Noname manuscript No.
(will be inserted by the editor)

Towards Artificial Emotions to Assist Social Coordination in HRI

Jekaterina Novikova · **Leon Watts**

Received: date / Accepted: date

Abstract Coordination of human-robot joint activity must depend on the ability of human and artificial agencies to interpret and interleave their actions. In this paper we consider the potential of artificial emotions to serve as task-relevant coordination devices in human-robot teams. We present two studies aiming to understand whether a non-humanoid robot can express artificial emotions in a manner that is meaningful to a human observer, the first based on static images and the second on the dynamic production of embodied robot expressions. We present a mixed-methods approach to the problem, combining statistical treatment of ratings data and thematic analysis of qualitative data. Our results demonstrate that even very simple movements of a non-humanoid robot can convey emotional meaning, and that when people attribute emotional states to a robot, they typically apply an event-based frame to make sense of the robotic expressions they have seen. Artificial emotions with high arousal level and negative valence are relatively easy for people to recognise compared to expressions with positive valence. We discuss the potential for using motion in different parts of a non-humanoid robot body to support the attribution of emotion in HRI, towards ethically responsible design of artificial emotions that could contribute to the efficacy of joint human-robot activities.

Keywords HRI · Artificial emotions · Communication of emotions · Social robot

J. Novikova
Department of Computer Science, University of Bath, Bath BA2 7AY, United Kingdom
E-mail: j.novikova@bath.ac.uk

L. Watts
E-mail: l.watts@bath.ac.uk

1 Introduction

The evolving relationship between humans and machines was recently highlighted as a major challenge by Gartner [**?**], a trade forecasting consultancy specializing in information technology research. More specifically, Gartner emphasized the emergence of opportunities and challenges associated with "*Humans and machines working alongside each other*".

The idea of 'working alongside' could mean anything from rigidly dividing an environment into human and robot zones, through to humans and robots sharing of responsibility and exchanging control as fellow team members. Robots could act as members of a human team by assisting people who share a given physical workspace, by performing actions relevant to their joint goals. Research on human-robot interaction (HRI) must address a number of challenges to make such coordinated action possible. Robots must act in a way that is understandable to the people with whom they are working, through the way they move and interact with objects in the shared space.

Team members routinely monitor their collaborators' attitudes to their individual and joint activity, as well as expressing their own attitudes to progress, through the presentation and interpretation of emotional signals. Human emotions are known to contribute to cognition and action, as people appraise the situation in which they find themselves. As collaborators, it is important for team members to maintain mutual appreciation of attitudes towards the progress of both individual and collective elements of joint work. For people, this is a multidimensional 'articulation' problem, including expectations that are rooted in social conventions as well as understanding of the immediate practical arrangements for accomplishing joint work [**?**]. Dynamic human coordination depends on inferences drawn from evidence about such attitudes during the production of joint work. These infer-

ences combine evidence in the form of events people perceive in the shared space, and in the form of the expressions produced by their collaborators, allowing people to form beliefs about the challenges currently facing collaborators, and about their intended actions. For people, affective expressions and responses are highly intuitive processes: they bring together biological, social and cultural factors to their interpretation of transient emotional states.

In order to benefit coordination, robot emotional signals should first of all be clearly expressed in a way comprehensible for humans. For robot emotional signals to function effectively in human interactions, it is necessary to consider the robot's internal state with respect to its ongoing activities, so that human collaborators can create relevant mappings from the set of signals it produces. In other words, 'artificial emotions' are a necessary prerequisite for generating intelligible robot emotional signals. Without this step, robot emotional signals are unlikely to serve interactions well.

In this paper we consider the potential of artificial robot emotions to serve as coordination devices in human-robot teams. We report an investigation of the potential for simple features of robotic embodiment to facilitate dynamic emotional signalling in a manner that allows interpretation by human observers. The broad aim of our work is to try to find a general scheme for communicating task-relevant internal states of a robot in a way which is both meaningful and intuitive for humans, with the ultimate aim of supporting successful social coordination between human and robot collaborators.

1.1 Human Emotions in HRI

Visual cues such as facial or bodily expressions are important in human-human coordination because they assist people to make inferences about one another's task-relevant state. For example, a grimace might indicate difficulty or a smile may suggest some success. Knowledge of this kind can help co-workers to bring their actions together at particular points, or to reschedule or reallocate work in case of difficulty.

Psychological studies have shown that humans are capable of inferring affective state from body movement [**?**], [**?**]. There is an extensive literature on describing of the perception of affect from expressive movements, whether from full body movements and gait, or from upper body movements. For full body expressive dance movements, Boone et al. [**?**] found that six cues are used to recognize an affect: changes in tempo (anger), directional changes in face and torso (anger), the frequency with which arms are raised (happiness), the duration for which arms are held away from the torso (happiness), muscle tension (fear), the duration of time leaning forward (sadness). Another study [**?**] characterized the qualities of movement for five target emotions in people who were walking. They specified Effort-Shape qualities for each of the emotions, e.g. the style characteristics for anger were defined as forceful, controlled, focused and fast movements of expanded limbs and stretched torso. Other researchers have focused on the association of affect with individual body parts: hand and arm movements have been found to be most significant for distinguishing between affective states [**?**]. Velocity, acceleration, and finger motion range are frequently reported as important features in hand and arm movement for distinguishing amongst affective states [**?**].

