clarity, Performance Rating and Interaction ratings. This analysis yielded a significant association between the two groups ($\tau_b$ = 0.786, $p < 0.05$), showing that, on

**Table 3: Participant expressive behaviour (% of trials) according to native language**

| Category | Native English | Non-native English | $\chi$ |
|---|---|---|---|
| Neutral | 57.35% | 51.76% | $21.214^{***}$ |
| Positive | 14.71% | 17.65% | 0.703 |
| Negative | 7.36% | 1.18% | $12.410^{***}$ |
| Not sure | 20.59% | 29.41% | $6.172^{*}$ |

$Signif.codes: `.' p < 0.1. {}^{*}p < 0.05. {}^{**}p < 0.01. {}^{***}p < 0.001.$

the whole both groups evaluated the expressions and interaction similarly.

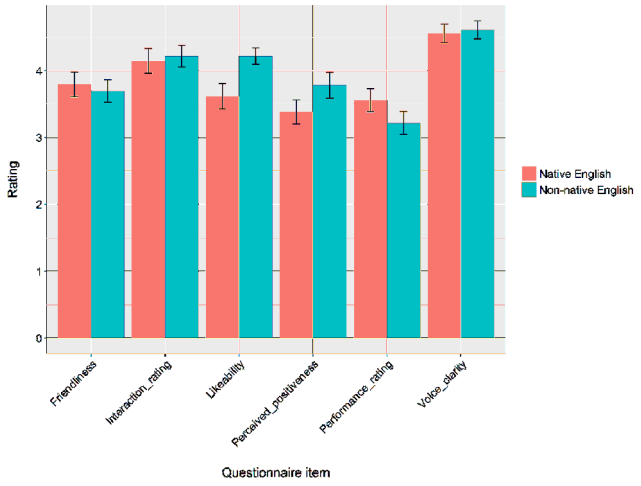



**Figure 4: Mean and confidence intervals of post-interaction questionnaire responses. Item responses were given along a 5-point Likert scale.**

## 3.3 Participant facial responses

Also of interest was the extent that participant's mimicked the facial expression of the robot having observed Alyx's reaction to the food item. However, preliminary data analysis revealed a low level of expressive correspondence (below 5%) between participants and Alyx for all expressions. As an alternative, we scrutinised the video footage for changes in participant's expression in response to Alyx's change in expression.

To obtain facial response data, video footage of participants was first reviewed and annotated in ELAN in accordance to the Facial Action Unit Coding System (FACS; [8]). A set of four descriptive facial categories was generated: Neutral, Positive, Negative, and Not Sure; see Figure 5. The experimenter examined the footage during and in the period following Alyx's expressive behaviour. Changes to participant's facial AUs were cross-referenced against the FACS action unit taxonomy. A second experimenter then reviewed the footage independently for each trial. Disputable expressions were analysed further until both experimenters came to a 100% agreement.

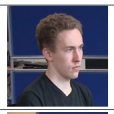
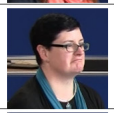
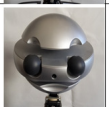
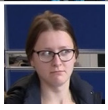
| Robot Expression | Participant Facial Response | Participant Facial Action Units (Ekman, 1978) | Descriptive Category | Percent of trials |
|---|---|---|---|---|
|  |  | N/A | Neutral | 55% |
|  |  | AU6 Cheek raiser<br>AU 12 Lip corner puller<br>AU13 Cheek puffer | Positive | 15.84% |
|  |  | AU9 Nose Wrinkler<br>AU15 Lip Corner Drepressor<br>AU17 Chin Raiser | Negative | 5% |
|  |  | AU4 Brow Lowerer<br>AU44 Squint<br>AU14 Dimpler<br>AU23 Lip Tightener<br>AU25 Lips Part | Not sure | 23.95% |

**Figure 5: Qualitative assessment of participant expressive behaviour immediately following robot expression.**

To determine if this propensity to spontaneously produce a facial expression was mediated by native language (English, non-native English) a series of binomial proportion tests were conducted for each expressive category identified from the initial qualitative analysis (see Figure 5).

The result of these binomial proportion tests (a form of Chi-square analysis) are presented in Table 3. We found that Native English speakers produced significantly more 'neutral' expressions, $\chi^2$ (1) = 21.214, $p < 0.001$, and 'negative' expressions, $\chi^2$ (1) = 12.410, $p < 0.001$ relative to their non-native English speaking counterparts. Participants in the non-native English group produced more 'not sure' expressions, $\chi^2$ (1) = 6.172, $p < 0.05$.

