Empathic Robots for Long-term Interaction

Evaluating Social Presence, Engagement and Perceived Support in Children

**Iolanda Leite · Ginevra Castellano · André Pereira ·
Carlos Martinho · Ana Paiva**

Abstract   As a great number of robotic products are entering people's lives, the question of how can they behave in order to sustain long-term interactions with users becomes increasingly more relevant. In this paper, we present an empathic model for social robots that aim to interact with children for extended periods of time. The application of this model to a scenario where a social robot plays chess with children is described. To evaluate the proposed model, we conducted a long-term study in an elementary school and measured children's perception of social presence, engagement and social support.

Keywords   Empathy · Social presence · Engagement ·
Social support · Long-term interaction

1 Introduction

Robots are becoming part our daily routines as commercial products, for example by helping us with tedious chores at home or augmenting the presence of colleagues working remotely. The application domains where these robots can assist people and improve their quality of life are innumerous, from health-care assistants for the elderly [41] to learning companions for children [1]. Breazeal [5] argues that for robots to become part of our lives, they should be able to communicate with people in similar ways people interact with

I. Leite (✉) · A. Pereira · C. Martinho · A. Paiva
INESC-ID and Instituto Superior Técnico, Technical
University of Lisbon, Oeiras, Portugal
e-mail: iolanda.leite@ist.utl.pt

G. Castellano
HCI Centre, School of Electronic, Electrical and Computer
Engineering, University of Birmingham, Birmingham,
United Kingdom

each other. This becomes even more relevant when robots are expected to interact with us for long periods of time, after the "novelty effect" fades away. For these reasons, research on social robots capable of engaging users for extended periods of time has received increasingly more attention in the recent years [24]. Previous studies have shown that robotic companions are able to keep children engaged in single-interaction sessions [18]. However, despite the remarkable empirical research in this field, there is a fundamental question that remains unanswered: which social capabilities should robots be endowed with to better engage users over repeated interactions?

In this paper, we aim to take a small step towards answering this question by exploring the role of empathy in long-term interaction between users, particularly children, and social robots. We argue that artificial companions capable of behaving in an empathic manner will be more successful at establishing and maintaining a positive relationship with users in the long-term. Hoffman defines empathy as "an affective response more appropriate to someone else's situation than to one's own" [17, p. 4]. To behave empathically, social robots need to understand some of the user's affective states and respond appropriately. However, empathic responses can go beyond facial expressions (e.g., mimicking the other's expression): they can also foster actions taken to reduce the other's distress, such as social supportive behaviours [11]. In fact, the perception of social support has been linked to positive outcomes in children's mental health and coping with traumatic events [40]. Thus, our aim is to study the effects of a computational model of empathy in the long-term relationship established between the robot and the user. Previous human–robot interaction (HRI) studies showed some positive effects of empathic robots [9,27], but these findings were obtained in studies where users interacted with robots only for a short period of time.

This paper is organised as follows. In the next section, we present some related work in long-term studies with children and interaction with empathic robots. After that, we describe our empathic model that includes socially supportive behaviours, and explain how the proposed model was implemented in a social robot that plays chess with children. We then present the setup, metrics and results of the long-term study conducted to evaluate the proposed model. Finally, we draw some conclusions and implications for future research in this area.

2 Related Work

Most of the existing longitudinal HRI studies with children were carried out in school environments. In this domain, one of the pioneer experiments was performed by Kanda et al. [19] with Robovie. They conducted a trial for two weeks with elementary school Japanese students. The study revealed that Robovie failed to keep most of the children's interest after the first week, although children who kept interacting with the robot after the first week improved their English skills. In a follow-up study [20], Robovie was improved with a pseudo-development mechanism and self-disclosure behaviours. In contrast to the results obtained in the previous experiment, Robovie was capable of engaging children after the second week (although with a slight decay), which the authors attribute to the new capabilities implemented in the robot. With children from a different age group, a longitudinal study where a QRIO robot [39] interacted with toddlers in a day care revealed that toddlers progressively started treating the robot as a peer rather than as a toy, and that they exhibited an extensive number of care-taking behaviours towards the robot. Kozima et al. [21] undertook a similar study to investigate the interaction between toddlers and a Keepon robot designed to interact with children through nonverbal behaviours such as eye contact, joint attention and emotions . The authors report that children's understanding of the robot changed over time, from a mere "moving thing" to a "social agent". In our own previous work [25], we measured perceived social presence towards a robotic chess companion over time. The results suggested that social presence decreased from the first to the last interaction, especially in terms of attentional allocation, and perceived affective and behavioural interdependence, that is, the extent to which users believe that their affective (and overall) behaviour affects and is affected by the robot's behaviour.

As we can see from the works mentioned above, none of the robots used in the long-term studies were programmed with empathic abilities. The only exception comes from the field of virtual agents, where Bickmore and Picard [2] developed Laura, a virtual conversational agent that employs relationship maintenance strategies while keeping track of user's exercise activities. Among other relational behaviours, Laura uses empathy in the attempt to maintain a long-term social-emotional relationships with users. The agent was evaluated in a study where approximately 100 users interacted daily with the exercise adoption system. After 4 weeks, the agent's relational behaviours increased participants' perceptions of the quality of the working alliance (on measures such as liking, trust and respect), when comparing the results to those of an agent without relational capabilities. Users interacting with the relational agent also expressed significantly higher desire to continue interacting with the system.

On the other hand, several authors investigated the effects of empathic social robots in single interaction studies. Cramer et al. [9], for example, investigated how empathy affects people's attitudes towards robots. In their study, two groups of participants saw a 4-minute video with an actor playing a cooperative game with an iCat robot. The experimental manipulation consisted in causing the robot to express empathic behaviour towards the actor in an accurate manner, or incongruent behaviour to the situation. They found a significant negative effect on users' trust in the inaccurate empathic behaviour condition. Conversely, participants who observed the robot displaying accurate empathic behaviours perceived their relationship with the robot as closer. In another study [33], where a robot with the form of a chimpanzee head mimics the user's mouth and head movements, the authors found that most subjects considered the interaction more satisfactory than participants who interacted with a version of the robot without mimicking capabilities. Similar results were found by Gonsior et al. [13], who showed that facial mimicry influences the degree of empathy that a person attributes to a robot. In terms of social support, Saerbeck et al. [34] investigated the effects of supportive behaviours of an iCat robot on children's learning performance and motivation. The results suggest that simple manipulations in the robot's supportiveness, while maintaining the same learning content, increased student's motivation and scores on a language test.

3 An Empathic Model for Social Robots

In this section, we present the main components of our proposed empathic model for supporting long-term human-robot interaction. After that, the scenario where the model was implemented is briefly described (Fig. 1).

