Towards more humane machines: Creating Emotional Social Robots

Ana Paiva Samuel Mascarenhas Sofia Petisca
Filipa Correia Patrícia Alves-Oliveira

June 15, 2017

1 Introduction

Robots are entering not only our workplace but also our homes. Research in human-robot interaction (HRI) is growing exponentially, with many systems and studies evaluating the acceptance of robots in different contexts and among different populations. Robots are now perceived as machines that not only will support humans in specific tasks, but will also complement them in activities that humans cannot perform. As such, robots will have to act autonomously, performing complex tasks in an intelligent way, as well as be able to interact with humans, adapt to them and perform activities together. One of the challenges that AI (Artificial Intelligence) and robotics face nowadays is how to create social AI and social robotics that interact with humans in an engaging, natural, and most importantly *humane* way, recognizing and respecting human values and social norms. To do that, researchers must develop new models, new algorithms and new techniques that will endow our machines with emotional and social competencies. This requires an understanding of people and their goals and emotions as well as the surrounding context, including the social context.

The advent of robotics and Artificial Intelligence (AI) techniques is also raising serious ethical concerns in our society. When the power of decisions is delegated into machines, questions concerning the morality of such decisions must be questioned. Robots should not only be able to estimate what actions are instrumentally "good" for their goals, but they should also be able to distinguish what actions are morally "good" from those that are morally "bad".

We argue that the question of morality in social robots can be partially answered through their capacity for empathy. As a biological force that makes humans care for one another, empathy has had a fundamental role in the survival and prosperity of the human race. Going into the future, it is imperative that robots are programmed to have empathy so they are endowed with a human-like urge to look after other people. In fact, robots can have a form of empathy that can even be more conducive to human prosperity as it can avoid the pitfalls of being modulated by factors such as proximity or in-group partiality.

Essentially, for robots to be able to have successful human-robot interactions they must be endowed with emotional processing, as well as be able to respond emotionally to the humans and adapt to their environmental and moral context. We will overview this challenging research area and present some application cases borrowed from educational and entertainment robotics, where empathy, emotion sharing, and group-based emotions have been explored to deepen the relation between humans and social robots. We then conclude with a discussion of how these emotional processes are a fundamental piece of the puzzle that is the creation of robots with the capacity for complex moral reasoning.

2 Emotional Processes in Social Robots

The importance of emotions in social robots is fully inspired by the role emotions have in human behaviour. Not only emotions were identified as a mean to form and maintain social relationships, as they may also establish a social position through intentions or decisions [33]. Therefore, including these notions from the social psychology into the social robotics field provides researchers ways of understanding and improving the interaction between humans and robots. Creating computational models of emotions was indeed pointed as having critical importance since one of the first social robots, Kismet [6]. The author claimed that the perception and the consequent interaction people have towards a social robot is shaped by the observable behaviour and the manner in which the robot reacts and responds to people.

Further studies have extended this belief in more detail by analysing which emotional processes can contribute to the successful creation of social robots. For instance, emotional expressions (through facial expressions of happiness, sadness and anger) had a positive influence on people's enjoyability when compared to non-emotional expressions [4]. Additionally, another user study with an educational scenario also revealed how the presence of emotional expressions can positively convey the learning performance of children by a robotic tutor [45]. Some of the main characteristics of this robotic tutor, regarding its social supportive behaviour, were non-verbal feedback, attention building, empathy, and communicativeness. Finally, among different extensions of these emotional processes, engagement gestures as tracking a human partner's face were reported as an appropriate ability for a robot that converses and collaborates [47]. This ability provides a way of maintaining the connection with one another, at the same time it constitutes a sophisticated and smoother interaction that can be perceived as more reliable.

The previous examples illustrate how social robots can enter our lives in several domains as domestic, entertainment, healthcare, education, etc. However, independently of being our companions, tutors, or simple operators, the collaboration and communication with them will be certainly required. Furthermore, the quality of this communication can even be more effective towards successful interactions between humans and robots, which may be achieved by introducing different emotional processes into the communication and collabo-

ration skills of the robots (as also seen in the previous examples). One way to do this, can be by trying to embed a kind of artificial empathy in the robot, since empathy constitutes one of the most relevant emotional processes able of promoting the relationship among one another. Previous findings also encourage this idea, where a robot with empathic behaviours was perceived as friendlier and, therefore, able to foster an improved relationship as a companion [32].

