



**University of  
Sunderland**

Obaid, Mohammad, Aylett, Ruth, Barendregt, Wolmet, Basedow, Christina, Corrigan, Lee J., Hall, Lynne, Jones, Aidan, Kappas, Arvid, Kuster, Dennis, Paiva, Ana, Papadopoulos, Fotis, Serholt, Sofia and Castellano, Ginevra (2018) Endowing a Robotic Tutor with Empathic Qualities: Design and Pilot Evaluation. *International Journal of Humanoid Robotics*, 15 (6). ISSN 0219-8436

Downloaded from: <http://sure.sunderland.ac.uk/id/eprint/10074/>

#### **Usage guidelines**

Please refer to the usage guidelines at <http://sure.sunderland.ac.uk/policies.html> or alternatively

contact [sure@sunderland.ac.uk](mailto:sure@sunderland.ac.uk).

*Preprint of an article submitted for consideration in International Journal of Humanoid Robotics ©  
2018 [copyright World Scientific Publishing Company]  
<https://www.worldscientific.com/worldscinet/ijhr>*

## Endowing a Robotic Tutor with Empathic Qualities: Design and Pilot Evaluation

Mohammad Obaid<sup>1</sup>, Ruth Aylett<sup>2</sup>, Wolmet Barendregt<sup>3</sup>, Christina Basedow<sup>4</sup>, Lee J. Corrigan<sup>5</sup>,  
Lynne Hall<sup>6</sup>, Aidan Jones<sup>5</sup>, Arvid Kappas<sup>7</sup>, Dennis Küster<sup>8</sup>, Ana Paiva<sup>9</sup>,  
Fotis Papadopoulos<sup>10</sup>, Sofia Serholt<sup>11</sup>, Ginevra Castellano<sup>1</sup>

<sup>1</sup>*Department of Information Technology, Uppsala University, Sweden*

<sup>2</sup>*Mathematics and Computer Science, Heriot-Watt University, UK*

<sup>3</sup>*Department of Applied IT, University of Gothenburg, Sweden*

<sup>4</sup>*Department of Psychology, Vancouver Island University, Canada*

<sup>5</sup>*Department of Electronic, Electrical and Systems Engineering, University of Birmingham, UK*

<sup>6</sup>*Faculty of Computer Science, University of Sunderland, UK*

<sup>7</sup>*Department of Psychology and Methods, Jacobs University Bremen, Germany*

<sup>8</sup>*Department of Computer Science, University of Bremen, Germany*

<sup>9</sup>*Instituto Superior Tecnico Universidade de Lisboa, Portugal*

<sup>10</sup>*School of Computing Electronics and Mathematics, Plymouth University, UK*

<sup>11</sup>*Dept. of Computer Science and Engineering, Chalmers University of Technology, Sweden*  
*mohammad.obaid@it.uu.se*

Received Day Month Year

Revised Day Month Year

Accepted Day Month Year

As increasingly more research efforts are geared towards creating robots that can teach and interact with children in educational contexts, it has been speculated that endowing robots with artificial empathy may facilitate learning. In this paper, we provide a background to the concept of empathy, and how it factors into learning. We then present our approach to equipping a robotic tutor with several empathic qualities, describing the technical architecture and its components, a map-reading learning scenario developed for an interactive multitouch table, as well as the pedagogical and empathic strategies devised for the robot. We also describe the results of a pilot study comparing the robotic tutor with these empathic qualities against a version of the tutor without them. The pilot study was performed with 26 school children aged 10-11 at their school. Results revealed that children in the test condition indeed rated the robot as more empathic than children in the control condition. However, as we explored several related measures, such as relational status and learning effect, no other significant differences were found. We further discuss these results and provide insights into future directions.

*Keywords:* robot; tutor; empathy; education; children.