Research on the recognition of emotion in human-human interaction has inspired the creation of artificial emotional expressions in virtual agents [**?**] and robots [**?**]. However, it is important to remember that robots do not always have a humanoid or human-like body, thus the direct transfer of human emotional body language to a robot is not always easy or straightforward. In our studies, we use a non-humanoid robot for expressing emotional signals. Non-humanoid robots form an extremely large class in the whole range of different robotic forms. The map presented in Fig.1 shows different robotic embodiments ranging from highly expressive robots towards low expressive ones, to illustrate the importance of non-humanoid forms in the space of possible designs.

Low and semi-expressive non-humanoid robots can be used more often for home-working tasks (e.g. a robotic vacuum cleaner Roomba), search-and-rescue [**?**], domestic assistance [**?**] and other tasks. The design of such robots is intended to match their purpose, e.g. designed to move across disaster zones to find and reach victims, or to be steady and move safely in order to help elderly or disabled people get out of bed and move around. Thus it's not always useful or possible for such robots to have human-like bodies. However, as social agents, it is still useful for robots to be able to generate cues that are capable of expressing aspects of their state that are relevant for social coordination. And although most studies on the expression of emotions in robots are make use of humanoid robots, it is well known that humans can perceive affective states from non-anthropomorphic demonstrators [**?**] and even from abstract geometrical shapes [**?**].

In HRI, research on emotion recognition, expression, and emotionally enriched communication is of great potential importance and has been the subject of significant research effort since the mid-1990s [**?**], [**?**], [**?**]. Most of the existing work in social and humanoid robotics focuses on the recognition of human emotions [**?**], [**?**] or mimicking their expression [**?**], [**?**]. However, from an interaction perspective, understanding of social cues and a social context should not be considered as a one-sided process. In addition to understanding human emotions, more work should be done on the role of artificial emotions in human-robot teams and their impact on interaction.

Fig. 1 Multitude of robotic embodiments along a dimension of Expressiveness. Robots on the left contain more degrees of freedom available for expressivity.

1.2 Artificial Emotions in HRI

There is a growing body of research on techniques for expressing artificial emotions via facial expression, in both human-like and non-humanoid robots. The work of [**?**] explored interaction with the Lego-based 70cm-tall 'humanoid' Feelix robot through tactile stimulation so that various kinds of stimulation elicited the robots emotional responses. Observation of spontaneous interactions with Feelix showed that humans anthropomorphize a lot when interacting with objects with human-like features, so only a few of human-like emotion-related features are needed to make the interaction believable.

Eddie [**?**] is another low-cost emotional robot developed in Germany. The 23 degrees of freedom (DoF) and actuators assigned to particular action units of the facial action coding system allow it to express emotions using eyes, eyebrows, ears, mouth and jaw, and the crown. This robot uses animal-like features (crown of a cockatoo and ears of a dragon lizard) to display basic human emotions, which are recognized well by users.

Emotional expressions of a non-humanoid robot are presented in the work of [**?**] with a huggable animal-like robot Probo. Probo has a fully actuated head, with 20 degrees of freedom, capable of showing facial expressions and making eye contact. Robot's basic facial expressions are represented as a vector in the 2-dimensional emotion space based on Russells circumplex model of affect [**?**]. Probo robot is focused on interaction with hospitalized children.

It is perhaps unsurprising that the majority of prior work on emotional signalling focuses on facial expression. People typically identify sadness, for example, with a frown. However, the influence of affective states in humans and in animals is experience throughout the whole biological system. Sadness may also be accompanied by lowering of shoulders, slumping, a reduced pulse and slowing of bodily movements.

The potential utility of using embodied robotic expressions of emotion has been examined in a small number of recent studies [**?**] and [**?**]. Li [**?**] demonstrates the communication of emotion by a social bear-like robot through only simple head and arm movements. Child-robot interaction with a humanoid NAO robot is described in [**?**] that focuses on giving humanoid robots the capacity to express emotions with their body. Whilst work is maturing on the analysis of facial expression for robot emotional signalling, research on bodily postures for social robots is in its infancy. To date, very few researchers have examined how recognition of emotional states might be supported by combining these two facets of emotional communication, especially in non-humanoid robots. Our work intends to close the gap by presenting the study of expressing artificial emotions through both bodily movements and a simple facial feature in a non-humanoid robot.

This paper is organized as follows. We begin by setting out our methodological approach, defining the main research questions and the measures we selected for addressing the problem. We then present two exploratory studies, the first based on still images of robot poses and the second based on 'live' episodes of embodied robot emotional signalling. Details of each study are given together with its results. Finally, we conclude with a discussion of the results and suggest both implications for HRI and directions for further work.

2 Expressing and Interpreting Artificial Emotions

We present our questions, experimental platform and choice of emotion concepts for human interpretation, before explaining the scheme we used for quantitative analysis in studies one and two. In the stated research questions we use the terms *expression* and *intention*.

Expression here means an affect-expressive movement or a set of movements. Karg (2013) [**?**], proposed the following categorization of movements as potentially expressive:

- Communicative movements that can convey emotional meaning and are often performed in daily life.
- Functional movements. These are task related movements, such as walking.
- Artistic movements, e.g. dancing. Such movements do not occur in a daily life, are often exaggerated and over-expressive.
- Abstract movement, that are neither task related, not specifically expressive, e.g. lifting a leg.

In our work, we take a more restricted viewpoint. We treat emotional robot expressions as postures, movements or a sequences of movements, which are explicitly designed to communicate an affective state. Such an expressions are not required for carrying out a robot's task work, neither are they intended to be artistically exaggerated or abstract.