The definition of empathy has evolved for the past years among the different scopes, interpretations, or possible theories it may include. Nevertheless, Cuff et al. [13] have recently reviewed this concept and proposed a definition that provides an answer to most inconsistencies from previous findings. It includes both affective and cognitive aspects since it produces an emotional response upon the understanding of an emotional stimulus whose source is not one's own. Moreover, the emotional response consists of an emotion that might be or not followed by an emotional behaviour, and although it is automatically elicited, this empathic process can later be controlled, reframed, suppressed, or even modified. Additionally, the resulting emotion on the observer is as similar as possible to the target's own emotion depending on the empathic accuracy by the observer, which will inevitably shape its perception of the situation. A final relevant consideration in their careful definition of empathy refers to the source of the emotional stimulus, which does not have to be another person and may generally be any element containing an emotionally-laden stimulus. For instance, it can be a fictional or imaginary person as in books or animated films, where there are no living entities and people still respond emotionally. Consequently, this last evidence also bridges the study of empathy in social robotics where previous findings strongly suggest that people conceive anthropomorphic models of interactive robots [28].

Another relevant consideration that Cuff et al. has exposed while reviewing the concept of empathy [13] refers to the Self-Other distinction during the empathic process. Some authors argue that for a complete awareness that the emotional experience comes from an external source, the observer should maintain a clear self-other distinction. Alternatively, neuroscience findings have shown a partial overlapped between the brain activations of someone taking the Self perspective and the Other perspective, which is coherent with the opposite Self-Other merging theory [24]. Decety and Sommerville (2003) have even mentioned that without some Self-Other merging the understanding the other's emotion would be difficult, compromising the cognitive empathy [15]. Moreover, the importance of referring Self-Other merging is naturally extended to the topic of social identification [49], where social coordination is facilitated [20], and the social bonds are fostered by including the other in one's own mental self-representation. This evidence is also definite as Tropp and Wright (2001) have demonstrated that the degree to which the in-group is included in the Self can measure the ingroup identification [50].

All the discussed ideas sustain the relevance of endowing social robots with empathic behaviours, as well as the ability to elicit empathy in humans during their interactions. Therefore, it is important to embed a robot with more bonding characteristics, to explore how an artificial cognitive empathy can be

implemented in robots and what effects we hope to produce from this implementation.

We know that since our childhood until we die we live in a social context, with day by day social interactions. Our own emotions propel us towards interacting with others and sharing them is a big part of our intimate relationships. So, could we improve the relationship between a human and a robot if it shared its emotions? Rimé claims that "the social sharing of emotion occurs in discourse, when individuals communicate openly with one or more persons about the circumstances of the emotion-eliciting event and about their own feelings and emotional reactions"[41, page 19] and presents support from studies showing how people after an emotional event try to share it with someone else. This in turn, contributes to a sense of closeness in the relationship and greater intimacy, which is something we look for in our most meaningful relationships. Therefore, we believe that giving a social robot the capacity to share its emotions can be an important implementation to facilitate human-robot interactions. This way, it may make the human much more involved in the interaction.

As empathy does not lead always to behavioural outcomes, it is not possible to guarantee that by eliciting empathy in humans towards robots, this will always reflect in a different behaviour towards the robot, or even a prosocial behaviour towards it. Still, we think that by endowing the robot with human-like characteristics and empathic capabilities, it can better do its role (e.g. supporting a student in class by understanding when it is having difficulties and helping him), and it can make the interaction more meaningful. We will now present some use case scenarios we developed that support this idea.

3 First Steps: Case Studies With Humane Social Robots

We are on the brink of a revolution were machines are compelled to act in an empathic manner, capable of understanding and sharing affective experiences that resonate with us. In this section we will describe three examples where social robots were programmed with the type of emotional processes we previously described.

The three examples we will explore are: an empathic social robot that tutors teenagers and was developed in the EMOTE project; a social robot that shares its emotions with its users; and a social robot that plays a team-based card game, while expressing group-based emotions.

3.1 EMOTE: Creating Robotic Tutors With Empathy

One of the most important qualities of a good teacher is *empathy*. For a teacher to be accepted, to motivate, to engage, and to fully be inspirational, empathic qualities are needed. Perceiving, listening attentively, motivating, encouraging, looking into the eyes, placing oneself into the learner's shoes are necessary

characteristics that all influential teachers have. Although there has been significant work has been devoted to the design of artificial tutors with human capabilities with the aim of helping increase the efficiency achieved with a human instructor, these systems often lack the personal, empathic and human elements that characterise a traditional teacher and fail to engage and motivate students in the same way a human teacher does. *Empathy and engagement*, abilities that are key to influence students' learning, are often forgotten when such learning systems are created. To address this issue, research on intelligent tutoring systems has recently shifted towards a more learner-centric approach to endow artificial tutors with the ability to perceive the emotions experienced by learners and incorporate these into pedagogical strategies to build more effective computer-based learning systems [8]. Examples include determining the appropriateness of affective interventions by means of empathic strategies as a response to a learner's emotional state [42]. Recent research on socially intelligent robots shows that robots are increasingly being studied as partners that collaborate and do things with people [7], making the use of robotic platforms as tools for experimental learning more approachable [31].