### 1. Introduction

In recent years, Human-Robot Interaction (HRI) has made advances in the design of robots that can take on a broad range of social roles in different domains, including for education, as socio-emotional support, and in therapy<sup>27,19,50,80,48</sup>. One of these

efforts focuses on the design of educational robotic assistants. The aim is to provide such robots with similar perceptive, expressive, and educational capabilities to those a tutor requires to effectively help learners. One aspect of this aim is to equip robots with the ability to establish affective loops with children<sup>14</sup>, so that they are able to generate the socio-emotional response behavior required to be perceived as empathic and helpful<sup>11</sup>.

The emergence of robots as educational assistants is grounded in a substantial amount of previous work on virtual agents in education<sup>41,60</sup>. It is argued that virtual agents can offer greater motivation to students for a given task in a technologically-enhanced environment compared to the same environment without such an agent<sup>41</sup>. For example, Yilmaz et al.<sup>88</sup> have demonstrated that animated agents can contribute to the learning experience of students and positively impact their grades, attitudes and retention of learning. Research has also shown that introducing social behaviors to such virtual tutors may improve their effectiveness<sup>42,69</sup>. The embodiment and physical presence of artificial entities play an important role in how people perceive them<sup>57</sup>, and a large body of research has suggested that a real-world physical embodiment has advantages compared to virtual agents<sup>64,3,47,55</sup>. This has increased interest in using social robots as tutors, instead of virtual agents or other types of educational software<sup>65</sup>. Mounting evidence on the supportive effects of robots on students' learning has led to more systematic testing of such systems in schools<sup>3,72</sup>.

Children have a great capacity to engage with robots, and to anthropomorphize them<sup>6</sup>. This may render the task of endowing robots with empathic capabilities for enhanced learning less daunting. If empathy is understood as an interactive process between two agents, rather than as a capacity residing solely in the robot<sup>6</sup>, then an *as-if* level of empathic competence might indeed be sufficient to improve learning with social robotic tutors. While the child should perceive the robot as empathic, the robot need not be able to truly share the child's emotional experience. This approach should, of course, be discussed in relation to the ethical implications it holds. For example, empathic relations with robots could potentially be understood as deceptive<sup>17</sup>, and present a number of implications, which in the longer term may be deemed undesirable by, e.g., teachers<sup>73</sup>. Notwithstanding, it has also been questioned whether robots could even reach a stage where they can be considered empathic<sup>76,75</sup>.

In this research, which is part of the EU funded project EMOTE<sup>a</sup>, the focus is to explore the possibility and effectiveness of equipping robotic tutors with empathy. Thus, this paper contributes to the HRI field by providing opportunities for design of artificial empathy for robotic tutors through a detailed description of technical architecture, as well as lessons learned from a pilot study with children to explore its impact on perceptions and short-term learning effects.

<sup>a</sup><http://www.emote-project.eu/>

## 2. Related Work

### 2.1. Educational Robots and Social Behaviours

Research on how robots can function as tools in education has intensified in recent years<sup>7</sup>. Robots are now used in personalized socially-assistive scenarios: as therapeutic tools for children with autism, as educational facilitators and companions<sup>28,59</sup>, as teachable agents<sup>79</sup>, and as tutors in the classroom<sup>48</sup>. In an overview of research on educational robots, Mubin et al.<sup>61</sup> highlighted the tutoring role as one of the main expectations for an educational robot.

Several researchers have worked towards understanding the exact qualities needed in such a robotic tutor. Saerbeck et al.<sup>72</sup> presented a study on the influence of supportive behaviors in a robotic tutor on learning efficiency. They implemented supportive behaviors in the iCat robot to help students in language learning, and compared it with a version of iCat that did not have any supportive behaviors. They concluded that the introduction of social supportive behaviors increased students' learning efficiency. Kennedy et al.<sup>47</sup> likewise investigated the effects of adopting social behaviors. Their results suggested that the presence of a robot capable of tutoring strategies may lead to better learning. However, they also cautioned that social behaviours in robotic tutors can potentially distract children from the task at hand.

