You got the feeling: attributing affective states to dialogical social robots

Silvia De Marchi[1], Cristina Gena[1[0000−0003−0049−6213]], and Antonio Lieto[2,3[0000−0002−8323−8764]]

[1] University of Turin, Dept of Computer Science, Italy
silvia.demarchi/cristina.gena@unito.it
[2] University of Salerno, DISPC, Cognition Interaction and Intelligent Technologies Laboratory (CIIT LAB) and ICAR-CNR alieto@unisa.it

Abstract. In this paper we report the result of an analysis aiming at investigating, among different virtually embodied social robots (endowed with different degrees of dialogical complexity), the perceived difference in emotion attribution and understanding by the human users interacting with them. In particular, in our case study, the most complex dialogical modality - using a emotional content to vehiculate its messages - has been based entirely on the adoption of a Large Language Model (i.e. chatGPT in our case) whilst the simplest one has been based on a manual simplification of the generated text. We report the obtained results based on the adoption of a number tests and standardized scales and highlight some possibile future directions.

Keywords: Social Robotics · HRI · Affective Computing

1 Introduction

Empathy is an important aspect of human-human communication and building emphatic robots (i.e. robots able to elicit empathy in human users) represents a crucial challenge in the field of Human-Robot Interaction [20]. According to Hoffman's theory [34], one of the components of empathy is the affective one: it concerns, in detail, the emotional experience aroused by a stimulus of the same nature. Similarly Strayer [43], already in 1990, pointed out the co-participation of the affective component as the very content of empathy.

In the context of this paper we report a preliminary study assessing to what extent the use of affective content in dialogues during a human-robot interaction sessions impacts on the recognition and attribution of emotional and mental states to robots. More specifically: two different robots were employed: a virtual NAO and a virtual Pepper. Such virtual robots were endowed with different communication modalities: the NAO was able to provide answers by using an informative but neutral tone, communicating emotion simply and directly with a minimum of empathy; the Pepper, on the other hand, made use of emotion-driven and affective-charged content to deliver its messages, showing emotional participation and actively involving the user in a more articulate and constructive conversation. Both dialogues were generated by using a Large Language Model. The objective of this work was to detect which modality, used by the two different robots, was the most effective in making the user understand which emotion NAO and Pepper really wanted to convey and express. The analysis of this aspect represents one of the priority objectives in the study of social robotics. The correct affective interpretation and attribution of robot dialogues (from the user perspective) provides, indeed, essential elements for implementing and improving the communicative aspect and the entire process of empathetic interaction between man and robot. In other words, one could say that the correct recognition of the affective content (if any) that a robot deliver in its dialogues is an indicator that humans can attribute the correct *affective theory of mind* to the talking robots. And this element is of paramount importance for planning, from the robotic point of view, emphatic dialogues based in the the sharing of the same affective mood interpreted by the involved actors. Robots will undoubtedly become increasingly present in schools [14], factories [2], and homes [25] and, in our vision, their empathetic behavior certainly encourages their acceptance [13].

2 Empathy and Emotions Theories

According to Preston and De Waal [33] empathy can be defined as "the capacity to (a) be affected by and share the emotional state of another, (b) assess the reasons for the others' state, and (c) identify with the other, adopting his or her perspective". Following a shared categorization in psychology [29], empathy can be divided in three major categories: (1) empathy as an affective response to others' emotional states (*affective empathy*), (2) empathy as the cognitive understanding of others' emotional states, as well as the ability to put oneself in the other person's shoes (*cognitive empathy*), and (3) empathy as composed of both an affective and a cognitive component. Other perspectives [8, 42, 44] distinguish empathy in *dispositional* and *situational empathy*. While the former is a character trait, i.e. a person's general tendency to empathize, the latter is the empathy that a human perceives towards another agent in a specific situation.
Indeed, empathy is a concept that affects multiple fields of knowledge, from social to developmental, from clinical psychology to neuroscience. Since the discovery in 1996 of mirror neurons [12], interest in the concept of empathy has increased exponentially, also involving the field of human-robot interaction, see for instance [17, 19, 31, 40]. Similarly, during a human-robot interaction, we speak of the cognitive process when a robotic agent appears to individuals as being able to understand and imitate the emotions of others. The affective process occurs when the robotic agent manifests its emotions through voice, body posture, movements and gestures, adapted to the context of the situation.
According to several neurological and psychological researches [5, 15, 36] the involvement of mirror neuron system is implicated in neurocognitive functions, such as social cognition, language, *empathy*, and *Theory of Mind (ToM)* [4, 45], which is a human-specific ability that allows the attribution of mental states –intentions, thoughts, desires, and emotions– to themselves and others to explain

and predict behavior. As a consequence of this state of affairs, emotions (and their recognition), have been acknowledged as a primary component for building empathic robots.