Based on these recent findings, and aiming to achieve fruitful empathic interactions with learners, the EMOTE project[1] designed, developed and evaluated a new generation of artificial embodied tutors that have perceptive capabilities to engage in empathic interactions with learners in a shared physical space. EMOTE adopted a learner-centric approach, applied to the design of curriculum-driven learning scenarios, where personalised and pedagogically sound learning strategies were employed by the tutor in order to successfully adapt to the learner's engagement and progress in the learning task. Towards this end, two learning scenarios were developed related with geography: a map-reading task and an activity to learn about sustainability. The EMOTE project adopted personalised strategies to generate tutor interventions targeted to a specific user and their needs. The tutor's interventions took place at the *pedagogical and empathic level* and these interventions occurred in both learning scenarios of the project. The main difference between the learning scenarios of EMOTE concerns the number of learners that is included in the task. Therefore, the map-reading task was performed at an *individual* level, in which one student solved the task guided by a robotic tutor; while the sustainability activity is a *collaborative* one in which two students and a robot (see Figure 1) need to build a sustainable city together using a serious game called EnerCities [29].

In a long term study with the individual map reading activity, Serholt and Barendregt (2016), have explored children's social engagement to the empathic robotic tutor [46]. This was performed by analysing their behavioural reactions to socially significant events initiated by the robot, such as greeting, feedback/praise and when questioning learners. The results seem to show that children reveal behaviours that indicate social engagement using a range of communicative channels. Thus, while gaze towards the robot is the most common indication for all types of social events, verbal expressions and nods are the

[1] http://www.emote-project.eu/

Figure 1: Sustainability Scenario in EMOTE.

most common for questions, and smiles appear usually after positive feedback. In conclusion, this study shows that the behavioural responses that children express reveal engagement with an empathic social robot, which could either be understood as a developing social bond (e.g., [36] [26]) or a reaction to perceiving the robot as a social actor [35] [48].

Role assignment is a way to organize interpersonal encounters and can result in uncertainty decrease when facing a novel interaction with someone we just met, or even to rediscover new roles within previous relationships [30] [43]. In fact, most people have never interacted with a robot and specially in EMOTE, most children have never seen a robot. As robots being created to fulfil specific roles, such as of a robotic tutor, it would then be expected that users too would assign the same role to that robot. Alves-Oliveira and collaborators (2016) have studied how learners assign roles to an educational empathic robot whose role is established from the beginning of the interaction [2]. This study compares the role that children assign to an empathic robot who they have been interacting with in the context of the collaborative sustainability task for a period of 2 months (long term interaction); and also compares the role that children assign to a non-empathic robot in a short-term interaction with the same learning activity. The results show that before knowing the empathic robot well, children attribute the role of a friend and at the end of the long term interaction most of children consider it a classmate. The role shift in children seems to be adapted to the role of the robot, since they interact with it exclusively to learn. In the second study, in which they have compared the role that children assign to a non-empathic robot (compared to an empathic one, in a short term interaction), we can see that they think the empathic robot is a friend, while the non-empathic robot is perceived as a tutor. This can change the learning process, in which the results suggest that children can feel more close in the interaction when learning

with an empathic robot tutor compared to a non-empathic robot tutor.

In another study, it was investigated the expectations and satisfaction of learners before and after having interacted with the empathic robotic tutor in a school classroom. Students interacted with the empathic robot in the collaborative scenario of sustainability and the results seem to show that they had high expectations towards the robot. When asked to rate their satisfaction level after the experience of the learning interaction with the empathic tutor, their satisfaction levels were also considerably high. Thus, this study demonstrates that children feel satisfied to learn with an empathic robotic tutor [1].

As a conclusion remark, it seems important to reflect on the different roles that robots can acquire in our societies. Some of these roles with require more sophisticated and somewhat human-like characteristics, such as the ability to show empathy. In EMOTE project, an empathic robotic tutor was created and developed to teach children about sustainability and map reading abilities. The results are promising but more research is needed. A special focus should also be given to ethical and moral concerns of robots in the schools.