The term *intention* is used in our studies as a synonym of a *predicted action*. This is an inference drawn by an observer and it is merely an observer's beliefs about the likelihood of the robot's next action, regardless of whether it is in fact planned by the robot. Previous studies have linked affective expressions with fundamental behavioural form of

approach-avoidance [**?**], [**?**] and showed that basic behavioural intentions could be forecasted from emotional expressions [**?**]. This motivated one of our research questions.

2.1 Method

A series of studies was conducted in order to better understand whether a non-humanoid robot can express artificial emotions in a manner that is understandable for human. The studies have been conducted to examine three questions:

1. What meaning do people assign to the observed non-humanoid robot expressions?
2. Can people consistently recognize as emotional non - humanoid robot expressions presented to observers in a static or dynamic manner?
3. Can people consistently recognize robot intentions based on observed robot expressions?

In the first study participants were presented with static pictures of different robot expressions and asked to guess the observed robot emotion. In the second study, participants viewed dynamic expressions of the robot in a real time and were asked 1) to describe what the robot was doing in their own words (deliberately without asking participants to use emotional terms); 2) to guess the meaning of the observed expression by choosing from a controlled list of emotional terms, and 3) to guess the possible future robot actions, based on their beliefs about the meaning of the expression they had just seen.

The robot we have been experimenting with is shown in Figure **??**. It was implemented using Lego Mindstorms NXT and was based on a Phobot robot's design [**?**]. It includes a head element, with articulated 'eyebrows', that is mounted on a 'neck' element, and two limbs ('hands') attached to its control module. The robot was equipped with two motors that allowed it 1) to move forwards and/or backwards on a flat surface, and 2) to move its upper body part. The upper body part was constructed in such a way that the robot's hands were connected and moved together with robot's neck and eyebrows. The robot's neck section could move forward and backwards, its hands could move up and down, and its eyebrows could also rise and fall. Figure **??** presents three design sketches to illustrate the range of movement available for presenting emotional signals [**?**].

For programming robot's behaviours the RWTH – Mindstorms NXT Toolbox for MATLAB [**?**] was used. This software is a free open source product and is subject to the GPL. The RWTH toolbox was developed to control Lego Mindstorms NXT robots with Matlab via a wireless Bluetooth connection or via USB.

We prepared a controlled list of emotional terms which was presented to the participants as a list of possible options

Fig. 2 Lego robot used in the studies.

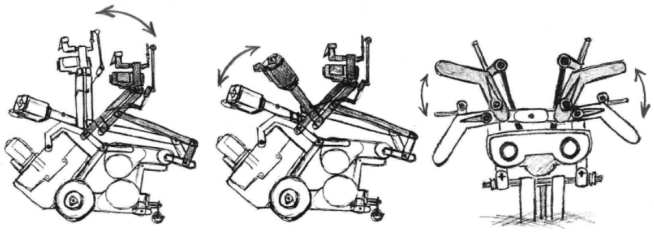

Fig. 3 A sketch of Lego robot's expressive movements (left - neck, middle - hands, right - eyebrows).

to choose from when characterizing the robot expressions. The list was created with an intention to balance proposed options in term of both valence and arousal. The main list consisted of seven emotional terms - *scared, surprised, excited, angry, neutral, happy* and *sad*. Later we have included additional terms *other* and *don't know* to the main list in order to provide the participants with additional options to express their opinions. The emotions from the main list were balanced in the valence-arousal circumplex model [**?**] over the dimensions of both valence and arousal, as shown in Figure **??**. Three options i.e. *scared, angry* and *sad*, belonged to a negative valence section V1; two options i.e. *surprised* and *neutral*, belonged to a no-valence section V2; and two more options i.e. *happy* and *excited*, belonged to a positive valence section V3. On the arousal dimensional area the *sad* option belonged to a low arousal section A3; *scared, surprised* and *excited* belonged to a high arousal section A1; and *angry, neutral* and *happy* were in the middle section that corresponds to an average-to-none arousal level in the section A2.

2.2 Measures

Two statistical measures were used to estimate the extent to which the robot emotional signals were interpreted consistently by our participants. These measures were used in both study 1 and study 2. However, we adopted a mixed-methods approach to our exploration of human responses to robot emotional signalling in study two by conducting a thematic analysis of the qualitative data provided by our

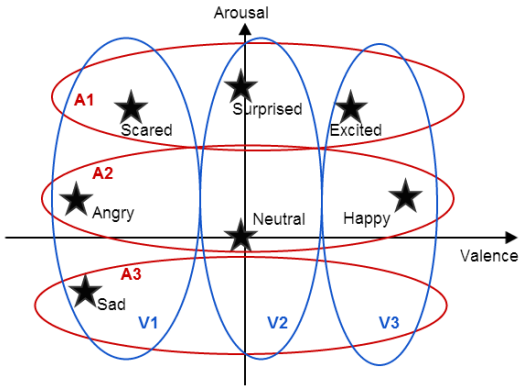

Fig. 4 Proposed emotional terms in a valence-arousal circumplex model. *A1, A2* and *A3* sections correspond to high, average-to-none and low arousal respectively. *V1, V2* and *V3* sections correspond to negative, neutral and positive valence respectively.

participants. The additional qualitative data was of great importance in providing meaning to the statistical results we found, given that we are committed to relating inferences about emotional signals to socially coordinated patterns of action from the perspective of human collaborators.