In addition to their role, emotions also provide an universal language through which people convey their experience, well beyond words. Despite the differences in the expression of emotions across languages, and the influence of cultural factors, in fact, emotions own an universal origin [10]: rooted in evolution, they provide the basis for intercultural communication, as effectively demonstrated by the advancements in face expression recognition [6, 35]. In this sense, emotions can provide a suitable means for connecting robots with people belonging to different groups, intended as culture, age, education, and different sensory characteristics. Pervasive in human communication, emotions are expressed through multiple channels, ranging from face expression and body posture to spoken and written language. Emotion theories broadly belong to three main categories, partly derived from different research traditions. The expression of emotions through language, in particular, lies at the basis of several models of emotions. Categorical models focus on the definition of primary emotion types, which are assumed to be the result of phylogenesis. These emotion types are typically discrete and can be mapped straightforwardly onto face expressions. Sometimes referred to with the term 'basic emotions' to emphasize their innate nature, they appear at specific stages of the evolution of the child, progressively acquiring cognitive content. Depending on the reference theories, primary emotion types range from 5 to 6 [9] [26] including joy, anger, fear, disgust, sadness and sometimes surprise. Thanks to the tight relation with the preverbal (and postulated cross cultural) expression of emotions, these theories have deeply influenced the research on face expression recognition, through models such as the Facial Action Coding System (FACS), on which face expression datasets are built [23]. The model of six basic emotion prototypes proposed by Shaver et al. [41] has affinities with this group of models, but significantly differs from most of them from the methodological point of view: aimed at investigating the intuitions behind the human conceptualization of emotions, its design has been driven by the analysis of linguistic data.

Dimensional models represent emotions as the product of a set of predefined component dimensions, which axes such as polarity (often termed hedonic) and arousal. Historically derived from Wundt's three-dimensional definition of the emotional experience in terms of pleasure (pleasantness/unpleasantness), tension (tenseness/relaxedness), and excitement (excitement/depression) contemporary dimensional theories are usually represented through circumplex models, with significant variations: Plutchik's wheel of emotions acknowledges 8 bipolar emotions, derived from theoretical assumptions and rooted in behaviour [32], while Russel's circumplex model arranges empirically collected emotion labels in the continuous space generated by the two basic dimensions of arousal and polarity [38]. Depending on the dimensions considered (e.g., Mehrabian added dominance to the standard bidimensional space) different emotion types emerge from the intersections of the dimensions in the 2D or 3D space; in some models (e.g., Plutchik) secondary emotion types are generated by combining the primitive emotion types. Mainly geared to the subjective description and consequent expression of feelings, these models have influenced the creation of lexical resources for the analysis of sentiment [3].

Appraisal theories describe the subjective process of assessment of a situation which leads to the activation of an emotional state in a subject. These theories focus on the cognitive dimension of emotions [30], [18], [11], [39], [37], describing analytically the parameters that affect the emotional appraisal process. According to appraisal models, subjective motivations, or goals, and cultural factors, such as moral norms, affect the assessment of a given situation by an agent: each emotional category, then, is the result of a specific configuration of appraisal parameters, usually represented in the form of an activation rule. Appraisal models allow the same situation to be appraised differently by different individuals, and postulate complex emotions as the result of the activation of multiple appraisal processes on the same situations: for example, in the OCC model [30], the activation of distress and reproach yields the emotion of anger. The emotional coping process [37], the natural complement of appraisal, describes how the agent responds to the activated emotions at the mental and behavioural levels, in continuity with the appraisal parameters. Due to their cognitive background, appraisal models lend themselves to the integration with agent models, and to mentalistic models such as the Belief-Desire-Intention model (BDI) [7].

3 The experiment

In the present work, we compared two different expression modalities of diverse virtual robots (NAO and Pepper) built by using the Aldebaran Choreographe software [1]. Both robots had to express messages, within a structured conversation, able to convey six different types of emotions extracted from the Plutchik's wheel, namely: joy, envy, sadness (three basic emotions) and surprise, disapproval and curiosity (three complex emotions). Positive emotions produce pleasant effects in individuals, fostering a state of emotional and psychological well-being, whereas negative emotions lead to unpleasant sensations, causing frustration and emotional discomfort. All six emotions were presented to the users during a dialogue. Such dialogues, in both cases, were generated by a Large Language Model (chatGPT in its GPT 3.5 Turbo Version). However, for the NAO robot they were manually simplified in order to provide the more neutral and as simple as possible information. In order to elicit the above mentioned emotions, we requested to the LLM to use, within the generated sentences, a set of keywords resulting as highly associated with them by using the DEGARI affective-based reasoner [21], already successfully applied for suggesting stories from multiple affective viewpoints in museums [22], and its associated NRC lexicon [28].