3.2 Robots That Share Their Emotions

It is undeniable at this point that emotions are a big part of human lives and will be important to exist in human-robot interactions. It will be important for robots to be able to ascertain from the others their emotions as well as share their own, since a major step for meaningful interactions occurs when people are able to communicate to others how they feel and are able to understand other's social signals (e.g. how he/she is feeling).

Starting from this idea, we aimed to explore emotional sharing in human-robot interactions (for more information see [39]). For this, we approached its role in a competitive setting, where people had to play individually a version of the dots and boxes game [5] called Coins and Strings against a social robot (see Figure 2). In this game, a set of coins are attached one another through strings and each turn a participant can only take one string from the board. If a string releases a coin (from other strings attached to it) the player wins that coin for himself and gets an extra turn. The game ends when all coins have been collected, the player with more coins, wins the game. Participants would each play five boards, with its difficulty increasing. The robot would autonomously play the game and comment on the actions taken with the emotional sharing behaviour being manipulated. Participants would either be in the Sharing Condition (the robot would do small talk and share its emotions towards the game at the end of each board game) or the No Sharing Condition (the robot would only do small talk- e.g. "This is going to be a hard game").

For this to be possible, the FAtiMA Emotional Agent Architecture [16] was used, integrated with Thalamus Framework [40] which in turn, connected to the game (Unity3D) and the Emys robot [25]. The FAtiMA would update its own internal state of the game and consequently the robot's emotional state towards the game actions. This emotional state would trigger emotion expression that depending on its intensity could also trigger speech, for example if an action

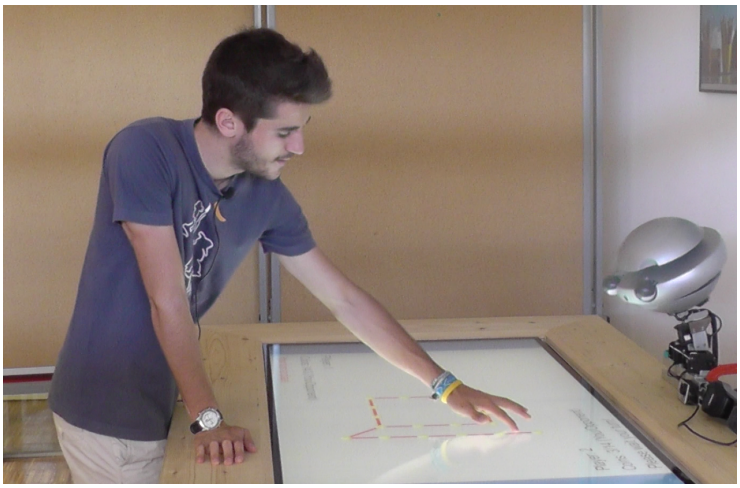

Figure 2: Experimental setting for the emotional sharing scenario

made Emys very happy with it, like winning, it would display a joyful facial expression and at the same time say "Great!". The robot cognitive behaviour was done by FAtiMA with an AI for gameplaying, integrating a standard Minimax Algorithm [44] to decide the best move to play and the desirability of an event. This would generate different emotions according to OCC Theory of Emotions [37]. On the other hand, in the Sharing Condition the FAtiMA agent had the goal of sharing a summary of the more relevant emotions elicited in that board (e.g. "Several minutes ago, I wanted to win the game which made me feel frightened. Afterwards I played an important move. I was feeling really glad").

We were interested in seeing if the presence of emotional sharing from the robot to the user, would change the user perception of the robot (making it more humanlike instead of machinelike) and the level of closeness felt towards it. We found no significant differences between the user's perceptions of the robot in both conditions, suggesting that participants did not felt the robot more humanlike by having the emotional sharing component. In order to test our robot performance, we observed that participants in both conditions considered the robot to be intelligent and competent in the task (which was also supported by only having 4 participants that were able to beat Emys in the game). Regarding the level of closeness felt towards the robot, there were also no significant differences between the conditions. This seems to suggest that the emotional sharing behaviour is not having the effect we were hoping for. But taking into account that most of the participants were loosing the games against Emys we wondered if the robot competence could have been having some kind of effect towards the emotional sharing behaviour. With Emys winning almost all the games, it means it was also making victory summaries at the end of each board game. Participants might have more easily seen the robot as a machine because of that, a machine that always beats humans, nullifying the effect that

we aimed for with the emotional sharing behaviour in this specific context.