3.1 Model Components

The empathic model follows a traditional perception and action loop, and includes the following components:

Fig. 1 Architecture of the empathic model applied to the iCat robot
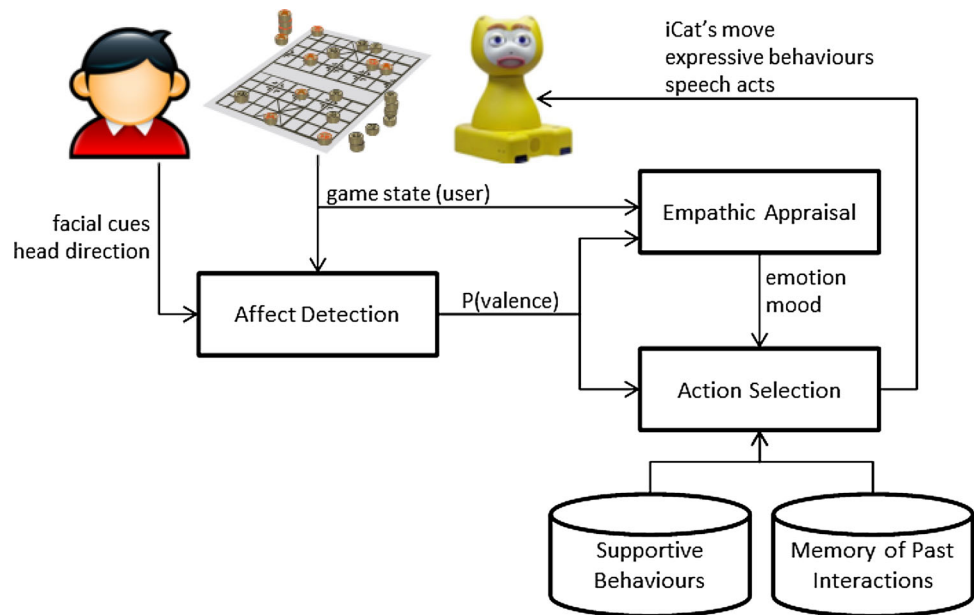
1. **Affect Detection**: a real-time prediction of the current affective state of the user who is interacting with the robot. This component returns the probability of the user's positive and negative valence of feeling taking into account both visual cues from the user and information about the state of the game. More details on the multimodal affect detection system can be found in [6].

2. **Empathic Appraisal**: based on the current affective state of the user, the robot appraises the situation and generates an empathic response (e.g., a facial expression in tune with the user's affective state) using "perspective-taking", that is, appraising the situation that the user is experiencing from his/her own point of view. Taking inspiration from Scherer's work [35], which divides the affective states in five different categories (emotion, mood, interpersonal stances, attitudes, and personality traits), the empathic appraisal of the agent incorporates the first two, emotions and mood. Emotional reactions have a short duration, but they are quite explicit. These reactions are computed based on the *emotivector* model [30], resulting in one out of nine different emotional facial expressions in the robot [29]. On the other hand, mood is a longer lasting affective state. It is less intense but remains in the robot's facial expression for longer periods of time.

3. **Supportive Behaviours**: empathy also includes actions to reduce the other's distress. Therefore, the robot has a series of supportive behaviours that can employ when the user's affective state is negative. Considering the framework of Cutrona et al. [10], social support can be separated in four different categories: "information support" (advice or guidance), "tangible assistance" (concrete assistance, for example by providing goods or services), "esteem support" (reinforcing the other's sense of competence) and "emotional support" (expressions of caring or attachment).

4. **Memory of Past Interactions**: remembering past interactions is extremely relevant for people to build rapport with each other. As such, and since we are interested in long-term human-robot interaction, the robot remembers simple aspects of previous interactions with the user (e.g., if they played a game, it remembers who won the game), and uses such information to generate dialogue that aims to give the user the feeling of "being cared for".

5. **Action Selection**: this module selects the most appropriate actions (expressive behaviours and speech utterances) for the robot based on the modules 2, 3 and 4. The mechanism for selecting the supportive behaviours is adaptive and takes into account previous interactions with that same user (for more details, please see [27]).

3.2 Application Scenario

This model was implemented in a social robot (iCat from Philips) that plays chess with children using an electronic chessboard. Each interaction starts with the iCat waking up. If the user is interacting with the robot for the first time, the iCat simply invites the user to play (e.g., by saying "Let's play chess!"), otherwise it greets the user by his/her name (e.g., "Hello Maria, nice to see you again!") and makes a comment about their previous game (e.g., "It's been 6 days since we played together. I won our last game, have you been practising?"). After every child's move, the robot provides feedback on the moves that children play by conveying empathic facial expressions determined by the Empathic Appraisal component. If the child's affective state is negative, one of the supportive behaviours described in Table 1

Fig. 2 Child playing with the iCat

is displayed. Since the iCat cannot move its pieces, it asks the user to play its move by saying the move in chess coordinates. After that, the robot waits for the children's next move, until one of the players checkmates the other. At the end of the interaction, the iCat comments the game result and/or the child's progress (e.g., "It was a good game! You are doing very well: in the four times we played together, you could beat me three times!"). The robot's behaviour is fully autonomous, except for an initial parametrisation where the name of the child needs to be typed in.

Initial experiments with children showed that they perceived the robot as more engaging and helpful when it reacted to their emotions by displaying empathic behaviour [27]. Moreover, the robot's empathic behaviour affected positively how children perceived the robot, when compared with the same version of the robot without empathic capabilities [23]. However, these results were obtained with a single interaction with the robot. In the following section, we present the results of a study in which the same group of users played with the robot over repeated interactions.

4 Method

To evaluate the impact of the proposed model in long-term interaction between children and social robots, we conducted a long-term study in a Portuguese school (Fig. 2).

Our main hypothesis for the study was the following: *children's perception of social presence, engagement and support towards the robot will remain constant from the first to the last interaction session.*

4.1 Participants

The participants of this study belonged to a Portuguese elementary school where children have chess lessons as part of their extra-curricular activities. A total of 16 participants from the 3rd grade were selected: nine girls and seven boys. Their ages varied between 8 and 9 years old ($M = 8.5$) and

their chess level was similar, as all of them had chess lessons at least since the 1st grade.

The study took place in the school after the official school hours (from 4 p.m. to 6 p.m.), with the 3rd grade children who stayed in the school during that period doing their homework and other activities supervised by a teacher. None of the children had interacted with the iCat before.