Specifically, the goal of the experiment was to provide a preliminary response to the following questions: what emotions did individual robots express **(RQ1)**? This aspect has been assessed based on the robot's ability, as perceived by the user, to coherently express different emotions in an understandable manner.

What differences are observable (if any) in the emotional expressions manifested by NAO and Pepper robots **(RQ2)**? This aspect investigates whether there are significant differences in how the two robots express emotions and if one is more effective in terms of clarity and expressiveness. Did subjects report different level of engagement experiencing for the specific emotions conveyed by the robot **(RQ3)**? The answers to these questions provides a basis for reflection on the complex and articulated world of social robotics, despite the limited number of experiences collected.

We report the obtained results based on the adoption of a number tests and standardized scales and highlight some possible future directions.

Fig. 1. Experimental setting of a user interacting with the virtual NAO. On the desk are visible the different QR codes leading to post-test questionnaire and the document explaining the experiment (to read before starting it).

3.1 Method and Interaction Steps in the Dialogues

The sample considered for this study consists of 32 Master's degree students in Communication, ICT, and Media, and students from the Faculty of Social Innovation, Communication, and New Technologies at the University of Turin. There were 20 females, with an overall age range between 19 and 32 years.

The experiment was conducted in a dedicated space within the Luigi Einaudi Campus in Turin with the aid of a PC station connected to the Choregraphe software. Due to a malfunction of the NAO robot, the experiment was carried out using virtual robots available through the programming software mentioned above, instead of directly employing the two robots as initially hypothesized.

For this reason, the user from the identified sample is invited to sit in front of the computer and start interacting with the robot previously selected by the researcher (as shown in Figure 1). The robot initiating the interaction changes from subject to subject (e.g., user 1 begins the experiment with the NAO robot, user 2 begins with Pepper), as does the order of presentation of the emotions expressed by the two artificial agents, to randomize potential effects from preset sequences that are the same for all users.

The procedure we followed has the following structure: i) the user seated at the dedicated PC station; ii) users are introduced to the experiment through a document; iii) the participant is asked to fill out a first empathetic evaluation questionnaire (Interpersonal Reactivity Index [16]), accessible via a specific QR Code, placed on the work table. The user begins the interaction with the assigned robot. For simplicity, we will use user number 1 as an example from now on.

NAO will attempt to express curiosity and then proceed with the remaining five emotions. At the end of each dialogue with NAO (and the related emotion expressed by the robot), user 1, through a second QR Code, can access a second questionnaire that will remain active for the duration of the experiment. This questionnaire aims to identify the emotion perceived by the human interlocutor from the interaction and their level of involvement.

The subsequent step involves the user interacting with the second robot and similarly completing the questionnaire introduced in the previous point.

Pepper will attempt to express surprise and then proceed with the remaining five emotions. Thus, user number 1, after interacting with NAO, will proceed with Pepper, who will seek to express the feeling of surprise and then continue with the remaining five emotions.

The greeting is the main input to initiate communication with the robot, marking the moment when the latter detects the presence of the user and activates to interact with them. NAO responds to the greeting and expresses the emotion it is feeling. At the end, NAO asks the human a specific question, which will vary depending on the emotion expressed (e.g., "Would you like to get to know each other better?", "Do you agree with me?", "Have you ever felt this sensation?"). The user's response can mostly be affirmative or negative; different ways for the user to respond to the robot's questions are also anticipated.

Pepper, by managing a conversation in a more personalized way, returns the greeting and asks the user's name. This information, stored by the robot, facilitates the creation of a relationship of knowledge and trust and can be recalled at any time during the interaction. It is hypothesized that the user may respond to the question using various expressions. Expressions a user can use to answer Pepper's question "What's your name?" Once the user's response is received, Pepper introduces the scenario for the emotion it will express. Subsequently, the robot asks the user a specific question, which, again, will vary depending on the dialogue (e.g., "Would you like?", "Can I explain how I feel?", "Can I tell you about this event?", etc.). Depending on the user's response, the dialogue can proceed in two directions: a negative response leads to the end of the conversation, where the robot bids farewell to the user and concludes the interaction. An affirmative response leads Pepper to delve into the expression of the emotion felt, opening up various arguments based on what the user has expressed.

4 Evaluation

In our evaluatiuon we have used a *within-subject design* method and the Interpersonal Reactivity Index (IRI), created in 1983 by Mark H. Davis [16]. IRI is a tool used in psychology to gather evaluative elements about empathy and how people detect, understand, and react to others' emotions and experiences. It is used in this work as the first test administered to the user before the human-robot interaction.