The first statistical measure represented the frequency of the term most often selected by participants, without regard to any initially intended emotion, and was based on the recognition ratio for each expression. The recognition ratio $r(p_i, e_j)$ for each picture or real-time expression was calculated as defined by Eq. **??**.

$$r(p_i, e_j) = \frac{N_{ij}}{N} \tag{1}$$

where p_i = picture or expression number i , e_j = selected emotional code number j; N_{ij} = number of responses (p_i, e_j); N = total number of respondents.

The second measure was used to estimate consensus of judgement among participants: the Fleiss' Kappa (κ) value [**?**]. The Fleiss' kappa value was used for measuring the agreement between the users regarding the observed robot emotion, as well as an expected robot's intention. The kappa value is a statistical measure for assessing the reliability of agreement between a fixed number of raters and is defined by Eq. **??**.

$$\kappa = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e} \tag{2}$$

$$\bar{P} = \frac{1}{Nn(n-1)} \left(\sum_{i=1}^{N} \sum_{j=1}^{k} n_{ij}^2 - Nn \right) \tag{3}$$

$$\bar{P}_e = \sum_{j=1}^{k} p_j^2 \tag{4}$$

The factor 1 - $\bar{P}_e$ gives the degree of agreement that is attainable above chance, and, $\bar{P} - \bar{P}_e$ gives the degree of agreement actually achieved above chance. If the raters are in complete agreement then $\kappa = 1$. If there is no agreement among the raters (other than what would be expected by chance) then $\kappa \leq 0$.

Table 1 Benchmark for strength of agreement indicated by κ value [**?**]

Kappa Statistics	Strength of Agreement
< 0	Poor
0.01–0.20	Slight
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Substantial
0.81–1.00	Almost perfect

In our studies: $i = 1, \ldots, N$ represents the participants, n is the number of pictures of Lego robot in the first study and the number of dynamic real-time robot expressions in the second study (with n_{ij} the number of ratings per picture/expression) and $j = 1, \ldots, k$ represents the possible answers (given in questionnaires). An interpretation of the κ values has been suggested by [**?**], and is presented in Table **??**. This table is however not universally accepted, and can only be used as an indication [**?**].

2.3 Study 1

2.3.1 Study 1 Apparatus

We programmed six combinations of robot's movements using them based on a basic arousal-valence underlying model [**?**], with approach and avoidance of the robot's neck and its whole body as a metaphor for valence and reflecting the arousal concept by raising its eyebrows. Then we photographed each combination from two angles – front and $^3/_4$ views. These two views were selected for presenting robot's expressions as these views are considered to be canonical for a large number of objects [**?**]. Moreover, the combination of the two views was proved to produce better face recognition performance [**?**]. The six pairs of pictures were used to construct a questionnaire provided to participants.

2.3.2 Study 1 Participants

27 people (14 females and 13 males) agreed to participate in a study to determine whether our simple set of valence-arousal robotic gestures could be interpreted as emotional signals. 18 had no previous experience with any kind of robots, 4 considered themselves as roboticists, and the rest had some previous interaction experience with robots. 18 were over 40 years old, 3 were between 30 and 39 years old, and six were between 20 and 29 years old.

Table 3 Participants' agreement regarding the robot's emotions in Study 1

Emotional Description	Fleiss' κ value	Interpretation of κ value
Scared	0.08	Slight agreement
Not emotional at all	0.05	Slight agreement
Surprised	0.14	Slight agreement
Angry	0.01	Slight agreement
Excited	0.05	Slight agreement
Sad	0.19	Slight agreement
Happy	0.01	Slight agreement

2.3.3 Study 1 Procedure

For each pair of images, participants were asked to select the most appropriate emotional term from a set of possible responses: sadness, happiness, anger, surprise, excitement, fear, other, no specific emotion and don't know. They were also asked to use a five-point Likert scale to rate their degree of confidence making that judgment.

2.4 Results of Study 1

The most frequently selected codes for these expressions were *surprised, sad, scared* and *excited*. The values of recognition ratio for each presented expression are given in the Table **??**. The recognition ratio for such emotions as *surprise, fear* and *sadness* were the highest (52, 42% and 42% respectively). The lowest recognition ratio was for the emotion of *anger*, as shown in the Table **??**.

The values of participants' confidence of the observed robot's emotion, on average, were quite similar for each emotional expression and differed in the range between 3.29 (SD = .80) and 3.79 (SD = 1.15), where '1' was *the least confident* and '5' was equal to *the most confident*, as presented in Table **??**. The confidence levels for the options *don't know* were ignored because this option does not represent any specific emotion.

The recognition ratio for each expression observed by the participants was significantly higher than the recognition ratio expected by chance ($p < .001$) However, the Fleiss' κ value calculated for each expression only showed a slight agreement between participants for each of recognized emotions, as show in Table **??**.

Reflection on study 1 identified three major methodological limitations: 1) an image of the end point of an expressive state may not convey the same meaning as the experience of seeing it performed in real time; 2) although it is assumed that people will naturally use anthropomorphic terms to describe non-human agents, forcing participants to use emotional labels undermines the validity of claims that emotional terms are spontaneously appropriate for robot signals, and 3) there was no context given to participants within which to interpret the signals.