In order to clarify this, we replicated our study [38] but this time we also manipulated the robot competence: in one study the robot would have a high competence in the game, in the other a low competence. In each study there would be two conditions (Sharing or No Sharing) as in the previous study reported. Results confirmed our manipulation of competence level and participants reported feeling more affect in the Sharing condition comparing to the No sharing one, but again, the sharing behaviour did not show any difference in the perception of the robot. When taking into account only the Competence level of the robot, participants gave more attention to Emys when it had a lower competence in the game (even though Emys spoke more in the high competence study) and reported higher scores of empathic concern (other-oriented feelings of concern for others) in Emys.

Taken this all together, we can conclude that for a competitive setting, emotional sharing does not have the desired effect. The nature of this task, surpasses the effect of hearing the other emotional state. Instead of bringing them together, emotional behaviour looses its effect, with participants being concentrated in winning the game. We still believe emotional sharing can be an important component of human-robot interactions but the context were it is implemented, strongly defines the effect it can have. Furthermore, emotional disclosure is also more normally accepted between females, comparing to males [17]. In both studies we had more males than females, this can also add to our results as a strange behaviour to occur in a short interaction. Suggesting that emotional sharing behaviour might be better implemented in a collaborative scenario and perhaps in longer interactions.

3.3 Robotic Partners With Group-Based Emotions

Appraisal theories argue that emotional experiences are triggered by how individuals subjectively evaluate the events they perceive. So, for instance, when two persons play a card game against one another, they are likely to have contrasting emotions during the game, provided they are both invested enough in winning. More precisely, when one is happy for being ahead, the other is likely to be sad or angry for being behind. Most computational models of emotions are able to capture this notion quite satisfactorily. However, what if there are two teams playing against each other rather than just two individuals? Then, members of each team are now likely to also experience emotions based on how the team itself is being affected by the events that happen. Such emotions are referred to as group-based emotions [27]. While this may, at first, seem incongruent with the notion of individual subjectivity put forth by appraisal theories, it is possible to reconcile the two if we consider the process of group identification, based on social identity theory [51]. This is a process that makes the individual see himself or herself as a member of a social identity that is contextually salient. When this happens, the subjectivity of the emotional appraisal process is based on the social identity that is being adopted by the individual at the moment, rather than the individual's personality [21]. It is through this

process that one can explain why someone can feel ashamed by immoral actions conducted by members of their ingroup that they themselves did not commit.

As robots and humans start working together in teams to solve problems it is important that the emotional apparatus of social robots is capable of simulating an appraisal process that is also able to encompass social identities, in order to generate group-based emotions. The potential of doing so is that it will enable the creation of robots that are less self-centred and, consequentially, more trustworthy. So far, most research on human-robot interaction has typically focused on dyadic interactions, i.e., a single robot interacting with a single human. However, there is a current growing trend that focuses on more complex scenarios that involve multiple robots interacting with multiple people [9, 19, 18]. These more complex scenarios are better suited to test the potential impact of group-based emotions.

Our first case-study that focuses on group-based emotions in social robots is based on a tabletop card game, named *Sueca*[2]. This is a very popular Portuguese card game that is played by four players that are divided in two teams. The first version of this scenario was developed in the PArCEIRO project[3], with the aim of understanding how much people would trust a robot partner they had interacted before compared to a robot partner they never interacted with. An experiment was conducted where several groups of three participants played *Sueca* with the EMYS robot for about an hour (see Figure 3), with one of them being randomly assigned to be the robot's partner. The obtained results showed evidence that the scenario was able to increase how much trust was attributed to the robot, but only for the group of participants that had a previous interaction with it [12].

Figure 3: Sueca scenario in the trust experiment.

In the aforementioned experiment, the EMYS robot is able not only to decide which card to play during its turn, but also to subjectively evaluate the

[2]https://en.wikipedia.org/wiki/Sueca (card game)
[3]http://gaips.inesc-id.pt/parceiro/

state of the game and trigger appropriate emotional responses that are conveyed both verbally and non-verbally. Similar to the emotion sharing scenario, the emotional appraisal process of the robot is determined by the use of the emotional agent architecture FAtiMA [16]. However, for being able to generate group-based emotions, this appraisal process of FAtiMA was modified to have a group identification mechanism, which makes the robot consider itself responsible for not only the cards it chooses to play but also for the cards that are played by its human partner. In a sense, it is as if the robot considers its partner as an extension of itself. Consequentially, this means that the robot is able to feel group-based pride when its partner does a great play and also feel group-based shame when the partner does a terrible move. In contrast, without the group identification bias in the appraisal process, the robot will instead feel admiration or reproach for the good or bad moves of its partner. To test this new appraisal process, we plan to conduct an experiment using the same *Sueca* scenario but with two human players, each playing with their own robotic partner. One of the robots will have the group-based appraisal process activated and the other robot will not. Our main hypothesis for this experiment is that the subjects who play with the robot that has group-based emotions will have a significantly higher identification with his or her team and consequentially a higher intrinsic motivation associated to the task.