The IRI consists of four subscales, each comprising seven items that probe each component of the empathic process: perspective-taking (PT), empathic concern (EC), personal distress (PD), and fantasy (FS). The total of 28 items are presented in the form of statements, to which the subject can respond using a Likert scale ranging from "does not describe me at all" to "describes me very well." The total score obtained for each of the four scales, ranging from 0 to 35 since each scale includes seven statements with scores ranging from 0 to 5, can be summed to achieve a comprehensive score between 0 and 140. A high score indicates a greater level of empathy and a higher relational sensitivity. This scale's measurement highlights whether an individual spontaneously considers others' viewpoints in daily life. A low score suggests a limited ability to consider others' perspectives and difficulty in understanding emotions different from one's own. A high score, on the other hand, suggests an adequate understanding of emotional states and viewpoints of others, essential for an empathic relationship. The Interpersonal Reactivity Index (IRI) scales are tools that, alongside a more detailed evaluation, can contribute to enriching the puzzle that constitutes the complex multidimensional picture of empathy. These scales acquire significance only when interpreted in their entirety and are listed below.

The perspective-taking scale. The perspective-taking (AP) scale refers to an individual's intellectual or imaginative ability to put oneself in another's shoes [24]. In particular, a subject who scores high on the AP scale shows predisposition towards communication, tolerance and interpersonal relationships. Subjects with a high AP score also tend to have a flexible mindset, which allows them to adapt their thinking to different situations. An extremely high score in this area can be negative as it can interfere with the ability to make decisions. Conversely, a low score is a sign of poor cognitive empathy, and is typical of individuals who exhibit little mental flexibility and are not good at understanding the mental state of others. An extremely low score on this scale can be related to a significant deficit in interpersonal and communication skills [24].

The Fantasy Scale. The utilization of this scale aims to probe the extent to which an individual can immerse themselves in fictional situations through

imagination, as might occur, for instance, through reading a book or watching a film. A low score signifies a reduced inclination towards engagement in imaginary scenarios and is more commonly observed in individuals who are highly grounded in reality, pragmatic, and less willing to envision themselves in fantastical situations. These traits alone are not sufficient to delineate a person's level of empathy. Conversely, a high score typically characterizes highly creative individuals who are more disposed to immerse themselves in imaginary stories.
Empathic Concern scale (EC). The adoption of this scale facilitates the assessment of the predisposition exhibited by the interviewed subjects towards experiencing feelings of compassion and emotional engagement towards others who are undergoing negative experiences. A low score indicates a lesser degree of involvement, whereas a high score suggests that the individual is interested and actively concerned with the emotional well-being of others. In the former case, individuals might experience difficulties in engaging empathetically; in the latter, the outcome suggests a genuine interest and concern for others, with these empathic subjects feeling emotionally involved and eager to alleviate the emotional distress others may face.
Personal Distress Scale (PD). The Personal Distress Scale evaluates situations in which witnessing unpleasant events involving others generates anxiety and dismay in the observer. Such situations can lead to a genuine loss of control in some individuals. A low score serves as an indicator of a certain capacity for emotional management, wherein individuals might be better equipped to handle complex situations. Conversely, a high score identifies individuals who react to similar situations with discomfort and suffering, accompanied by a high level of stress. These subjects feel deeply emotionally implicated and recognize the need to reduce the distress experienced by others.

During the interaction phase, at the end of each dialogue between the robot and the user, the latter is administered a **second questionnaire** accessible with the dedicated QR Code, consisting of two questions. These questions are presented at the end of each task to capture the immediate memory of the interaction just occurred with the robots. The first question aims to verify which emotion was expressed by the robot, and the second assesses the level of user involvement during the interaction:

- **Question 1**: Identification of the emotion expressed by the robot "*In your opinion, which emotion did the robot express*?". In this case, the user must select an emotion from a dropdown menu;
- **Question 2**: Evaluation of user involvement in the interaction "*How involved did you feel in the emotion expressed by the robot?*". The question is posed to the user after identifying the perceived emotion following the dialogue with the robot. In this case, their response can range from 1 ("not at all") to 5 ("very much"), and the purpose is to measure how involved the user felt in the emotion expressed. It also aims to observe the emotional response following positive or negative emotions, considering that the latter typically have a greater impact on humans compared to positive emotions.

.

AMS (Attribution of Mental State Questionnaire. As an additional tool for post-interaction evaluation we administered to the to the users, upon completion of the experimental tasks conducted with both NAO and Pepper, The AMS (Attribution of Mental States) Questionnaire [27]. AMS is used for measuring and assessing the attribution of mental states (see also [13]) to the two robots.

In the context of this work, the questionnaire is employed to measure the mental states that experiment participants assign to the robot and, in particular, to highlight the user's perception regarding the robot's mental characteristics in comparison with those of humans.