More recently, researchers started investigating how robots can be used to support personalized learning<sup>16</sup>. Examples include studies exploring the effect of personalized teaching and timing strategies delivered by social robots on learning gains<sup>54,66</sup>, and affective personalization of social robotic tutors to facilitate student learning<sup>30</sup>. Some work proposed to use personalized robotic tutors to promote the development of students' meta-cognitive skills and self-regulated learning<sup>43</sup>. Previous work by some of the authors investigated the effect of empathic and supportive strategies by a robot acting as a game companion in an educational scenario on children's perceived quality of the interaction<sup>12</sup>. However, to our knowledge, no previous research has explored effects of personalization via empathy in a robotic tutor on learning gains and students' perception of the robotic tutor's empathic skills.

### 2.2. Empathy

There is no consensus on the term empathy in the literature<sup>4</sup>. It is often understood as an inter-subjective process that involves the capacity to share someone's affective experiences while remaining aware of whose feelings belong to whom<sup>23,77</sup>. Empathy has been found to be associated with positive outcomes of the interaction, as well as a positive perception of the interaction partner, e.g., a therapist<sup>32,29</sup>. It is generally conceived of as a multidimensional construct<sup>21</sup> that is an important prerequisite of relationship formation between people<sup>9,22</sup>. Empathy can be measured based on dispositional and/or situational empathic processes<sup>78</sup>. *Dispositional empathy* is con-

sidered a character trait, whereas *situational empathy* is connected to responding with empathy in a specific situation.

When considering the relevance of empathy in education, both situational and overall dispositional empathy are likely to play a role. Teacher empathy, in the dispositional sense, has been shown to be highly relevant to educational outcomes<sup>26</sup>, including the teacher’s capacity to minimize adverse student outcomes<sup>84</sup>. However, when the aim is to program empathic behavior for a robot, situationally appropriate empathy becomes much more relevant. Situationally empathic behavior often involves a matching and expression of another’s perceived emotion, regardless if it is a positive or a negative emotion<sup>37,23</sup>. Nevertheless, an automatic synchronization of emotional expressions alone<sup>35</sup>, i.e., *motor mimicry*, may not be sufficient for empathy, as evidence suggests that mimicry may require additional contextual information<sup>37</sup>.

In our view, the potential role of empathy in HRI exceeds that of mere robot expressiveness because it highlights the importance of a match between context and expression. Sometimes, such a match may be implicitly assumed. For example, robot expressiveness has recently been shown to enhance learning and retention in narrative storytelling<sup>85</sup>. When instructing an actress to narrate a story in an expressive way<sup>85</sup>, what really happens is not an injection of random expressions but a careful synchronization of tone to context. In more interactive learning tasks, involving more degrees of freedom for the child to take part, this matching becomes more difficult. In particular, the emotional responses may no longer be as predictable, and thus need to be assessed reliably. Nevertheless, if an empathic bond between the child and a robotic tutor can be created in such a situation, it may lead to positive outcomes, acceptance, and better perceived learning<sup>44</sup>. For a robotic tutor to display such convincing empathic behavior, it must close the affective loop<sup>11,15</sup>.

Automatic affect sensing may be based on a variety of affective cues<sup>89</sup>. Nevertheless, the ability to automatically recognize affect in HRI frameworks is still limited. Exceptions include work by Liu et al.<sup>56</sup>, who developed an affect inference mechanism based on physiological data for real-time detection of affective states in children, and work by Rich et al.<sup>68</sup>, involving automatic recognition of engagement in HRI based on a set of “connection events” such as directed gaze, mutual facial gaze, conversational adjacency pairs, and backchannels. In educational scenarios, Castellano et al.<sup>13</sup> developed a computational framework for the real-time recognition of affective states experienced by children playing chess with an iCat robot. In their work, the robot autonomously sensed affective states related both to the game and to the social interaction with the robot, such as feelings experienced during the game, level of interest, and engagement with the robot, using different combinations of behavioral (e.g., eye gaze, facial expressions, expressive postural features) and contextual (e.g., task- and robot-related) features<sup>13,10</sup>.