2.5 Study 2

2.5.1 Study 2 Apparatus

The second study was designed to address the limitations discussed above. In the second study we programmed five dynamic expressions each intended as an emotional signal behaviour based on the combinations of the two movements of the same Lego robot and presented them to the participants in real-time, providing them with an emotionally neutral statement of the context in which the robot was acting. We also give our participants the opportunity to describe the robot's behaviour in their own words before asking them specific questions about emotional expression. A paper form was provided to the participants for them to describe the robot behaviour in their own words. A Matlab programmed questionnaire was presented to participants for selecting Likert scale responses to a set of questions (see Procedure below).

2.5.2 Study 2 Participants

The second study was conducted during a Bath University Open Day. 28 people (6 females and 22 males) agreed to participate in a study, ranging in age from 17 to 53 (M = 17.8, SD = .99), interested in human-robot interaction.

2.5.3 Study 2 Procedure

In the second study, conditions 1 and 2 were examined by presenting the five dynamic signal behaviours to participants successively in real-time. Each condition took approximately five minutes to complete. By way of context, participants were asked to consider that the robot was exploring an unfamiliar space when it noticed something. The language used to state context was deliberately intended to avoid leading participants to use emotional terminology rather than any other form of description.

Condition 1 required the participants first to explain in their own words what the dynamic expressions meant to them by writing whatever they liked on a paper form. Condition 2 repeated the same presentation of dynamic expressions but this time asked them to select a term of best fit from a fixed list of emotional terms. The participants were also asked to use a five-point Likert scale to rate their degree of confidence (1 - least confident, 5 - most confident) making that judgement. Finally, the participants were asked to choose the most likely "what happens next" option from another prepared list. All the questionnaires provided to the

Table 2 Recognition ratio for the expressions observed in Study 1.

Expression No / Code	Surprised	Scared	Excited	Sad	Neutral	Happy	Angry	Other pos.	Other neg.	Don't know
Expression 1	**29.6%**	3.7%	14.8%	11.1%	22.2%	11.1%	3.7%	3.7%	0.0%	0.0%
Expression 2	3.7%	11.1%	11.1%	**40.7%**	14.8%	0.0%	14.8%	0.0%	3.7%	0.0%
Expression 3	**51.9%**	22.2%	18.5%	0.0%	0.0%	7.4%	0.0%	0.0%	0.0%	0.0%
Expression 4	**33.3%**	22.2%	18.5%	0.0%	3.7%	14.8%	3.7%	3.7%	0.0%	0.0%
Expression 5	3.8%	**42.3%**	0.0%	30.8%	11.5%	0.0%	3.8%	0.0%	7.7%	0.0%
Expression 6	16.0%	4.0%	**36.0%**	0.0%	0.0%	12.0%	12.0%	8.0%	0.0%	12.0%

participants were in an electronic form in a Matlab environment.

2.5.4 The Thematic Analysis

The thematic analysis [**?**] was conducted for analysing qualitative data collected under the Condition 1 of the second study. Thematic analysis was advantageous for this purpose as it could offer an accessible and theoretically flexible approach to analysing qualitative data, produce a useful summary of key features, patterns and themes of a body of data, highlight similarities and differences across the data set and allow for social interpretations of data [**?**]. As a result of thematic analysis we produced an initial thematic map of five main themes shown in Figure **??**. The main themes developed at this stage of the analysis were: 1) emotional robot's state, 2) emotional robot's behaviour, 3) non-emotional robot's behaviour, 4) mental maps, and 5) interaction.

From this early stage thematic map we realized the relationship between themes (presented as circles in the Figure **??**) and different levels of sub-themes. A number of participants described the robot's expressions as an internal robot's emotional state emerged as a consequence of previous robot's interaction with its environment. The same explanation was very often seen in the descriptions of robot's behaviour, both when explained in an emotional and non-emotional tone. It is likely that people associated the changes of internal state with a previous interactional experience of the robot and made assumptions regarding that interaction. The same way, many participants made associations between the emotional state of the robot and its behaviour they observe. For some of the participants, the behaviour was a predecessor of a soon interactive act, for others it was a consequence or an accompanier. We explain such assumptions as a process when participants were creating mental maps about the presented robot and its surroundings in both place and time.

The interaction itself was described by participants in several different ways. The majority of participants described the object of the imagined interaction, which was a person himself, other unspecified people, non-human actors like pets and cats, different objects like table legs, parts of the environment like walls and floor. However, several participants

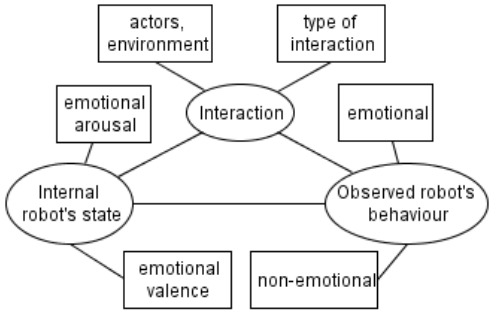

Fig. 6 Final thematic map, showing three final main themes

were more specific about the type of the interaction rather than the object the robot interacted with. In the description of robot's expressions they mentioned the words "investigating" and "investigate", thus defining the type of interaction they imagine. One person mentioned that the robot "was ignored" previously thus suggesting the previous unsuccessful interaction between the robot and some actor. The importance the concept of interaction had in the descriptions of participants means that people tend to directly relate emotional states and emotional behaviour with interactive acts, either previous, current or future. If such an interaction wasn't observed people just created it in their mind and related to the future or the past.