Another aim is to gather information to understand if the attribution of mental states to the robot is influenced by the expressive mode enacted by the robot itself during dialogue. That is, to detect if the modes manifested by the two robots, one more direct and mechanical versus the other more articulated and complex, affect the attribution of mental states expressed by the participant at the end of the experiment.

The questionnaire consists of 25 questions with response options: "a lot: score 2", "a little: score 1", "not at all: score 0". The sum of all responses determines the overall score, which ranges from 0 to 50.

Furthermore, the AMS questionnaire comprises five specific dimensions, with partial scores being the sum of responses within each dimension, ranging from 0 to 10. The dimensions are:

- ***Epistemic***: related to participants' concept of the robots' cognitive intelligence: can the robot understand? Can it make decisions? Can it think?
- ***Perceptual***: the dimension related to the possibility that the robot experiences sensations, such as smell, sight, taste, etc.
- ***Emotional***: can the robot experience feelings of anger, happiness, fear?
- ***Desires***: can the robot express desires or preferences?
- ***Imaginative***: is the robot capable of dreaming or imagining?

The questionnaire, presented in a single form, was administered twice: once to gather information on NAO and a second time for Pepper. The sum of scores in each area allowed the comparison of results through the T-test and if there are significant differences between the two questionnaires, and in which areas.

In addition to the questions in the questionnaire, further inquiries specifically chosen to probe more deeply into the level of interaction the experiment participant had with the robot were included: "Was this the first time you interacted with Pepper?"; "Was this the first time you interacted with NAO?"; "Have you interacted with other robots in the past?"; "If you answered 'yes' to the previous question, which one(s)?"; "Which type of interaction (NAO or Pepper) did you prefer?"; "Indicate the factors that determined your choice."

5 Results and Future Works

We obtained, for the above mentioned scales composing the IRI index, the following results. For the Perspective-Taking scale (PT): the total score obtained, equal to 806, indicates a considerable level of empathic involvement, a sign of a high understanding of the emotional needs manifested by others. The value, compared to the collected data, identifies a sample sensitive to the needs and emotions of other people. For the Fantasy Scale (FS): the total of 500, obtained from the sum of the scores of this subscale, suggests that the sample is mostly composed of pragmatic subjects, less inclined to engage in emotional fantasies or immerse themselves in imaginary roles. For the Empathic Concern scale (EC): the score of 901, being the highest in the overall assessment of the questionnaire, suggests that the sample has a high propensity to put themselves in others' shoes and to welcome and understand the emotional expressiveness shown by other people. Finally, for the Personal Distress scale (PD), the total score detected (623) indicates that the sample possesses a moderate level of discomfort in relation to emotionally stressful situations involving other people. The overall data suggest a sample of subjects who feel emotionally involved and subjected to a high degree of stress when others experience discomfort or suffering.

In summary, for the sample of 32 subjects to whom the questionnaire was administered, it can be hypothesized that they possess an adequate level of empathy. This interpretation is supported by significant values referred to the Perspective-Taking and Empathic Concern scales, suggesting a high level of acceptance and understanding towards others from an emotional perspective. The values from the Personal Distress scale suggest a moderate sensitivity in perceiving others' discomfort, while a medium-low ability is detected regarding the Fantasy subscale, which could indicate that the interviewed subjects tend more towards a rational and concrete approach to emotional experiences, rather than opting for a more fanciful and imaginary view of reality.

Based on such assessment, we also analyzed the results of the above mentioned questionnaire that allowed us to check to what extent it was possibile to identify different levels of *affective theories of mind* attributed by the users to the different robots and robotic modalities they interacted with.

The additional data collected through the questionnaires complete the picture of information intended to be collected following the experiment. After identifying the emotion the robot wanted to transmit, it was intended to verify whether the different communicative attitude attributed could determine, in the interviewed subjects, an emotional implication capable of influencing their choice regarding the preference between the two robots.

In particular, for what concern the correct emotion identification and attribution we analyzed - for the two testing conditions - the total total number of users who correctly identified ("Yes") and those who did not identify ("No") the individual emotions that the robots NAO and Pepper tried to express during the interactions Overall, the data show that the error rate, pertaining to the users' failure to recognize the emotion the robot intended to convey, is higher for NAO (13 errors out of 192 interactions, equivalent to 6.8%) compared to Pepper (6 errors out of 192 interactions, equivalent to 3.1%). This difference may be attributed to Pepper's more complex dialogue structure, which could have facilitated a better understanding of the intended emotional transmission.

However, it is important to note that the detected errors are mostly related to the users' selection of a different emotion, albeit similar to the one the robot initially aimed to express.

We conducted a paired t-test to determine if there were significant differences in the identification of individual emotions as reported separately by the two robots. The result ($p = .135$) suggests that even the more synthetic and direct expressive mode deployed by the NAO robot was effective in making understandable the emotion it intended to convey; the simplicity of the dialogues attributed to it might have positively influenced this result.