## 4 DISCUSSION

Designing interpretable expressive behaviours of robots is a task that requires careful consideration of the participant's native language, their cultural background, the robot's appearance, and the nature of the interaction. Here, we attempted to address these issues by designing and testing contextually relevant expressive behaviour (i.e. during an interaction) with a low DOF robot head, and by considering how native language affects expression interpretation.

The value of adopting a social signal recognition approach is that our expressive behaviour was evaluated as part of an ongoing interaction, rather than in a de-contextualised or post-hoc manner. As such, the behaviour elicited by the robot is in direct response to a persons actions, giving greater meaning to the interaction as a whole. The majority of responses (81.57%) were correct, further validating our *approval* and *disapproval* expressive behaviours. Also, evaluations of the interaction in the questionnaire were very positive, indicating that they enjoyed the food offering interaction, despite the robot producing a disapproval expression on 50% of trials. This result is important in the context of creating a social skills training robot, as the final system will need to produce interpretable expressions as well as engage users positively for role playing workplace social scenarios.

It is likely a sample of adults with an ASC will include individual's from different cultural backgrounds with different native languages. So, as well as adopting a social signals methodology we considered the impact that native language would have on expression recognition. Drawing on the literature of culture and emotion, and native language, we predicted that there would be some group differences in accuracy, but on the whole, both groups would evaluate expressions and the interaction similarly given the universality of emotion.

In line with our expectations, native English speakers overall showed better accuracy of the four expressions. This could be explained by both native language and culture. Given the indication that emotional native words are favoured over non-native emotional words [4, 28], and that Alyx told participants in English that they would be indicating whether it "...liked or disliked the food", it is possible that native English participants were more perceptually attuned to Alyx's expressions. That is, both the language used by the experimenter and the robot primed participants emotion recognition and they were able to judge the expressive more accurately. However, future work should specifically investigate how the language of a robot modifies the signal recognition of the participant. For example, the robot could give instructions using both Western and Eastern typical emotional words, and ask that items be placed in certain locations (e.g. to their left or right) to avoid semantic priming.

Differences in accuracy could also be explained by the value orientation of participants. Many Western societies support an individualist value orientation, where emotions are important for guiding behaviour [27]. As such, Westerners show greater sensitivity to emotional expressivity - particularly positive emotions - in faces, whereas participant's from Eastern cultures traditionally rely less on emotion and more on in-group norms [19]. If it follows that most Western cultures speak English as their main language, the findings suggest that expressive behaviour as a response to offering food is more relevant in individualist cultures, where information about another's mental state provides a reference for behaviour. In support of this idea, non-native English speakers showed fewer negative facial responses and a greater number of uncertain facial responses, indicating Alyx's social signals did not offer additional information to guide future action. As such, social skills trainings for Eastern cultures will have a different strategy that match in-group expectations. An example would be a collaborative exercise with a robot that helps to establish a strong bond with the participant. Once this bond is made clear the emotional responses could be made incrementally.

Whilst we highlighted differences that occur between people of a different native language (e.g. in response accuracy) we also investigated whether the performance trend between native and non-native English speakers was similar. This analysis was conducted to question the emotion universality hypothesis [6, 41], work indicating that native language does not affect emotion evaluation [9], and to give a more general overview of participant's response behaviour - above scrutiny of specific differences. In a manner similar to [29], we performed a correlation analysis between the means between of each of the outcome variables. This analysis found a strong association, showing that response strategy was similar between non-native and native English speakers. The result also demonstrates that the expressive behaviours themselves were recognisable as either approving or disapproving regardless of native language, supporting the second hypothesis. However, it should be noted that universality in this instance only applies to the expressions examined here - that our robot's expressions of approval and disapproval were understood in the context of an ongoing interaction. We chose these two expressions because of their frequency in the workplace. It is reassuring to know that people of a different native language would understand if a robot looks pleased or upset at an outcome. The next step is to determine if these expressive behaviours provide useful social signal training for peoples with social skill difficulties (e.g. adults with an ASC), and if this training leads to generalisation outside of the laboratory setting.