4.2 Procedure

Each child played a total of five chess exercises with the iCat—one exercise per week over 5 consecutive weeks. The exercises consisted in playing from a predefined chess position until the end of the game (i.e., either the child or the iCat checkmates the other) and were suggested by the school's chess instructor so that the difficulty was appropriate to the chess level of the children. After approximately 20 min, if none of the players had checkmated the other, the iCat either proposed a draw to the user (if it was in advantage) or gave up (if it was in disadvantage). The difficulty of the exercises varied over the sessions: in the first, third and fifth weeks the exercises were easier (i.e., the child started with advantage), whereas in the second and fourth weeks the exercises were more challenging to the child, since the iCat started with some advantage.

The procedure was the same every week: at the scheduled time, the child was guided to a room where she was alone with two experimenters and was asked to play a chess exercise with the iCat. Each game lasted, on average, 20 min, ranging between approximately 10–25 min. After playing with the robot, in the first and last weeks of interaction children filled in a questionnaire and were interviewed in a different room by another experimenter. All the interaction sessions were video recorded. Additionally, for each child, interaction logs were automatically saved in every interaction. The logs contain not only information about the game (e.g. all the moves played by both the child and the iCat, captured pieces, the game results, etc.), but also information related to the affective states of the children and the empathic behaviours employed by the robot in the different moments of the game.

4.3 Measures

This section presents the main measures of this study, obtained by combining two different data collection methods: questionnaires and open-ended interviews. The questionnaire contained several assertions to specifically evaluate the measures presented next. The open-ended interviews were used mainly to complement the questionnaire analysis, understand whether the behaviours implemented in the iCat were well understood by the children, and understand children's general motivations, expectations and suggestions to improve the iCat.

Table 1 Examples of supportive behaviours implemented in the iCat based on the theory of Cutrona et al. [10]

Social support category	Supportive behaviours	Examples of implementation in the iCat
Information support	Suggestion/advice	"Need help? Touch my paw so I can suggest you a move."
	Teaching	"That wasn't your best move, because now I can capture your Queen."
Tangible assistance	Direct task	(Play a bad move)
	Tension reduction	"Shall we start this exercise all over again?"
		"I always say, lucky in love, unlucky in chess."
Esteem support	Compliment	"That was professionally done!"
	Validation	"Well done, you played what I have played!"
	Relief of blame	"Don't worry, you didn't have better options."
	Reassurance	"Something's not quite right, but it will get better for sure."
Emotional support	Relationship	"I really enjoy playing with you!"
	Undestanding	"I understand how you're feeling, I've been through similar situations."
	Encouragement	"Come on, I still believe in you!"

The robot spoke to children in Portuguese; these examples are translations of the Portuguese utterances

4.3.1 Social Presence

Social presence measures "the degree to which a user feels access to the intelligence, intentions, and sensory impressions of another" [4], and it has been widely used to measure people's responses towards different technological artefacts, such as text-to-speech voices [22], virtual reality environments [16] and social robots [15,36]. This measure was also used in a preliminary long-term study conducted with an initial version of this scenario [28], where we employed the same questionnaire items. In this questionnaire, social presence is measured through six different dimensions:

– Co-presence: the degree to which the observer believes she/he is not alone;
– Attentional allocation: the amount of attention the user allocates to and receives from an interactant (in this case the iCat);
– Perceived message understanding: the ability of the user to understand the message from the interactant;
– Perceived affective understanding: the user's ability to understand the interactant's emotional and attitudinal states;
– Perceived affective interdependence: the extent to which the user's emotional and attitudinal state affects and is affected by the interactant's emotional and attitudinal states; and
– Perceived behavioural interdependence: the extent to which the user's behaviour affects and is affected by the interactant's behaviour.

In our initial long-term study, the empathic model was not implemented in the iCat and the robot simply displayed expressive emotions according to its own performance in the game. We found that some of the social presence dimensions decreased over time, namely perceived affective and behavioural interdependence and attentional allocation.

4.3.2 Engagement

Engagement can be defined as "the process by which two (or more) participants establish, maintain and end their perceived connection" [38], and it is a very important metric for long-term interaction [3]. If users are engaged, it is more likely that they will keep interacting with the robot for longer periods of time. The questionnaire items we used for Engagement are based on the questions developed by Sidner et al. [38] to evaluate users' responses towards a robot capable of using several social capabilities to attract the attention of users.

4.3.3 Help and Self-Validation

Help and Self-Validation are dimensions from a Friendship Questionnaire [31] employed in a study where the iCat observes and comments the chess match between two human players [29]. With these two metrics, we intend to evaluate how helpful children perceived the robot (help), and to what extend they consider the iCat as encouraging and able to help children to maintain a positive image of themselves (self-validation). These measures are related to the perception of social support that will be described next.

4.3.4 Social Support

Perceived social support can be seen as "the belief that, if the need arose, at least one person in the individual's circle would be available to serve one or more specific functions" [10]. Cutrona and colleagues argue that "it is neces-

Fig. 3 Cards containing different supportive behaviours displayed by the iCat

sary for the person to have experienced a number of interactions with the individual that communicates support", and therefore perceived social support can only be measured after repeated interactions. As such, we measured perceived support in the final questionnaire (in the 5th interaction session). Children were presented with a series of assertions adapted from the Social Support Questionnaire for Children (SSQC) [14], a self-report measure designed to evaluate children's social support via five different scales: parents, relatives, non-relative adults, siblings, and peers. In this case, we adapted the Peer scale by translating the items from English to Portuguese and changing "a peer" to "iCat".

4.3.5 Preferences on Support Behaviour Types

In addition to perceived support in general, we were also interested in understanding the impact that the different support behaviour categories—information support, tangible assistance, esteem and emotional support—had on children. To do so, in the final interview we gave participants four different cards containing a picture of the iCat and some speech bubbles (see Fig. 3). Each card contains speech bubbles with sentences that the iCat says when it is employing behaviours from one of the categories. Participants were asked to order the four cards from the one they liked the most to the one they preferred less.

5 Results and Discussion

In this section, we present and discuss the results of this study. The quantitative results obtained through the questionnaire data are presented first. These results are then complemented with the analysis of the open-ended interview questions.

5.1 Quantitative Results

To measure social presence, engagement, help and self-validation, the questionnaire contained a set of assertions that

children had to rate using a 5-point Likert scale, whereas for measuring social support, a 4-point Likert scale was used to ensure comparability with results of the original scale. The Likert scales were anchored by a *smileyometer* [32] to facilitate the interpretation of the scale. The English translation of the final questionnaire items used in this study can be found in Appendix.