Regarding the level of involvement that the interviewed subjects declared having felt for the specific emotions transmitted by the robots (Question 2 of the second questionnaire), the data reveal that the mode used by Pepper was more effective than that of NAO, with statistically significant differences between the two robots ($p < .005$). This datum is aligned with the fact that more articulated, and expressed through a more complete and engaging expressive mode, were attributed to Pepper robot; on the contrary, the input provided to NAO involved very synthetic and direct responses, not supported by engaging dialogue.

Regarding the subcategories of the AMS questionnaire, the responses provided by users towards were as follows.

Epistemic Dimension. This category pertains to the cognitive capabilities recognized in the NAO robot. 37.5% of participants attributed an absolute lack of such capability to the robot; 46.3% recognized it as having limited cognitive ability, and 16.2% attributed a high cognitive capacity to NAO. For the Pepper, 3,8% of the subjects attributed to the robot the complete lack of such capability; 21,9% recognized it as having limited cognitive ability and 74,3% attributed a high cognitive capacity. Here the result of the paired T-test showed a significant difference for the first ($p = 0.01$)and the last item ($p = 0.0003$).

Perceptual Dimension. This dimension acknowledges the robot's ability to experience sensations related to the five senses. In this case, 73.8% of the surveyed users did not find NAO to possess such ability; 22.5% recognized the robot as having a reduced perceptual capacity, while 3.7% defined NAO as a robot with a high ability to perceive sensations. Concerning the Pepper: 39.3% did not find robot to possess such ability; 28,8% recognized the robot as having a reduced perceptual capacity and 31,9% defined the robot as having a high ability to perceive sensations. Here the result of the paired T-test showed a significant difference for all the items ($p < 0.002$).

Emotional Dimension. This dimension explores the level at which the robot can experience feelings such as anger, fear, happiness, and sadness. 21.9% of the respondents did not recognize this capability in NAO; 61.9% attributed it with a low capacity, while the remaining 16.2% believed that NAO is capable of experiencing such sensations. For the Pepper, 6.3% of the respondents did not recognize this capability in the robot, 11.3% attributed it with a low capacity, and 82.5% believed that Pepper is capable of experiencing such sensations. Here the result of the paired T-test showed a significant difference for the first and the last item (both $p = 0.002$).

Desire Dimension. This dimension assesses the robot's ability to express desires or preferences. 61.3% of respondents did not recognize this capability in NAO; 31.2% acknowledged it a limited capacity, while 7.5% attributed it a high capacity. For Pepper, the results were (in the same order) of 8.8%, 30% and 61.2% respectively. Here the result of the paired T-test showed a significant difference for the first ($p = 0.01$) and the last item ($p = 0.003$).
Imaginative Dimension. This dimension evaluates the robot's ability to imagine. For this aspect, 70% of respondents did not recognize any imaginative capacity in NAO; 20.6% acknowledged it as having a limited capacity, while the remaining 9.4% attributed a high imaginative capacity to the robot. For Pepper, the results were (in the same order) of 21.89%, 24.3% and 53.8% respectively. Here the result of the paired T-test showed a significant difference for the first ($p = 0.01$)and the last item ($p = 0.003$).

The AMS data show that the users attribute greater capabilities in Pepper across all five dimensions of the AMS questionnaire. The interaction with Pepper was also the most preferred one; among the main reasons supporting this preference were: good emotional management by the robot (22.3%), a high degree of engagement (19.4%), and constructive communication (15.8%) as key elements for the success of a human-robot interaction. Regarding the reasons that led five users to prefer interacting with NAO, the most influential factors were: clarity of exposition by the robot (25%), the ability to synthesize (25%), and simple communication (25%).

Overall, the main findings of this preliminary experimentation concern the fact that emotions seem to be detected in both the robots (RQ1), but their identification does not appear to be related to the different mode adopted by the two robot expressing it (RQ2), suggesting that the expressive mode, more synthetic and direct, employed by the NAO robot, was nevertheless effective in conveying the intended emotion. This result could have been positively influenced by the simplicity of the dialogues attributed to it.

Regarding the level of engagement that the interviewed subjects reported experiencing for the specific emotions conveyed by the robots (RQ3), the data analysis reveals that the mode used by Pepper was more effective than that of NAO, with statistically significant differences between the two robots. It is interesting to note how the more structured dialogue, expressed by Pepper, also determined in the users an attribution of the robot's capabilities, in term of identified mental states, even superior to those actually present in the dialogue. The data confirm that the type of dialogue between man and robot influenced the perception that users have of them.