4 Future Directions: Exploring Empathy for Creating Moral Robots

The use cases that were previously described serve to illustrate the importance of endowing social robots with different types of emotional capabilities for them to be successful in helping or collaborating with people. But going into the future, as autonomous robots become more pervasive in our society, and start to be able to interact with people, and make decisions that impact our physical world, the public in general and the research community in particular needs to address important questions that concern the morality and ethics of human-robot behaviour. One must consider that robots will have the power of making decisions that will implicate humans and can range from moral decisions in a military or government context for example, to more simpler decisions as deciding if that student work reflects in a good grade or not.

A common starting point in the discussion of moral robots are the three "Laws of Robotics", which were introduced in the classical literary work of Asimov [3]. Briefly, the first law states that robots should never harm a human being, the second law dictates that robots should always obey human orders unless these orders violate the first law and, finally, the third law states that robots should protect their own existence, except if it conflicts with either the first or the second law. As they have been referenced in books and films multiple times, these laws have shaped society's expectations of what it means for a robot to act in a moral manner. However, as hinted even in Asimov's stories, these laws

are an oversimplification and upon closer inspection, several researchers have pointed out many practical and theoretical issues concerning their implementation [10, 11]. To give an example, the first law is incompatible with the notion that a robot designed to rescue people would need to, sometimes, temporarily harm a person in order to save his or her life.

In response to the issues found in Asimov's laws, some authors have proposed an alternative set of rules [10], but others have criticized the notion altogether of having the morality of robots based just on a fixed ruleset [11, 52]. Indeed, one must be careful with the limitations of applying conscious reasoning to morality. While we are certainly capable of reaching moral conclusions starting from a set of explicit principles, it does not mean that such principles are the ultimate source of human morality. To further illustrate this point, psychopaths are able to follow rules but they lack the ability to feel that something is morally wrong.

The 18^{th} century philosopher David Hume was the first to bring forth the idea that emotions are the basis for human morality. In his view, emotions are responsible for triggering moral judgements, which are only then explained and rationalized through reason. More recently, neuroscientists have found empirical support for Hume's idea that moral judgement involves emotional engagement [22]. Following Hume's footsteps, Hoffman developed a theory of morality that places empathy at its core. In his words, empathy is "the spark of human concern for others, the glue that makes social life possible." [23, page 23]. One of its most important aspects for moral reasoning, is that empathy prevents us from acting out moral violations that would make rational sense, from an utilitarian perspective. Empathy is thus not only beneficial for improving the ability of social robots to engage with humans, as was previously illustrated, but it is also an important component to consider when trying to answer the question of what kind of capabilities should robots have to be moral.

Still, even though empathy plays a crucial role in human morality, it also has its limitations, and sometimes may even interfere with moral decisions due to its partiality [14]. More specifically, people feel more empathic distress for in-group members and towards victims that are physically present compared to those that are out of sight. Nevertheless, without empathy there would not be any empathic distress at all. Moreover, it is possible that the artificial empathy that is encoded in social robots is able to overcome current limitations of human empathy, while keeping its social benefits.

Finally, asides from empathy, another key competence that will be required for social robots to be moral is the ability to reason and communicate about social norms [34]. The communication aspect is particularly important if we consider that, no matter how thorough the robot designers are, it will be impossible for them to specify and program the set of all relevant norms into the robot's architecture. Instead, robots should be able to evolve their set of norms by observation and communication with people. When robots become able to learn from us, we will be able to learn from them as well.

References

[1] P. Alves-Oliveira, T. Ribeiro, S. Petisca, E. Di Tullio, F. S. Melo, and A. Paiva. An empathic robotic tutor for school classrooms: Considering expectation and satisfaction of children as end-users. In *International Conference on Social Robotics*, pages 21–30. Springer, 2015.

[2] P. Alves-Oliveira, P. Sequeira, and A. Paiva. The role that an educational robot plays. In *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*, pages 817–822. IEEE, 2016.