The questionnaire data was analysed by comparing the results from the 1st and 5th weeks of interaction. The only exception was in the social support measure, because these items were present only in the questionnaire of the 5th interaction session. Considering our hypotheses, our goal is to show that the questionnaire results of the first session are indistinguishable from the results of the last session. In this case, traditional statistical tests of significance are not appropriate [42]. As such, we performed equivalence tests by comparing the means and confidence intervals between the two groups (1st and 5th weeks).
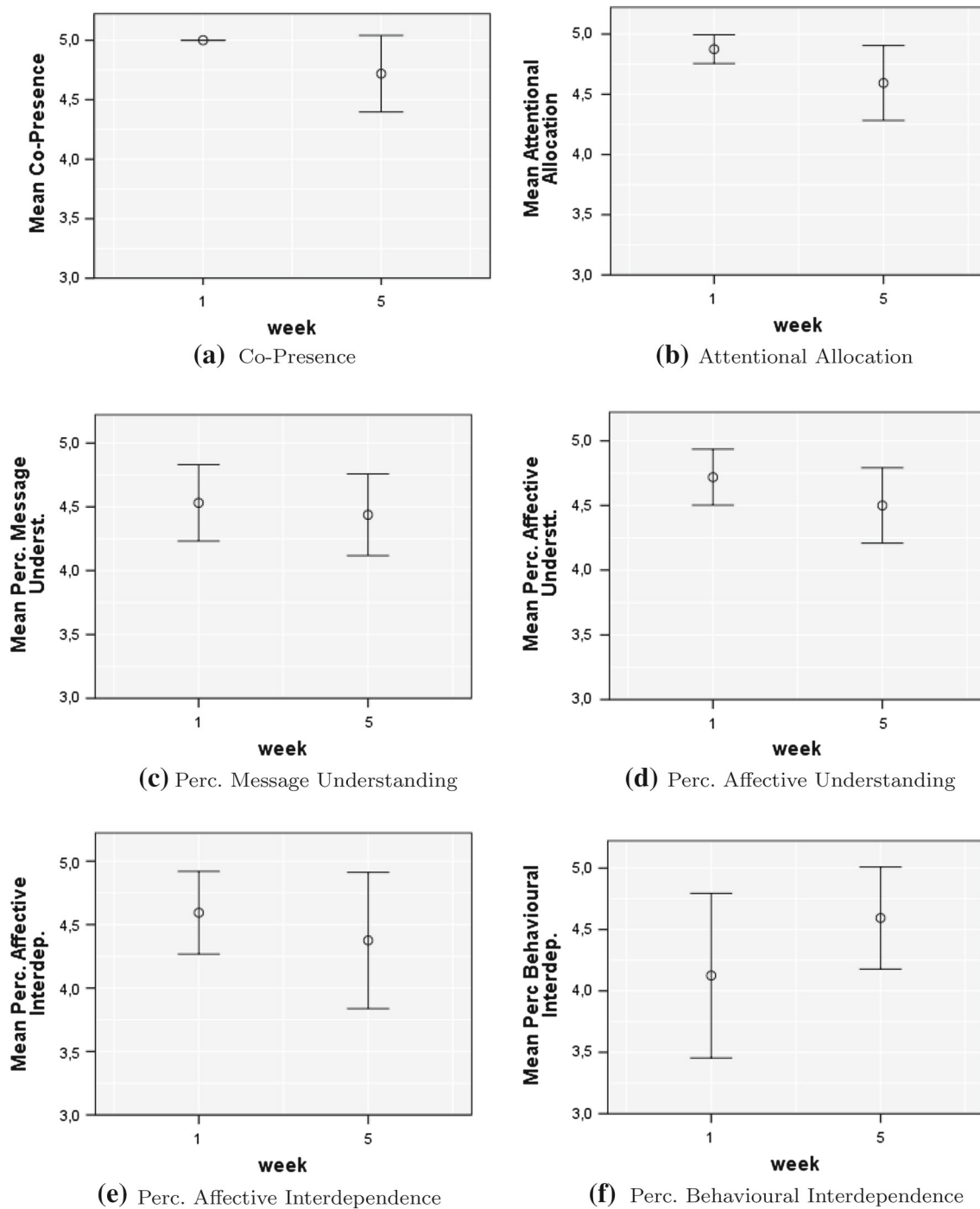
5.1.1 Social Presence

The means and confidence intervals of the Social Presence dimensions between the 1st and 5th week of interaction are displayed in Fig. 4. In contrast with the results obtained in the first long-term study [26], the ratings remained roughly the same between the 1st and 5th interaction sessions, even in the dimensions that decreased over time in the earlier study —attentional allocation, perceived affective interdependence and perceived behavioural interdependence. Note that, in our earlier study, participants also provided high ratings in the questionnaires of the first session, but they decreased their ratings after the five interaction sessions. In fact, in this present study, for perceived behavioural interdependence, the ratings even increased from the first to the last week (see Fig. 4f). Given that the confidence intervals overlap in most of the cases, there is strong evidence that in the two conditions (first and last week), children provided equivalent answers in terms of Social Presence.

5.1.2 Engagement

Similar results were obtained for engagement, as we can see from Fig. 5. After 5 weeks, children's ratings of engagement were very similar to the ones provided in the first week. To further analyse these results, we are planning to complement this analysis with observation and annotation of the video recordings of the sessions.

5.1.3 Help and Self-Validation

The results for help and self-validation followed the same trend as the ones for social presence and engagement, as illustrated in Fig. 6. In particular, the ratings for self-validation

Fig. 4 Means and 95 % confidence intervals for the ratings of the Social Presence dimensions in the 1st and 5th weeks of interaction

were considerably high in both sessions for all users, and the confidence intervals are very small (see Fig. 6b).

5.1.4 Perceived Social Support

We ran Cronbach alpha test to examine the internal consistency of our adapted version of the SSQC, since this was the first time that the adaptation of this scale was used. The results revealed an acceptable consistency ($\alpha = 0.52$), although the original Peer scale from the SSQC had a higher reliability ($\alpha = 0.91$).