Some of the limitations of the current work concern the use of virtual (instead of physical) robots. Here, we plan to repeat the experiment with embodied robots. In addition, we plan to extend the experimental sample of the tested users and to extend the evaluation to the entire emotional spectrum of the Plutchik's theory of emotion.

Bibliography

[1] Aldebaran: Nao software 1.14.5 documentation

[2] Brunetti, D., Gena, C., Vernero, F.: Smart interactive technologies in the human-centric factory 5.0: A survey. Applied Sciences **12**(16), 7965 (2022)

[3] Cambria, E., Li, Y., Xing, F.Z., Poria, S., Kwok, K.: SenticNet 6: Ensemble application of symbolic and subsymbolic AI for sentiment analysis. In: Proceedings of the 29th ACM International Conference on Information & Knowledge Management. pp. 105–114 (2020)

[4] Carlson, S.M., Koenig, M.A., Harms, M.B.: Theory of mind. Wiley Interdisciplinary Reviews: Cognitive Science **4**(4), 391–402 (2013)

[5] Cattaneo, L., Rizzolatti, G.: The mirror neuron system. Archives of neurology **66**(5), 557–560 (2009)

[6] Cordaro, D.T., Sun, R., Keltner, D., Kamble, S., Huddar, N., McNeil, G.: Universals and cultural variations in 22 emotional expressions across five cultures. Emotion **18**(1), 75 (2018)

[7] Dennett, D.C.: The intentional stance. MIT press (1989)

[8] Eisenberg, N., Fabes, R.A., Murphy, B., Karbon, M., Maszk, P., Smith, M., O'Boyle, C., Suh, K.: The relations of emotionality and regulation to dispositional and situational empathy-related responding. Journal of personality and social psychology **66**(4), 776 (1994)

[9] Ekman, P.: Basic emotions. Handbook of cognition and emotion **98**(45-60), 16 (1999)

[10] Ekman, P., Friesen, W.V.: Constants across cultures in the face and emotion. Journal of personality and social psychology **17**(2), 124 (1971)

[11] Frijda, N.H., others: The emotions. Cambridge University Press (1986)

[12] Gallese, V., Fadiga, L., Fogassi, L., Rizzolatti, G.: Action recognition in the premotor cortex. Brain **119**(2), 593–609 (1996)

[13] Gena, C., Manini, F., Lieto, A., Lillo, A., Vernero, F.: Can empathy affect the attribution of mental states to robots? In: Proceedings of the 25th International Conference on Multimodal Interaction. pp. 94–103 (2023)

[14] Gena, C., Mattutino, C., Perosino, G., Trainito, M., Vaudano, C., Cellie, D.: Design and development of a social, educational and affective robot. In: 2020 IEEE Conference on Evolving and Adaptive Intelligent Systems, EAIS 2020, Bari, Italy, May 27-29, 2020. pp. 1–8. IEEE (2020). https://doi.org/10.1109/EAIS48028.2020.9122778, https://doi.org/10.1109/EAIS48028.2020.9122778

[15] Iacoboni, M.: Imitation, empathy, and mirror neurons. Annual Review of Psychology **60**(1), 653–670 (2009). https://doi.org/10.1146/annurev.psych.60.110707.163604, https://doi.org/10.1146/annurev.psych.60.110707.163604, pMID: 18793090

[16] Keaton, S.A.: Interpersonal reactivity index (iri) (davis, 1980). The Sourcebook of listening research: Methodology and measures pp. 340–347 (2017)

[17] Kwak, S.S., Kim, Y., Kim, E., Shin, C., Cho, K.: What makes people empathize with an emotional robot?: The impact of agency and physical embodiment on human empathy for a robot. In: 2013 IEEE RO-MAN. pp. 180–185 (2013). https://doi.org/10.1109/ROMAN.2013.6628441

[18] Lazarus, R.S.: Progress on a cognitive-motivational-relational theory of emotion. American psychologist **46**(8), 819 (1991), publisher: American Psychological Association

[19] Leite, I., Pereira, A., Mascarenhas, S., Martinho, C., Prada, R., Paiva, A.: The influence of empathy in human–robot relations. International journal of human-computer studies **71**(3), 250–260 (2013)

[20] Lieto, A.: Cognitive design for artificial minds. Routledge (2021)

[21] Lieto, A., Pozzato, G.L., Zoia, S., Patti, V., Damiano, R.: A commonsense reasoning framework for explanatory emotion attribution, generation and re-classification. Knowledge-Based Systems p. 107166 (2021)

[22] Lieto, A., Striani, M., Gena, C., Dolza, E., Marras, A.M., Pozzato, G.L., Damiano, R.: A sensemaking system for grouping and suggesting stories from multiple affective viewpoints in museums. Human–Computer Interaction **39**(1-2), 109–143 (2024)