For hypothesis three, From an initial qualitative assessment of facial AUs, four common (and from our assessment, objective) expressive categories were generated: neutral, positive, negative, and not sure. Of these, neutral and positive were the most easy to spot; neutral expressions did not present and change in AU state, and positive expressive behaviour was more salient. For negative and not sure responses, a change of state was clear, but whether it could be categorised as such proved difficult to ascertain.

Each change of state was carefully examined both from static photographs and video clips multiple times between two researchers. This was the best approach to take at the time. Results from group comparisons were mixed, with native English speakers showing a little more positive expressivity, but also more inactivity (greater number of neutral faced responses). That the non-native group produced a greater number of 'not sure' expressions substantiates our claim that this group found the expressive behaviour more ambiguous. Future work should expand from the approach adopted here, to include more expressive behaviours and social signal types. Using approval and disapproval behaviours likely limited the scope for mimicry by participants. Although suitable for our purposes, it could be that robots with more DOF (like KOBIAN-R) designed to produce a range of expressive behaviours would have greater success eliciting social cognitive phenomena, like facial mimicry. This is especially important given the indication that different cultures process facial expressions with unique strategies, like attending to the mouth more than the eyes, and vice-versa [40].

Together, the findings highlight a number of important features concerning social signal recognition for robot-based social skills training in this domain. Firstly, that native language is an important factor in the design of robot expressive behaivour, as participants in the non-native English group found the expressive behaivour of our English speaking robot more ambiguous. This may be in part due to the language used by the robot and the individual's cultural background. If robot's are to be deployed successfully in human environments they will need to accommodate the customs of different cultures. Added functionality could include the option for users to state their cultural background, initiating a separate script in the robot that is equipped with that particular culture's practices. On the other hand, participant's response trend was similar between groups, indicating that robot's can erode some of the human boundaries created by culture. So perhaps it would be more pertinent to explore how robot's can create their own cultural

practice, to guide users interactions over pre-conceived notions of culture.

Our desire is to design a robot for social skills training, by modifying the social signals of the user, to recognise socially salient cues during an interaction. Here, we demonstrated that this is possible in a simple food offering exchange between people of different native languages. The next challenge is to examine the systems effectiveness in a more complex, work-like context.

## REFERENCES


[1] Christian Becker-Asano and Hiroshi Ishiguro. 2011. Evaluating facial displays of emotion for the android robot Geminoid F. In *IEEE SSCI 2011 - Symposium Series on Computational Intelligence - WACI 2011: 2011 Workshop on Affective Computational Intelligence*. DOI: http://dx.doi.org/10.1109/WACI.2011.5953147

[2] Casey C. Bennett and Selma Šabanović. 2014. Deriving Minimal Features for Human-Like Facial Expressions in Robotic Faces. *International Journal of Social Robotics* (2014). DOI: http://dx.doi.org/10.1007/s12369-014-0237-z

[3] StÃĽphanie Buisine, S Abrilian, and Jean Claude Martin. 2004. Evaluation of multimodal behaviour of embodied agents - Cooperation between speech and gestures. (2004).

[4] Catherine L. Caldwell-Harris. 2015. Emotionality Differences Between a Native and Foreign Language: Implications for Everyday Life. *Current Directions in Psychological Science* (2015). DOI: http://dx.doi.org/10.1177/0963721414566268

[5] Laurence Chaby, Mohamed Chetouani, Monique Plaza, and David Cohen. 2012. Exploring multimodal social-emotional behaviors in autism spectrum disorders: An interface between social signal processing and psychopathology. In *Proceedings - 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust and 2012 ASE/IEEE International Conference on Social Computing, SocialCom/PASSAT 2012*. DOI: http://dx.doi.org/10.1109/SocialCom-PASSAT.2012.111

[6] Charles Darwin. 1872. The expression of the emotions in man and animals. *The American Journal of the Medical Sciences* (1872). DOI: http://dx.doi.org/10.1097/00000441-195610000-00024

[7] Celso De Melo and Ana Paiva. 2006. Multimodal expression in virtual humans. In *Computer Animation and Virtual Worlds*. DOI: http://dx.doi.org/10.1002/cav.127

[8] Paul Ekman and Wallance V. Friesen. 1978. The Facial Action Coding System. *Consulting* (1978).