Table 2 contains, for each questionnaire item, the Means and Standard Deviations obtained in our study and the values obtained by Gordon [14] with a sample of 416 American
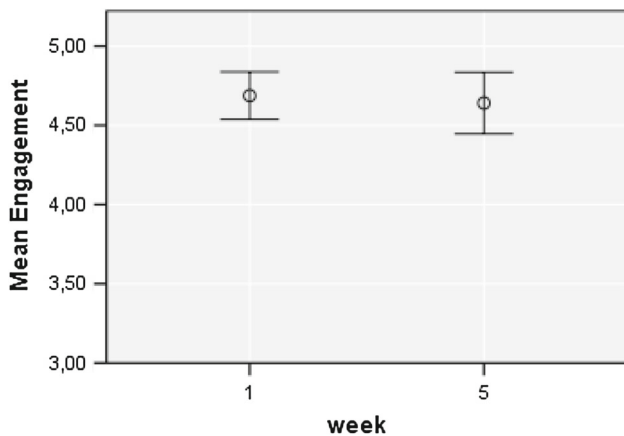
Fig. 5 Means and 95 % confidence intervals of the engagement ratings in the 1st and 5th weeks of interaction

children during the phase of selecting the final items for the questionnaire. In this latter sample, the mean age of children was 13 years old and, for this particular set of questions (Peer scale), children were asked to answer thinking on "anyone around your age who you associate with such as a friend, classmate, or teammate", whereas in our case they were asked to answer in relation to the iCat. Despite the differences in the sample, the positive results obtained in our study (all of them above the baseline mean values) indicate that, in this particular setting, the robot was perceived as supportive in a similar extent to what children in general consider being supported by their peers.

5.2 Qualitative Results

In this section, we present the most interesting findings obtained during the analysis of the open-ended interviews of the first and last weeks of interaction. Children were interviewed in a different room by another experimenter. The interview questions were separated by themes. For each theme, there was a common set of questions in both interviews and, in the final interview, additional questions were included. Table 3 contains the interview questions for the different themes.

The interview transcriptions were coded according to an iterative process [8]. An initial coding scheme was obtained while reading and highlighting the main concepts in the text. Subsequent iterations allowed us to refine the initial coding scheme. After that, related codes were grouped into themes and categories. To ensure reliability during the coding process, two complete coding iterations were performed with an interval of 1 month by the same coder. Differences between these two coding steps were addressed by asking another coder to classify the conflicting segments. In the analysis, we refer to the different participants of the study as P1, P2, ..., P16.

5.2.1 Perception of Supportive Behaviours

The questions in this theme worked as a manipulation check to understand whether children understood correctly the behaviours implemented in the robot. In general, children perceived the robot's behaviour as we intended to: most of them answered that the robot helped them when they were experiencing difficulties in the game (Q1), and that the iCat praised them when they played well (Q2). We found no substantial differences to these questions between the answers from the first and the last interviews.

Similar results were obtained in the question regarding the expressive behaviour of the iCat in the final interview (Q3). All children answered "yes" when we asked them if they understood the iCat's expressions. However, when asked to elaborate on their answer, only 14 out of the 16 children provided a valid answer. The remaining 2 children could not say why did they understand the expressions or provided an incorrect answer, for example, by saying that the robot got happy when they played bad moves.

(a) Help
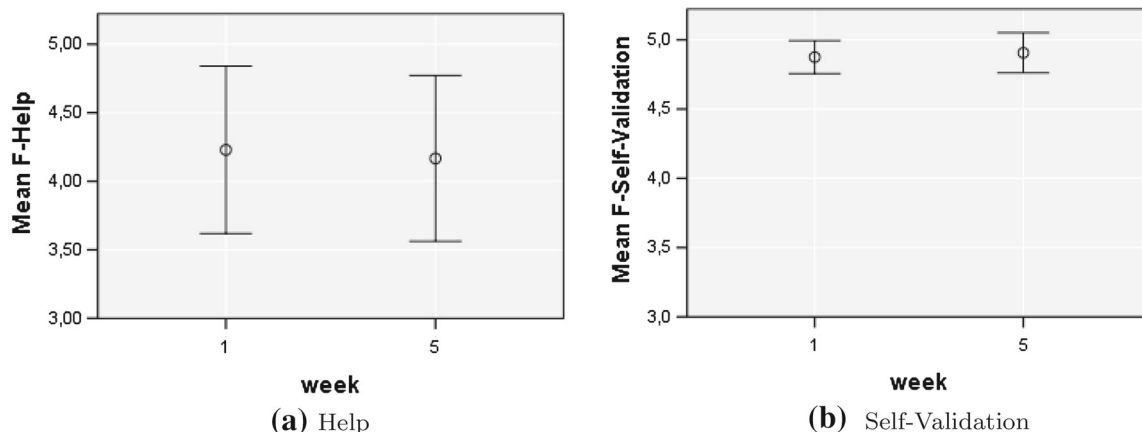
(b) Self-Validation

Fig. 6 Means and 95 % confidence intervals of the help and self-validation ratings in the 1st and 5th weeks of interaction

Table 2 Means and standard deviations of each questionnaire item obtained in our study (2nd column) and the baseline values (3rd column) obtained by Gordon [14]

The scale ranges from 0 (strongly disagree) to 3 (strongly agree)

Questionnaire items	Obtained results		Baseline	
	Means	S.D.	Means	S.D.
iCat comforts me when I am upset.	2.88	0.48	2.25	0.82
iCat cares about me.	2.75	0.75	2.27	0.85
iCat gives me good advice.	2.44	0.86	2.15	0.83
iCat accepts me for who I am.	3.0	0.0	2.47	0.73
iCat supports my decisions.	2.94	0.24	2.27	0.76
I can count on iCat.	3.0	0.0	2.51	0.69
iCat encourages me.	2.81	0.39	2.35	0.8
iCat understands me.	2.63	0.78	2.46	0.74
iCat praises me when I've done something well.	2.94	0.24	2.21	0.89

Table 3 Interview questions divided by themes

Theme	Question	Response Type
Perception of supportive behaviours	1. How did the iCat behave when you had problems during the game?	open
	2. How did the iCat behave when you played well?	open
	3*. Did you understand the iCat's facial expressions and comments? How so?	yes-no, open
	4*. What was the weirdest thing the iCat said or done?	open
Preferences on supportive behaviours	5*. These cards contain sentences that the iCat said to you while you were playing together. Please order these cards from the one you liked the most to the one you like the least.	ranking
Advantages, disadvantages, suggestions	6. Enumerate the 3 things you liked the most and the 3 things you liked the least in the iCat.	open
	7*. If iCat could help you in other tasks, what would you chose?	open
	8. Do you have any suggestions to improve the iCat?	open

The questions marked with * were only asked in the final interview

As for Q5 (*What was the weirdest thing the iCat said or done?*), 6 participants referred to concrete moments of the game. For example, P10 answered:

> He let me capture his Queen, and I didn't understand why!

These answers suggest that the supportive behaviour "Play Bad Move" was not completely understood as a deliberative action of the robot, but rather as a mistake. Additionally, 7 participants considered the tension reduction comment "Lucky in love, unlucky in chess" as the weirdest behaviour of the iCat. The remaining participants did not find anything strange in the robot's behaviour.