[3] I. Asimov. Runaround. *Astounding Science Fiction*, 29(1):94–103, 1942.

[4] C. Bartneck. Interacting with an embodied emotional character. In *Proceedings of the 2003 international conference on Designing pleasurable products and interfaces*, pages 55–60. ACM, 2003.

[5] E. R. Berlekamp. *The Dots and Boxes Game: Sophisticated Child's Play*. AK Peters/CRC Press, 2000.

[6] C. Breazeal. Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies*, 59(1):119–155, 2003.

[7] C. Breazeal. Role of expressive behaviour for robots that learn from people. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364(1535):3527–3538, 2009.

[8] W. Burleson. *Affective learning companions: strategies for empathetic agents with real-time multimodal affective sensing to foster meta-cognitive and meta-affective approaches to learning, motivation, and perseverance*. PhD thesis, Massachusetts Institute of Technology, 2006.

[9] W.-L. Chang, J. P. White, J. Park, A. Holm, and S. Šabanović. The effect of group size on people's attitudes and cooperative behaviors toward robots in interactive gameplay. In *RO-MAN, 2012 IEEE*, pages 845–850. IEEE, 2012.

[10] R. Clarke. Asimov's laws of robotics: implications for information technology-part i. *Computer*, 26(12):53–61, 1993.

[11] M. Coeckelbergh. Moral appearances: emotions, robots, and human morality. *Ethics and Information Technology*, 12(3):235–241, 2010.

[12] F. Correia, P. Alves-Oliveira, N. Maia, T. Ribeiro, S. Petisca, F. S. Melo, and A. Paiva. Just follow the suit! trust in human-robot interactions during card game playing. In *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*, pages 507–512. IEEE, 2016.

[13] B. M. Cuff, S. J. Brown, L. Taylor, and D. J. Howat. Empathy: a review of the concept. *Emotion Review*, 8(2):144–153, 2016.

[14] J. Decety and J. M. Cowell. Friends or foes: Is empathy necessary for moral behavior? *Perspectives on Psychological Science*, 9(5):525–537, 2014.

[15] J. Decety and J. A. Sommerville. Shared representations between self and other: a social cognitive neuroscience view. *Trends in cognitive sciences*, 7(12):527–533, 2003.

[16] J. Dias, S. Mascarenhas, and A. Paiva. Fatima modular: Towards an agent architecture with a generic appraisal framework. In *Emotion Modeling*, pages 44–56. Springer, 2014.

[17] K. Dindia and M. Allen. Sex differences in self-disclosure: a meta-analysis. *Psychological bulletin*, 112(1):106, 1992.

[18] F. Eyssel and D. Kuchenbrandt. Social categorization of social robots: Anthropomorphism as a function of robot group membership. *British Journal of Social Psychology*, 51(4):724–731, 2012.

[19] M. R. Fraune, S. Kawakami, S. Sabanovic, P. R. S. De Silva, and M. Okada. Three's company, or a crowd?: The effects of robot number and behavior on hri in japan and the usa. In *Robotics: Science and Systems*, 2015.

[20] A. D. Galinsky, G. Ku, and C. S. Wang. Perspective-taking and self-other overlap: Fostering social bonds and facilitating social coordination. *Group Processes & Intergroup Relations*, 8(2):109–124, 2005.

[21] A. Goldenberg, E. Halperin, M. van Zomeren, and J. J. Gross. The process model of group-based emotion: Integrating intergroup emotion and emotion regulation perspectives. *Personality and Social Psychology Review*, 20(2):118–141, 2016.

[22] J. D. Greene, R. B. Sommerville, L. E. Nystrom, J. M. Darley, and J. D. Cohen. An fmri investigation of emotional engagement in moral judgment. *Science*, 293(5537):2105–2108, 2001.

[23] M. L. Hoffman. *Empathy and moral development: Implications for caring and justice*. Cambridge University Press, 2001.

[24] P. L. Jackson, E. Brunet, A. N. Meltzoff, and J. Decety. Empathy examined through the neural mechanisms involved in imagining how i feel versus how you feel pain. *Neuropsychologia*, 44(5):752–761, 2006.

[25] J. Kędzierski, R. Muszyński, C. Zoll, A. Oleksy, and M. Frontkiewicz. EmysâĂŤemotive head of a social robot. *International Journal of Social Robotics*, 5(2):237–249, 2013.

[26] J. Kennedy, P. Baxter, E. Senft, and T. Belpaeme. Social robot tutoring for child second language learning. In *Human-Robot Interaction (HRI), 2016 11th ACM/IEEE International Conference on*, pages 231–238. IEEE, 2016.