[23] Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I.: The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In: 2010 ieee computer society conference on computer vision and pattern recognition-workshops. pp. 94–101. IEEE (2010)

[24] López-Pérez, B., Fernández-Pinto, I., Abad, F.: TECA. Test de Empatía Cognitiva y Afectiva. Tea Ediciones, S.A., Madrid (2008)

[25] Macis, D., Perilli, S., Gena, C.: Employing socially assistive robots in elderly care. In: UMAP '22: 30th ACM Conference on User Modeling, Adaptation and Personalization, Barcelona, Spain, July 4 - 7, 2022, Adjunct Proceedings. pp. 130–138. ACM (2022). https://doi.org/10.1145/3511047.3537687, https://doi.org/10.1145/3511047.3537687

[26] Mehrabian, A.: Comparison of the PAD and PANAS as models for describing emotions and for differentiating anxiety from depression. Journal of psychopathology and behavioral assessment **19**(4), 331–357 (1997), publisher: Springer

[27] Miraglia, L., Peretti, G., Manzi, F., Di Dio, C., Massaro, D., Marchetti, A.: Development and validation of the attribution of mental states questionnaire (ams-q): A reference tool for assessing anthropomorphism. Frontiers in Psychology **14** (2023)

[28] Mohammad, S.: Word affect intensities. In: Calzolari, N., Choukri, K., Cieri, C., Declerck, T., Goggi, S., Hasida, K., Isahara, H., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J., Piperidis, S., Tokunaga, T. (eds.) Proceedings of the Eleventh International Conference on Language Resources and Evaluation, LREC 2018, Miyazaki, Japan, May 7-12, 2018. European Language Resources Association (ELRA) (2018), http://www.lrec-conf.org/proceedings/lrec2018/summaries/329.html

[29] Omdahl, B.L.: Cognitive appraisal, emotion, and empathy. Psychology Press (2014)
[30] Ortony, A.: Are emotion metaphors conceptual or lexical? Cognition and Emotion **2**(2), 95–104 (1988), publisher: Taylor & Francis
[31] Paiva, A., Leite, I., Boukricha, H., Wachsmuth, I.: Empathy in virtual agents and robots: A survey. ACM Trans. Interact. Intell. Syst. **7**(3) (sep 2017). https://doi.org/10.1145/2912150, https://doi.org/10.1145/2912150
[32] Plutchik, R.: A general psychoevolutionary theory of emotion. In: Theories of emotion, pp. 3–33. Elsevier (1980)
[33] Preston, S.D., De Waal, F.B.: Empathy: Its ultimate and proximate bases. Behavioral and brain sciences **25**(1), 1–20 (2002)
[34] Raboteg-Saric, Z., Hoffman, M.: Empathy and moral development: Implications for caring and justice. Contemporary Sociology **30**, 487 (09 2001). https://doi.org/10.2307/3089337
[35] Revina, I.M., Emmanuel, W.S.: A survey on human face expression recognition techniques. Journal of King Saud University-Computer and Information Sciences **33**(6), 619–628 (2021)
[36] Rizzolatti, G., Craighero, L.: Mirror neuron: a neurological approach to empathy. In: Changeux, J.P., Damasio, A.R., Singer, W., Christen, Y. (eds.) Neurobiology of Human Values. pp. 107–123. Springer Berlin Heidelberg, Berlin, Heidelberg (2005)
[37] Roseman, I.J.: Appraisal determinants of emotions: Constructing a more accurate and comprehensive theory. Cognition & Emotion **10**(3), 241–278 (1996), publisher: Taylor & Francis
[38] Russell, J.A.: A circumplex model of affect. Journal of personality and social psychology **39**(6), 1161 (1980), publisher: American Psychological Association
[39] Scherer, K.R.: Criteria for emotion-antecedent appraisal: A review. Cognitive perspectives on emotion and motivation pp. 89–126 (1988), publisher: Springer
[40] Seo, S.H., Geiskkovitch, D., Nakane, M., King, C., Young, J.E.: Poor thing! would you feel sorry for a simulated robot? proceedings of the tenth annual ACM/IEEE international conference on Human-Robot interaction pp. 125–132 (2015)
[41] Shaver, P., Schwartz, J., Kirson, D., O'connor, C.: Emotion knowledge: further exploration of a prototype approach. Journal of Personality and Social Psychology **52**(6), 1061 (1987)
[42] Stephan, W.G., Finlay, K.: The role of empathy in improving intergroup relations. Journal of Social issues **55**(4), 729–743 (1999)
[43] Strayer, J.: 10 affective and cognitive perspectives on empathy. Empathy and its development p. 218 (1990)
[44] Stueber, K.: Empathy, mental dispositions, and the physicalist challenge (2009)
[45] Wellman, H.M.: The child's theory of mind. The MIT Press (1992)