5.2.2 Preferences on Support Behaviour Types

As mentioned before, we asked children to order a set of four cards containing utterances illustrating the different types of supportive behaviours displayed by the iCat. To analyse children's rankings, we classified their answers according to the following procedure: we attributed 3 points to the most preferred support behaviour type, 2 points to the second most preferred, 1 point to the third and 0 points to the least preferred. We summed the points for each category for every child, ending up with the ranking displayed in Fig. 7. As we can see in Fig. 7, the most preferred supportive behaviour category was esteem support, followed by emotional support, information support, and finally tangible assistance. The esteem support category contained behaviours in which the iCat praised the user. As such, these results are in line with previous findings in HCI, in which computers capable of some forms of flatter are perceived more positively by users [12]. The low rankings of tangible assistance might have been caused by the concrete behaviours implemented in this category such as *play bad move* and *tension reduction*. Since children from this chess level often commit mistakes

Fig. 7 Rankings of the preferred support behaviour categories
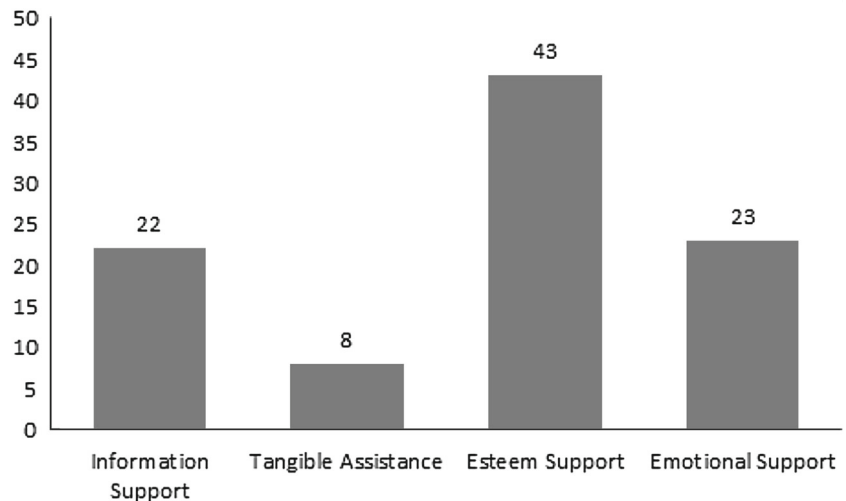
in the game (e.g., letting the other player capturing an important piece), children associated the robot's bad moves to an involuntary fault that happened not necessarily to help them, as shown in the answers from the previous theme. Regarding the use of humour (one of the possibilities for tension reduction), they may have become too repetitive over the course of the interactions because the robot only had two different jokes to say.

Overall, the emotion-oriented behaviours—esteem and emotional support—outranked the task-oriented behaviours —information support and tangible assistance. This result can be interpreted in two ways. First, it may be the case that, when playing competitive games, children prefer less tangible ways of support in contrast to being helped directly by the robot, which can reduce the merit of their victory (if they end up winning the game). The second interpretation is that the implemented task-oriented behaviours might not have been helpful enough. In the social support literature, task-oriented support often includes behaviours such as lending the other person something (e.g., money) or offering to take over of the other person's responsibilities while he/she is under stress, which are behaviours that are not applicable to this scenario (nor in most of the existing HRI scenarios).

5.2.3 Advantages, Disadvantages, Suggestions

In the answers to the questions about suggestions and enumerating positive/negative aspects of the iCat, we did not find significant differences between the answers from the first and the last interviews. In both interviews, children provided similar answers, often referring to physical capabilities of the robot (e.g., "It should have arms to move its pieces") rather than focusing on behavioural aspects.

The question *If iCat could help you in other tasks, what would you choose?* (Q8) was the one that yielded more interesting findings. 7 of the 16 children answered that the iCat

could help them with their homework, while 4 children would like to receive advice from the iCat on several matters, such as:

> He could help me in solving tough problems that I have in school.(P12)

> (...) help me when I was feeling sad. (P9)

> He could give me his opinion on different subjects (...)(P13)

One of the participants also expressed desire to do other things with the robot:

> Everything! I could play with him at school, and even invite him to sleep over sometimes. But since he is a robot, I don't know... does he sleep at all? (P2)

Additionally, two subjects wished that the iCat could play other board games with them. Tasks such as playing football, building other robots and doing housework were also mentioned.

6 Conclusion

In this paper, we investigated the issue of how to design and evaluate social robots that aim to interact with users for extended periods of time. We developed an empathic model for a social robot that plays chess with users and displays several prosocial behaviours resulting from its ability to emphasize with the user, and ran a long-term study to investigate children's perception of social presence, engagement and measures related to social support.

Overall, the obtained results are consistent with our experimental hypotheses: the developed empathic model had a positive impact in long-term interaction between children

and the robot. The ratings of social presence, engagement, help and self-validation remained similar after 5 weeks, contrasting with the results obtained in our initial exploratory long-term study where the robot was not endowed with the empathic model [26]. Moreover, the ratings of one of the social presence dimensions, perceived behavioural interdependence, even increased from the first to the final week, suggesting that over time, users were even more aware that their actions influenced the iCat's behaviour. We also found that children felt supported by the robot in a similar extent to what, in general, children feel supported by their peers. These results were complemented with the analysis of the interview data. In the interviews, we confirmed that the implemented supportive behaviours were well understood and valued by children. In general, children preferred the emotion-oriented behaviours of the iCat (esteem and emotional support) to the task-oriented behaviours (information support and tangible assistance).

6.1 Limitations

As with any user study with children, the results need to be interpreted in a cautious way. Children have an intrinsic tendency to please adults, a well studied phenomenon in the field of psychology known as *suggestibility* [7,37], which depends not only on the content and format of the questions, but can also be influenced by other factors such as the age and gender or the interviewer/experimenter [32]. To undermine suggestibility, we combined different data collection methods and analysis (questionnaires and interviews), but we are aware that there is a lot of pressure for children to behave in school contexts and this factor needs to be taken into account when interpreting the results.

Due to its long-term nature, several design decisions had to be considered because of practical limitations both in terms of time and allocated resources to this study. For example, there was only one study condition. We tried to overcome this limitation by approximating as much as possible the experimental design to the one in our first exploratory long-term study [26], to allow the comparison of some results.

6.2 Implications for Future Research

The findings reported in this paper suggest relevant implications for the design of social robots for children, particularly in applications where it is important that the robot is able to engage users over repeated interactions (e.g., education or robot-assisted therapy). If a social robot is able to display empathic and pro-social behaviours, children may see it as a peer and will eventually be more willing to continue interacting with the robot. Of course, this brings some ethical implications as well. For example, to what extent children believe that the support provided by the robot is sincere? Should they be told in advance that the robot's "concern" for them is not real? What happens if the robot suddenly breaks down or is not able to display appropriate supportive behaviours when faced with an apparent trivial situation? If robots are going to be present in our daily lives, these aspects should be carefully analysed and discussed in the near future.