At the final stage we developed the final thematic map showing three main themes - internal robot's state, observed robot's behaviour and interaction, as shown in Figure **??**. These three main themes were developed by combining the different sub-themes of similar types into more general groups. We decided to exclude the *Mental maps* theme from the diagram, because as we have explained earlier the creation of mental maps is a consistent process consisting of investigating robot's internal state, the meaning of its behaviour and its interaction with the environment. Thus, creating mental maps is an overwhelming continuous process covering both understanding robot's internal state and behaviour and actually interacting with a robot. The remaining three themes nicely represent the famous "sense-act" reactive robotic paradigm [**?**], where changes in the internal robot's state represent the *sense* part of the loop, and the theme represents the *act*, i.e.

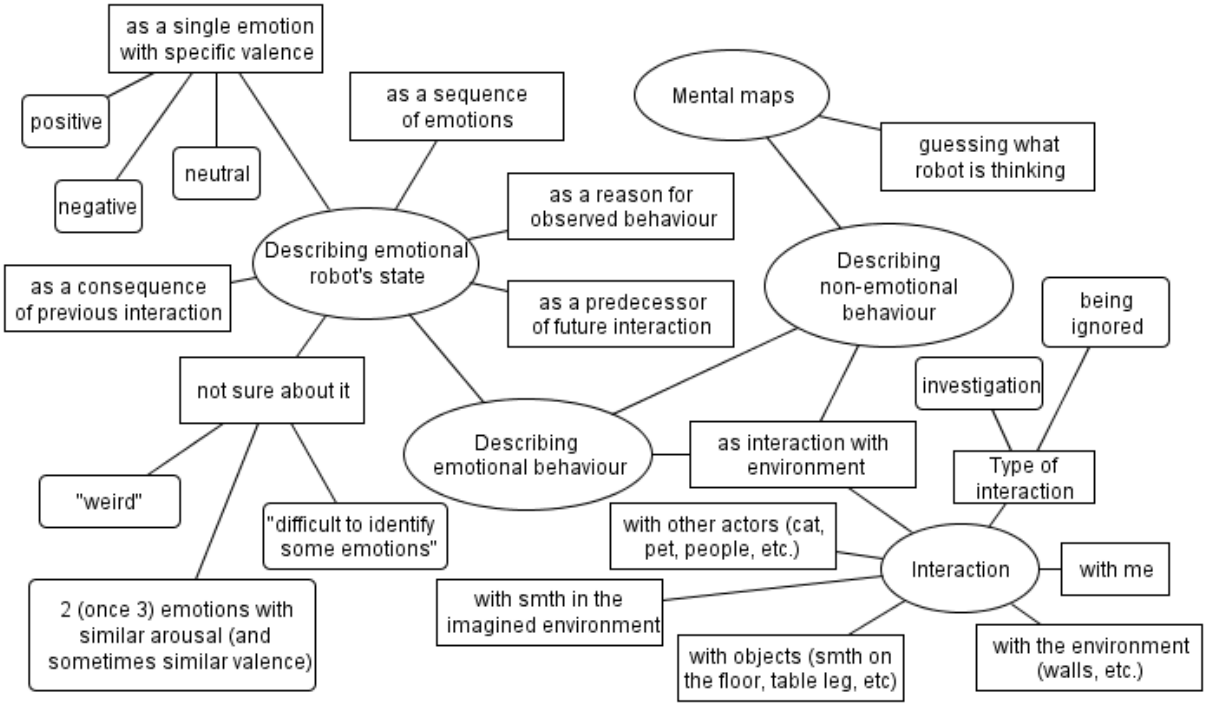

Fig. 5 Initial thematic map, showing five main themes

reactive response. The interaction theme here represents the loop itself.

2.6 Results of Study 2

For the dynamic robot expressions presented to the participants in the second study the recognition ratios were allocated as in the Table **??**, with the highest recognition ratio for the expressions 2 and 4 recognized as *scare* and *curiosity* respectively.

The values of participants' confidence of the observed robot's emotion, on average, were spread more widely comparing to the Study 1 and differed in the range between 1.50 (SD = .50) for *happiness* and 3.93 (SD = 0.81) for *surprise*, where '1' was *the least confident* and '5' was equal to *the most confident*, as presented in Table **??**.

The Fleiss' κ value calculated for each expression showed the moderate agreement for the emotion considered to be *scared* and for a non-emotional robot's expression. *Curious, surprised* and *angry* robot's emotions were interpreted with a fair agreement. Emotions interpreted as *excited* and *sad* had only a slight agreement, and for *pleased* participants didn't manage to agree, having a Fleiss' κ value smaller than 0, as shown in Table **??**.

There was a slight agreement between participants on the expectations of what the robot was going to do next – moving forwards/backwards, staying still, turning or doing something else. The highest values of agreement were presented for the choices *move forward* ($\kappa = 0.1322$) and *move backwards* ($\kappa = 0.1078$). However, none of the options ex-

Table 5 Confidence of the observed robot's emotion.

	Study 1		Study 2	
	Mean	St.Dev.	Mean	St.Dev.
angry	3.40	1.07	2.92	0.86
excited	3.48	0.70	3.54	0.63
happy/pleased	3.55	0.50	1.50	0.50
neutral	3.79	1.15	3.69	1.04
sad	3.57	0.66	3.08	0.95
scared	3.42	0.57	3.86	1.18
surprised	3.47	0.60	3.93	0.81
curious	-	-	3.69	1.18
other	3.29	0.80	3.69	1.04

Table 6 Participants' agreement regarding the robot's emotions in Study 2

Emotional Description	Fleiss' κ value	Interpretation of κ value
Scared	0.50	Moderate agreement
Not emotional at all	0.50	Moderate agreement
Curious	0.38	Fair agreement
Surprised	0.24	Fair agreement
Angry	0.24	Fair agreement
Excited	0.14	Slight agreement
Sad	0.01	Slight agreement
Pleased	-0.01	Poor agreement

Table 4 Recognition ratio for the robot's expressions observed in Study 2.