[27] T. Kessler and S. Hollbach. Group-based emotions as determinants of ingroup identification. *Journal of Experimental Social Psychology*, 41(6):677–685, 2005.

[28] S. Kiesler and J. Goetz. Mental models and cooperation with robotic assistants chi 2002 extended abstracts, 2002.

[29] E. Knol and P. W. De Vries. Enercities: educational game about energy. *Proceedings CESB10 Central Europe towards Sustainable Building*, 2010.

[30] T. Lan, L. Sigal, and G. Mori. Social roles in hierarchical models for human activity recognition. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 1354–1361. IEEE, 2012.

[31] I. Leite, G. Castellano, A. Pereira, C. Martinho, and A. Paiva. Modelling empathic behaviour in a robotic game companion for children: an ethnographic study in real-world settings. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pages 367–374. ACM, 2012.

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

[33] M. Lewis, J. M. Haviland-Jones, and L. F. Barrett. *Handbook of emotions*. Guilford Press, 2010.

[34] B. F. Malle. Integrating robot ethics and machine morality: the study and design of moral competence in robots. *Ethics and Information Technology*, 18(4):243–256, 2016.

[35] C. Nass and Y. Moon. Machines and mindlessness: Social responses to computers. *Journal of social issues*, 56(1):81–103, 2000.

[36] K. Oh and M. Kim. Social attributes of robotic products: observations of child-robot interactions in a school environment. *International Journal of Design*, 4(1):45–55, 2010.

[37] A. Ortony, G. L. Clore, and A. Collins. *The cognitive structure of emotions*. Cambridge university press, 1990.

[38] S. Petisca, J. Dias, P. Alves-Oliveira, and A. Paiva. Emotional sharing behavior for a social robot in a competitive setting. In *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*, pages 730–735. IEEE, 2016.

[39] S. Petisca, J. Dias, and A. Paiva. More social and emotional behaviour may lead to poorer perceptions of a social robot. In *International Conference on Social Robotics*, pages 522–531. Springer, 2015.

[40] T. Ribeiro, E. Di Tullio, L. J. Corrigan, A. Jones, F. Papadopoulos, R. Aylett, G. Castellano, and A. Paiva. Developing interactive embodied characters using the thalamus framework: A collaborative approach. In *International Conference on Intelligent Virtual Agents*, pages 364–373. Springer, 2014.

[41] B. Rimé. Emotion elicits the social sharing of emotion: Theory and empirical review. *Emotion Review*, 1(1):60–85, 2009.

[42] J. Robison, S. McQuiggan, and J. Lester. Evaluating the consequences of affective feedback in intelligent tutoring systems. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pages 1–6. IEEE, 2009.

[43] L. D. Ross, T. M. Amabile, and J. L. Steinmetz. Social roles, social control, and biases in social-perception processes. *Journal of personality and social psychology*, 35(7):485–494, 1977.

[44] S. Russell and P. Norvig. Artificial intelligence: a modern approach. 1995.

[45] M. Saerbeck, T. Schut, C. Bartneck, and M. D. Janse. Expressive robots in education: varying the degree of social supportive behavior of a robotic tutor. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1613–1622. ACM, 2010.

[46] S. Serholt and W. Barendregt. Robots tutoring children: Longitudinal evaluation of social engagement in child-robot interaction. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*, page 64. ACM, 2016.

[47] C. L. Sidner, C. Lee, C. D. Kidd, N. Lesh, and C. Rich. Explorations in engagement for humans and robots. *Artificial Intelligence*, 166(1-2):140–164, 2005.

[48] L. Takayama. Perspectives on agency interacting with and through personal robots. In *Human-Computer Interaction: The Agency Perspective*, pages 195–214. Springer, 2012.

[49] M. Tarrant, R. Calitri, and D. Weston. Social identification structures the effects of perspective taking. *Psychological Science*, page 0956797612441221, 2012.

[50] L. R. Tropp and S. C. Wright. Ingroup identification as the inclusion of ingroup in the self. *Personality and Social Psychology Bulletin*, 27(5):585–600, 2001.

[51] J. C. Turner and P. J. Oakes. The significance of the social identity concept for social psychology with reference to individualism, interactionism and social influence. *British Journal of Social Psychology*, 25(3):237–252, 1986.

[52] W. Wallach and C. Allen. *Moral machines: Teaching robots right from wrong*. Oxford University Press, 2008.