Appendix: Questionnaire Items

Engagement

1. iCat made me participate more in the game.
2. It was fun playing with iCat.
3. Playing with iCat caused me real feelings and emotions.
4. I lost track of time while playing with iCat.

Social Presence

1. I noticed iCat. (Co-presence)
2. iCat noticed me. (Co-presence)
3. I remained focused on iCat. (Attentional Allocation)
4. iCat remained focused on me. (Attentional Allocation)
5. My thoughts were clear to iCat. (Perceived Message Understanding)
6. iCat's thoughts were clear to me. (Perceived Message Understanding)
7. I could tell how iCat felt. (Perceived Affective Understanding)
8. iCat could tell how I felt. (Perceived Affective Understanding)
9. iCat was influenced by my mood. (Perceived Affective Interdependence)
10. I was influenced by iCat's mood. (Perceived Affective Interdependence)
11. iCat's behaviour was tied to mine. (Perceived Behavioural Interdependence)
12. My behaviour was tied to iCat's behaviour. (Perceived Behavioural Interdependence)

Perceived Support

1. iCat helped me during the game. (Help)
2. iCat's comments were useful to me. (Help)

3. iCat's comments were helpful when I needed them. (Help)
4. I felt that I could play better in the presence of iCat. (Self-Validation)
5. iCat praised me when I played well. (Self-Validation)
6. iCat comforts me when I am upset. (Social Support)
7. iCat cares about me. (Social Support)
8. iCat gives me good advice. (Social Support)
9. iCat accepts me for who I am. (Social Support)
10. iCat supports my decisions. (Social Support)
11. I can count on iCat. (Social Support)
12. iCat encourages me. (Social Support)
13. iCat understands me. (Social Support)
14. iCat praises me when I've done something well. (Social Support)

References

1. Belpaeme T, Baxter PE, Read R, Wood R, Cuayáhuitl H, Kiefer B, Racioppa S, Kruijff-Korbayová I, Athanasopoulos G, Enescu V et al (2012) Multimodal child-robot interaction: building social bonds. J Hum–Robot Interact 1(2):33–53
2. Bickmore T, Picard R (2005) Establishing and maintaining long-term human–computer relationships. ACM Trans Comput–Hum Interact 12(2):327
3. Bickmore T, Schulman D, Yin L (2010) Maintaining engagement in long-term interventions with relational agents. Appl Artif Intell 24(6):648–666
4. Biocca F (1997) The cyborg's dilemma: Embodiment in virtual environments. In: Cognitive Technology, 1997'.Humanizing the Information Age'. Proceedings., Second International Conference on, IEEE. pp. 12–26
5. Breazeal C (2009) Role of expressive behaviour for robots that learn from people. Philos Trans R Soc B 364(1535):3527–3538
6. Castellano G, Leite I, Pereira A, Martinho C, Paiva A, McOwan PW (2013) Multimodal affect modeling and recognition for empathic robot companions. Int J Hum Robotics 10(1)
7. Ceci S, Bruck M (1993) Suggestibility of the child witness: a historical review and synthesis. Psychol bull 113(3):403
8. Chi M (1997) Quantifying qualitative analyses of verbal data: a practical guide. J Learn Sci 6(3):271–315
9. Cramer H, Goddijn J, Wielinga B, Evers V (2010) Effects of (in)accurate empathy and situational valence on attitudes towards robots. ACM/IEEE international conference on human–robot interaction, ACM, pp 141–142
10. Cutrona C, Suhr J, MacFarlane R (1990) Interpersonal transactions and the psychological sense of support. Personal relationships and social support pp. 30–45
11. Eisenberg N, Fabes RA (1990) Empathy: conceptualization, measurement, and relation to prosocial behavior. Motiv Emot 14:131–149
12. Fogg B, Nass C (1997) Silicon sycophants: the effects of computers that flatter. J Hum–Comput Stud 46(5):551–561
13. Gonsior B, Sosnowski S, Mayer C, Blume J, Radig B, Wollherr D, Kuhnlenz K (2011) Improving aspects of empathy and subjective performance for hri through mirroring facial expressions. In: RO-MAN, 2011 IEEE, pp. 350–356. doi:10.1109/ROMAN.2011.6005294

14. Gordon A (2011) Assessing social support in children: development and initial validation of the social support questionnaire for children. Ph.D. Thesis, University of Houston
15. Heerink M, Kröse BJA, Evers V, Wielinga BJ (2009) Influence of social presence on acceptance of an assistive social robot and screen agent by elderly users. Adv Robotics 23(14):1909–1923
16. Heeter C (1992) Being there: the subjective experience of presence. Presence 1(2):262–271
17. Hoffman M (2001) Empathy and moral development: Implications for caring and justice. Cambridge University Press, Cambridge
18. Kahn PH Jr, Friedman B, Perez-Granados DR, Freier NG (2006) Robotic pets in the lives of preschool children. Interact Stud 7(3):405–436
19. Kanda T, Hirano T, Eaton D, Ishiguro H (2004) Interactive robots as social partners and peer tutors for children: a field trial. Hum–Comput Interact 19(1):61–84
20. Kanda T, Sato R, Saiwaki N, Ishiguro H (2007) A 2-month field trial in an elementary school for long-term human–robot interaction. IEEE Trans Robotics 23(5):962–971
21. Kozima H, Michalowski M, Nakagawa C (2009) A playful robot for research, therapy, and entertainment. Int J Soc Robot 1:3–18
22. Lee K, Nass C (2005) Social-psychological origins of feelings of presence: creating social presence with machine-generated voices. Media Psychol 7(1):31–45
23. Leite I, Castellano G, Pereira A, Martinho C, Paiva A (2012) 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, HRI '12, pp. 367–374. ACM, New York, NY, USA. doi:10.1145/2157689.2157811. http://doi.acm.org/10.1145/2157689.2157811
24. Leite I, Martinho C, Paiva A (2013) Social robots for long-term interaction: a survey. Int J Social Robotics 5(2):291–308. doi:10.1007/s12369-013-0178-y. http://dx.doi.org/10.1007/s12369-013-0178-y
25. Leite I, Martinho C, Pereira A, Paiva A (2009) As time goes by: long-term evaluation of social presence in robotic companions. In: Robot and human interactive communication, 2009. RO-MAN 2009, IEEE. pp. 669–674
26. Leite I, Martinho C, Pereira A, Paiva A (2009) As time goes by: Long-term evaluation of social presence in robotic companions. In: Robot and human interactive communication, 2009. RO-MAN 2009. The 18th IEEE international symposium on, pp. 669–674. doi:10.1109/ROMAN.2009.5326256
27. Leite I, Pereira A, Castellano G, Mascarenhas S, Martinho C, Paiva A (2011) Modelling empathy in social robotic companions. In: L. Ardissono, T. Kuflik (eds.) UMAP Workshops, Lecture Notes in Computer Science, Springer, vol. 7138, pp. 135–147
28. Leite I, Pereira A, Martinho C, Paiva A (2008) Are emotional robots more fun to play with? Robot Hum Interact Commun 2008: 77–82
29. Leite I, Pereira A, Mascarenhas S, Martinho C, Prada R, Paiva A (2013) The influence of empathy in humanrobot relations. Int J Hum–Comput Stud 71(3):250–260. doi:10.1016/j.ijhcs.2012.09.005
30. Martinho C, Paiva A (2006) Using anticipation to create believable behaviour. In: Proceedings of the 21st national conference on Artificial intelligence - Volume 1, pp. 175–180. AAAI Press
31. Mendelson MJ, Aboud FE (1999) Measuring friendship quality in late adolescents and young adults: Mcgill friendship questionnaires. Canad J Behav Sci 31(1):130–132. doi:10.1037/h0087080
32. Read JC, MacFarlane S (2006) Using the fun toolkit and other survey methods to gather opinions in child computer interaction. In: Proceedings of the 2006 conference on Interaction design and children, IDC '06, pp. 81–88. ACM, New York, NY, USA. doi:10.1145/1139073.1139096