Expression No / Code	Surprised	Scared	Excited	Sad	Neutral	Pleased	Angry	Curious	Other emotion
Expression 1	**57.1%**	7.1%	7.1%	10.7%	0.0%	3.6%	3.6%	3.6%	7.1%
Expression 2	21.4%	**67.9%**	7.1%	0.0%	0.0%	0.0%	0.0%	0.0%	3.6%
Expression 3	0.0%	0.0%	0.0%	14.3%	**57.1%**	0.0%	0.0%	28.6%	0.0%
Expression 4	3.6%	0.0%	0.0%	14.3%	0.0%	3.6%	3.6%	**67.9%**	7.1%
Expression 5	14.3%	3.6%	**32.1%**	3.6%	0.0%	0.0%	**35.7%**	3.6%	7.1%

Table 7 Participants' agreement regarding the robot's intentions in Study 2

Robot's intention	Fleiss' κ value	Interpretation of κ value
Move forward	0.132	Slight agreement
Turn	0.028	Slight agreement
Stay still	0.033	Slight agreement
Move backwards	0.108	Slight agreement
Something else	0.028	Slight agreement
Don't know	0.070	Slight agreement

ceeded the boundaries of only a slight agreement, as shown in Table **??**.

Table 8 Emotional and non-emotional interpretation of robot's expressions in Study 2

Expression	Emotional	Non-emotional	Chi-square statistics
1	22	5	$\chi^2(1, N = 27) = 10.70$, $p = .001$
2	21	6	$\chi^2(1, N = 27) = 8.33$, $p < .005$
3	11	17	$\chi^2(1, N = 28) = 1.27$, *ns*
4	19	9	$\chi^2(1, N = 28) = 3.57$, $p = .05$
5	20	8	$\chi^2(1, N = 28) = 5.14$, $p < .05$

3 Discussion

Let us examine how the study answered our different research questions.

1. What meaning do people assign to the observed non-humanoid robot expressions?

According to the results of the second study we can state that the majority of people assign the emotional meaning to the observed robot expressions, given a simple context. Table **??** shows that the majority of participants interprets robot's expressions in an emotional way. Chi-square test shows that the differences between an emotional and non - emotional interpretations are significant for all the expressions except one: there is the only expression where the non-emotional interpretation exceeds the emotional one, although the difference is not significant ($\chi^2(1, N = 28) = 1.27, p = .26$), and it is the *neutral* expression where the robot is not moving its hands, neck and eyebrows at all. For all the other expressions an emotional interpretation is selected significantly more often than non-emotional.

The tendency to assign emotions to the robot's expressions repeats in the other condition of the second study. The results of the qualitative data analysis show that 46% (13 out of 28) of participants describe the observed expressions as emotional behaviour, and another 46% (13 out of 28) – as an emotion itself. Less than 1% (2 out of 28) describe observed robotic expressions as a non-emotional behaviour.

The thematic analysis shows that in addition to assigning an emotional interpretation to the robot's expressions, people tend to relate robot's emotional state to the predicted future or previous interaction. 63% of those explain the observed robot's emotional as a consequence of a previous interaction, the rest of the answers distributes between explaining the meaning of the observed emotion as 1) a reason for observed behaviour, 2) a tool for interacting with people and 3) a predecessor of a future interaction.

2. Can people consistently recognise as emotional non - humanoid robot expressions presented to observers in a static or dynamic manner?

The values of recognition ratio exceed the chance level for each recognized emotion but the recognition ratios in our studies are not very high, comparing to similar previous experiments completed by [**?**] [**?**] and [**?**], as presented in the Table **??**. However, comparing the abilities to represent emotional states of our robot that has only three DoF with other robots presented in the table, we consider the given results being very satisfactory. The possible explanations for lower recognition levels could be 1) in our first study the participants viewed only the static pictures of the expressions and this decreased the recognition rate, 2) in our studies we used the movements of the whole robot body together with the only one 'facial' feature - eyebrow, while in the previous mentioned studies the emotions were represented by facial expressions only.

The results show a significant difference between an average recognition ratio for positive (section *V3* in Figure **??**) and neutral (section *V2* in Figure **??**) emotions, $t(6) = 2.25$,

Table 9 Emotion recognition rate of robot emotions, partly adopted from [**?**]

	Our study 1	Our study 2	Feelix	Probo	Eddie
surprise	52	57	37	70	75
fear	42	68	16	65	42
sad	41	14	70	87	58
happy/excited	36	32	60	100	58
disgust	-	-	-	87	58
anger	15	36	40	96	54

$p < .05$ (one-tail), as well as between an average recognition ratio for positive (section *V3* in Figure **??**) and non-positive (sections *V1+V2* in Figure **??**) emotions, $t(12) = 1.78$, $p < .05$ (one-tail), with a lower recognition ratio for positive emotions in both cases.