[9] Lin Fan, Qiang Xu, Xiaoxi Wang, Fei Xu, Yaping Yang, and Zhi Lu. 2018. The automatic activation of emotion words measured using the emotional face-word Stroop task in late ChineseâĂŞEnglish bilinguals. *Cognition and Emotion* (2018). DOI: http://dx.doi.org/10.1080/02699931.2017.1303451

[10] Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. 2003. A survey of socially interactive robots. In *Robotics and Autonomous Systems*. DOI: http://dx.doi.org/10.1016/S0921-8890(02)00372-X

[11] A. Hillier, H. Campbell, K. Mastriani, M. V. Izzo, A. K. Kool-Tucker, L. Cherry, and D. Q. Beversdorf. 2007. Two-Year Evaluation of a Vocational Support Program for Adults on the Autism Spectrum. *Career Development and Transition for Exceptional Individuals* (2007). DOI: http://dx.doi.org/10.1177/08857288070300010501

[12] Geert Hofstede. 1980. Culture and Organizations. *International Studies of Management & Organization* (1980). DOI: http://dx.doi.org/10.1080/00208825.1980.11656300

[13] P. Howlin. 2000. Outcome in Adult Life for more Able Individuals with Autism or Asperger Syndrome. *Autism* (2000). DOI: http://dx.doi.org/10.1177/1362361300004001005

[14] Sylwia Hyniewska and Wataru Sato. 2015. Facial feedback affects valence judgments of dynamic and static emotional expressions. *Frontiers in Psychology* (2015). DOI: http://dx.doi.org/10.3389/fpsyg.2015.00291

[15] Rachael E. Jack and Philippe G. Schyns. 2017. Toward a Social Psychophysics of Face Communication. *Annual Review of Psychology* (2017). DOI: http://dx.doi.org/10.1146/annurev-psych-010416-044242

[16] David O. Johnson, Raymond H. Cuijpers, and David van der Pol. 2013. Imitating Human Emotions with Artificial Facial Expressions. *International Journal of Social Robotics* (2013). DOI: http://dx.doi.org/10.1007/s12369-013-0211-1

[17] Sarah A. Koch, Carl E. Stevens, Christian D. Clesi, Jenna B. Lebersfeld, Alyssa G. Sellers, Myriah E. McNew, Fred J. Biasini, Franklin R. Amthor, and Maria I. Hopkins. 2017. A Feasibility Study Evaluating the Emotionally Expressive Robot SAM. *International Journal of Social Robotics* (2017). DOI: http://dx.doi.org/10.1007/s12369-017-0419-6

[18] Jan KÈĺdzierski, Robert Muszyński, Carsten Zoll, Adam Oleksy, and Mirela Frontkiewicz. 2013. EMYS-Emotive Head of a Social Robot. *International Journal of Social Robotics* (2013). DOI: http://dx.doi.org/10.1007/s12369-013-0183-1

[19] David Matsumoto, Seung Hee Yoo, and Sanae Nakagawa. 2008. Culture, Emotion Regulation, and Adjustment. *Journal of Personality and Social Psychology* (2008). DOI: http://dx.doi.org/10.1037/0022-3514.94.6.925

[20] Ifigeneia Mavranezouli, Odette Megnin-Viggars, Nadir Cheema, Patricia Howlin, Simon Baron-Cohen, and Stephen Pilling. 2014. The cost-effectiveness of supported employment for adults with autism in the United Kingdom. *Autism* 18, 8 (2014). DOI: http://dx.doi.org/10.1177/1362361313505720

[21] Daniel McDuff, Jeffrey M. Girard, and Rana el Kaliouby. 2017. Large-Scale Observational Evidence of Cross-Cultural Differences in Facial Behavior. *Journal of Nonverbal Behavior* (2017). DOI: http://dx.doi.org/10.1007/s10919-016-0244-x

[22] McKenna Peter E, Lim Mei Yii, Ghosh Ayan, Aylett Ruth, Broz Frank, and Gnanathusharan Rajendran. 2017. Do you think I approve of that? Designing facial expressions for a robot. In *International Conference of Social Robotics (ICSR)*. Ninth International Conference of Social Robotics (ICSR): Embodied Interactive Robotics, Tsukuba, Japan.