33. Riek LD, Paul PC, Robinson P (2010) When my robot smiles at me: enabling human–robot rapport via real-time head gesture mimicry. J Multimodal User Interfaces 3(1–2):99–108
34. Saerbeck M, Schut T, Bartneck C, Janse M (2010) Expressive robots in education: varying the degree of social supportive behavior of a robotic tutor. In: Proceedings of CHI 2010, ACM. pp. 1613–1622
35. Scherer K (2000) Psychological models of emotion. The neuropsychology of emotion
36. Schermerhorn P, Scheutz M, Crowell C (2008), Robot social presence and gender: Do females view robots differently than males? In: Proceedings of the 3rd ACM/IEEE international conference on human–robot, interaction, pp. 263–270. ACM
37. Scullin MH, Ceci SJ (2001) A suggestibility scale for children. Pers Individ Differ 30(5):843–856. doi:10.1016/S0191-8869(00)00077-5
38. Sidner C, Kidd C, Lee C, Lesh N (2004) Where to look: a study of human–robot engagement. In: Proceedings of the 9th international conference on Intelligent user interfaces, pp. 78–84. ACM
39. Tanaka F, Cicourel A, Movellan J (2007) Socialization between toddlers and robots at an early childhood education center. Proc Natl Acad Sci 104(46):17954
40. Vigil J, Geary D (2008) A preliminary investigation of family coping styles and psychological well-being among adolescent survivors of hurricane katrina. J Fam Psychol 22(1):176
41. Wada K, Shibata T, Saito T, Sakamoto K, Tanie K (2005) Psychological and social effects of one year robot assisted activity on elderly people at a health service facility for the aged. In: Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on, pp. 2785–2790. IEEE
42. Wellek S (2003) Testing statistical hypotheses of equivalence. CRC Press, Boca Raton

Iolanda Leite is a Postdoctoral Associate in the Social Robotics Lab at Yale University. She received her PhD in June 2013 from the Instituto Superior Tcnico (IST), Technical University of Lisbon, Portugal. Her research interests include human-robot interaction and affective computing, concretely in the areas of multimodal perception of social and affective behavior, user modeling, adaptive behavior generation and long-term human-robot interaction. She worked as a research assistant in EU-funded projects LIREC (Living with Robots and IntEractive Companions), MINDRACES (from Reactive to Anticipatory Cognitive Systems) and EMOTE (Embodied-perceptive Tutors for Empathy based learning). In the fall of 2012, she was a visiting researcher at Disney Research, Pittsburgh.

Ginevra Castellano is a senior researcher at the University of Birmingham, United Kingdom. Her research interests lie at the crossroads of affective behavioral computing and social robotics, and include the automatic analysis and recognition of human non-verbal behavior for adaptive human-computer and human-robot interaction. She has published more than 50 research articles on these topics. She is the coordinator of the EU FP7 EMOTE (EMbOdied-perceptive Tutors for Empathy-based learning) project (2012 2015). She is also co-investigator of the EU FP7 ILearnRW (Integrated Intelligent Learning Environment for Reading and Writing) project (2012 2015). She is a member of the management board of the HUMAINE Association; co-founder and co-chair of the AFFINE (Affective Interaction in Natural Environments) workshops; co-editor of special issues of the Journal on Multimodal User Interfaces and the ACM Transactions on Interactive Intelligent Systems; Program Committee member of several international conferences, including ACM/IEEE HRI, ACM ICMI, IEEE SocialCom, AAMAS, IUI, ACII.

André Pereira is a Ph.D. student at Instituto Superior Técnico, Technical University of Lisbon. He started his academic adventure by finishing his Master thesis at GAIPS (Intelligent Agents and Synthetic characters group). In his Master thesis he developed and studied a scenario where a social robot dotted with emotional behavior could play chess with human opponents. In his Ph.D. he is still studying how to create more believable and enjoyable board game opponents. He authored or co-authored more than 20 scientific papers, in areas such as human-robot interaction, digital games, intelligent virtual agents and multi-agent systems. Currently, he is working as a research assistant at EU-funded project EMOTE (EMbOdied-perceptive Tutors for Empathy-based learning).

Carlos Martinho is Assistant Professor at Instituto Superior Técnico, Technical University of Lisbon, and Senior Researcher in the Intelligent Agents and Synthetic Character Group at INESC-ID. His research interests include autonomous agents, synthetic characters, social robotics and user adaptation, with a focus on the creation of believable behaviour designed for interaction experiences. He has co-authored over 80 papers on these subjects and worked in more than 10 EU-funded and international projects related to his field of research.

Ana Paiva is an Associate Professor in the Department of Computer Science and Engineering of Instituto Superior Técnico, Technical University of Lisbon. She is also the group leader of GAIPS (Intelligent Agents and Synthetic Character Group) at INESC-ID. Her main scientific interests lay in the area of Autonomous Agents, Embodied Conversational Agents and Robots and Multiagent Simulation Systems. Prof. Ana Paiva has been researching in the area of artificial intelligence for the past twenty years, having also taught at IST during that period.