The results also show a significant difference between an average recognition ratio for high arousal (section *A1* in Figure **??**) and average arousal (section *A2* in Figure **??**) emotions, $t(10) = 2.43$, $p < .05$, as well as between an average recognition ratio for high arousal (section *A1* in Figure **??**) and other arousal (sections *A2+A3* in Figure **??**) emotions, $t(12) = 2.59$, $p < .05$, with a higher recognition ratio for high arousal emotions in both cases.

The participants observing dynamic emotions have in general a significantly higher level of confidence (M = 3.52, SD = 1.03) over those observing static emotions (M = 3.49, SD = .74), $t(251) = -.265$, $p < .001$. However, having in mind specific emotional expressions only for the emotion of *scare* there is a significant difference in a confidence level between the participants observing static images (M = 3.40, SD = .57) and those observing dynamic real-time expressions (M = 3.86, SD = 1.21), $t(45) = -1.71$, $p < 0.05$.

3. Can people consistently recognise robot intentions based on observed robot expressions?

The results of our qualitative analysis show that the participants relate their observations of the robot's emotional signals to its interaction with the environment, and some sense of its previous experience. They thus set their interpretation into an event timeframe, whether as a matter of feelings attributed to the robot at that moment, as a result of a recent activity, or in anticipation of the robot's next action. Based on these statements, it is clear that our participants were making systematic attempts to interpret the robot's state given its behaviour. We expected the observers to have at least a moderate agreement about robot's immediate intention to act, based on the emotion attributed to it. However, the results of the inter-rater agreement analysis show that the low overall agreement between participants regarding the robot's expected action, with a highest Fleiss' κ value of 0.132 for the agreement regarding robot's intention to moving forward. Thus, the results of the studies we have reported here cannot support the statement that people can consistently recognise robot intentions based solely on the set of robot expressions we designed. The question of robot's intention recognition from its behaviour raises interesting issues and should be explored in future research. Although it is not possible to draw definitive conclusions from this study, it underlines the importance of setting any expressive behaviour into a context of action. In our work, the context of action will be set by joint work and so inferences about artificial emotion must also include ethical considerations.

3.1 Responsible Design of Artificial Emotions for Social Coordination

We introduced our work with a focus on non-humanoid robots as potential members of human-robot teams. The decision to operate outside of the constraints imposed by humanoid forms have a number of advantages. It is possible to explore a very wide range of forms and scales, primarily driven by a concern to create robots whose form fits their functional purpose. At the same time, we have arguably created a more difficult interpretative problem for the human team member, who will perhaps be more ready to consistently attribute emotional expressions to humanoids than to the expressive behaviours enacted by robots that are transparently mechanical. In other words, the work of working together creates a requirement for collaborators to infer one another's concerns and attitudes and so there could be a strong social function afforded by adopting humanoid forms. Humanoid forms may promote anthropomorphic attributions of thoughts and desires.

Our treatment of affect has been deliberately framed in terms of task-related responses to events in the context of collaboration: we have not attempted to promote a model that could support the attribution of more durative moods (e.g. 'the robot is annoyed') or sentiments (e.g. 'the robot thinks I am unkind'). In our introduction, we refer to 'empathic competence', in part to suggest the ethical uncertainty of work in this research area. Researchers who are working towards the construction of emotional robots must consider the potential risk of creating a mechanism that fools human collaborators into believing robots are capable of moral agency and moral reflection. Although we are working towards the possibility of robots becoming team members, we are not attempting to create a framework for people to put themselves at risk in order to protect the interests of the robot, or to believe that robots are capable of intervening to protect them when such action is simply not possible within their programming. We believe that maintaining a strong task focus for the interpretation of emotion signals will help to confront this ethical problem. It has been argued that the machine-nature of a robot should be made apparent to people who encounter it, in part to guard against inappropri-

ate or dangerous attributions [**?**]. Creating an emotional signalling system for non-humanoid robots should retain their value as social coordination mechanisms whilst at the same time preserving their transparently mechanical nature.

4 Conclusion

This paper has presented initial research concerning expression of artificial emotions in human-robot interaction. As in human non-verbal communication, expressive movements of the body and the face play an important role in HRI.

The goal of this research was to explore the relatively new research topic of facial and bodily gestures communication in social robots using a simple Lego robot as a case study and thus find a way of communicating internal robot state to humans in a both meaningful and intuitive way. We posed three main research questions: What meaning do people assign to the observed non-humanoid robot expressions? Can people consistently recognise as emotional non-humanoid robot expressions presented to observers in a static or dynamic manner? Can people consistently recognise robot intentions based on observed robot expressions?

We investigated these questions using two paired studies. Studies 1 and 2 were exploratory in that they tested perception of artificial emotions in robot expressive movements of its body and one facial feature in a simple situational context.

The results from this study demonstrate that even very simple movements of a social robot with three DoF only can convey emotional meaning, showing promise for designing non-humanoid robots that could serve as socially coordinated members of human-robot teams. They suggest that it is possible to create effective robot collaborators without an expressive human-like face, legs, moveable fingers or wrists. We have further argued that such an approach could help researchers and designers to contain the risk of inappropriate attributing robots with durative affective states, and moral agency, by emphasising their machine-like nature.

The results of this research provide a reason to believe that, in a context of a joint human-robot activity, it should still be possible for interaction designers to use interface elements such as body movements or extremely simple facial expressions to increase the expressive power of robots and thus increase a social coordination between human and robot in a human-robot team.

Future work will explore the effect of moving the different parts of the robot body on the interpretation of artificial emotions in HRI, as well as an effect of expressing artificial emotion on the efficacy of a joint human-robot activity.