[23] Albert Mehrabian. 1996. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in Temperament. *Current Psychology* (1996). DOI: http://dx.doi.org/10.1007/BF02686918

[24] Laurent Mottron, Michelle Dawson, Isabelle Soulières, Benedicte Hubert, and Jake Burack. 2006. Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception. (2006). DOI: http://dx.doi.org/10.1007/s10803-005-0040-7

[25] B Mutlu and J Forlizzi. 2008. Robots in organizations: {The} role of workflow, social, and environmental factors in human-robot interaction. In *2008 3rd {ACM}/{IEEE} {International} {Conference} on {Human}-{Robot} {Interaction} ({HRI})*. DOI: http://dx.doi.org/10.1145/1349822.1349860

[26] National Autisitic Society. 2016. *The autism employment gap: Too Much Information in the workplace*. Technical Report. National Autistic Society. http://www.autism.org.uk/about/what-is/myths-facts-stats.aspx

[27] Daphna Oyserman, Heather M. Coon, and Markus Kemmelmeier. 2002. Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin* (2002). DOI: http://dx.doi.org/10.1037//0033-2909.128.1.3

[28] Stefano Puntoni, Bart de Langhe, and Stijn M. J. van Osselaer. 2009. Bilingualism and the Emotional Intensity of Advertising Language. *Journal of Consumer Research* (2009). DOI: http://dx.doi.org/10.1086/595022

[29] Chao Qu, Willem Paul Brinkman, Yun Ling, Pascal Wiggers, and Ingrid Heynderickx. 2013. Human perception of a conversational virtual human: An empirical study on the effect of emotion and culture. *Virtual Reality* (2013). DOI: http://dx.doi.org/10.1007/s10055-013-0231-z

[30] Gnanathusharan Rajendran and Peter Mitchell. 2006. Text Chat as a Tool for Referential Questioning in Asperger Syndrome. *Journal of Speech Language and Hearing Research* (2006). DOI: http://dx.doi.org/10.1044/1092-4388(2006/008)

[31] Ben Robins, Kerstin Dautenhahn, and Janek Dubowski. 2006. Does appearance matter in the interaction of children with autism with a humanoid robot? *Interaction Studies* (2006). DOI: http://dx.doi.org/10.1075/is.7.3.16rob

[32] Brian Scassellati, Henny Admoni, and Maja Matarić. 2012. Robots for Use in Autism Research. *Annual Review of Biomedical Engineering* (2012). DOI: http://dx.doi.org/10.1146/annurev-bioeng-071811-150036

[33] Jun'ichiro Seyama and Ruth S. Nagayama. 2007. The Uncanny Valley: Effect of Realism on the Impression of Artificial Human Faces. *Presence: Teleoperators and Virtual Environments* (2007). DOI: http://dx.doi.org/10.1162/pres.16.4.337

[34] MariÃńlle Stel and Ad Van Knippenberg. 2008. The role of facial mimicry in the recognition of affect. *Psychological Science* (2008). DOI: http://dx.doi.org/10.1111/j.1467-9280.2008.02188.x

[35] Gabriele Trovato, Tatsuhiro Kishi, Nobutsuna Endo, Kenji Hashimoto, and Atsuo Takanishi. 2012. A cross-cultural study on generation of culture dependent facial expressions of humanoid social robot. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. DOI: http://dx.doi.org/10.1007/978-3-642-34103-8{_}4

[36] Jeanne L. Tsai. 2007. Ideal Affect: Cultural Causes and Behavioral Consequences. *Perspectives on Psychological Science* (2007). DOI: http://dx.doi.org/10.1111/j.1745-6916.2007.00043.x

[37] Albert van Breemen, Xue Yan, and Bernt Meerbeek. 2005. iCat. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems - AAMAS '05*. DOI: http://dx.doi.org/10.1145/1082473.1082823

[38] Alessandro Vinciarelli, Maja Pantic, and HervÃĺ Bourlard. 2009. Social signal processing: Survey of an emerging domain. *Image and Vision Computing* 27, 12 (2009), 1743–1759. DOI: http://dx.doi.org/10.1016/j.imavis.2008.11.007

[39] Sherri C. Widen. 2013. Children's interpretation of facial expressions: The long path from valence-based to specific discrete categories. (2013). DOI: http://dx.doi.org/10.1177/1754073912451492

[40] Hui Yu, Oliver G.B. Garrod, and Philippe G. Schyns. 2012. Perception-driven facial expression synthesis. In *Computers and Graphics (Pergamon)*. DOI: http://dx.doi.org/10.1016/j.cag.2011.12.002

[41] Chang Yun, Zhigang Deng, and Merrill Hiscock. 2009. Can local avatars satisfy a global audience? A case study of high-fidelity 3D facial avatar animation in subject identification and emotion perception by US and international groups. *Computers in Entertainment* (2009). DOI: http://dx.doi.org/10.1145/1541895.1541901