Castellano et al.<sup>12</sup> showed that integrating empathic interventions with the automatic detection of children’s affective states in real-time led participants to perceive the robot as more helpful, more engaging, and more friendly. When children expe-

rienced negative feelings throughout the game, the robot adapted to the situation by employing empathic strategies such as encouraging comments, scaffolding (e.g., providing feedback on the child's last move, letting the child play again), suggesting a good move for the child to play in his or her next turn, or intentionally playing a bad move. The evaluation of this system highlighted that affect sensing and empathic abilities are necessary for the design of robots that are perceived as being as supportive to children as their human peers<sup>51</sup>.

There is only a limited body of previous work, however, that has explored the effects of empathy on learning performance in educational scenarios with a robot acting as an educational agent. In this paper, we mainly focus on whether situational empathic qualities exhibited by a robotic tutor have an impact on students' perceptions of the robot's empathy. In line with the work by Kennedy et al.<sup>47</sup> we also explore whether these empathic qualities influence their perceived relationship with the robot, as well as students' (perceived) learning with the robotic tutor.

### 3. Design of the Empathic Robotic Tutor and a Learning Scenario

In this Section the process to design the empathic robotic tutor is presented. This includes a description of the learning scenario and the tutor's pedagogical strategy, as detailed in the sections below.

#### 3.1. Design Process

In order to develop an empathic robotic tutor that drew on principles from educational science, while demonstrating effective HRI, our design approach was based on the following maxims, also described in<sup>44</sup>: (1) Involve both teachers and students in the design of the robot to understand the social and contextual structures inherent in the environment. (2) Identify core empathic and personalized pedagogical strategies from human interactions. Successful personalized tutoring has to identify those empathic and pedagogical components and strategies that are most effective in establishing, strengthening, and sustaining social bonds. HRI studies that are based on human interactions can be quite successful, e.g., by adopting human gaze behaviours to increase engagement with the robot. (3) Supplement Human-Human Interaction (HHI) based behaviours with new capabilities available to the robot. On top of the core components identified with the help of observation and design activities, the robot can perform behaviours that teachers might not be able to produce in this form, but could tap into the same underlying mechanisms. For example, the robot could produce robot-appropriate sounds that mimic a teacher's back-channelling efforts<sup>34 46</sup>. However, HRI is not routinely based on HHI due to the differences in how humans perceive robots and other humans. For example, Serholt et al. showed that children are less likely to ask a robotic tutor than a human tutor for help, even if both tutors act in a similar way<sup>74</sup>. (4) Test interactions using techniques such as Wizard of Oz (WoZ) studies, where a robot is controlled



6 *Obaid et al.*

by a human wizard to investigate how students interact with a robot before developing final automated behaviors<sup>20</sup>. (5) Prototype and test these capabilities in the robot iteratively and in situ<sup>71</sup>. (6) Design robot capabilities on the basis of well-supported psychological and pedagogical theories, and make full use of interdisciplinary expertise<sup>1</sup>. Pedagogical psychology is a discipline with many coexisting theories, and the development of personalized learning strategies should specifically target those concepts that have been shown to be empirically well supported. (7) Enable the robot to adapt to individual differences. Towards this aim, personalized learning approaches should seek to identify cues that teachers use to adapt their teaching styles to the individual students.

### 3.2. Learning Scenario

Based on the design process described in the previous section, a learning scenario where children learn map reading skills was developed. In this scenario the NAO<sup>b</sup> robot acts as a tutor while children perform a map reading exercise on a multitouch table. The use of a multitouch table was motivated by previous research suggesting that interactive tables facilitate collaboration, equal participation, and learning<sup>38,39</sup>, while they have also been used in order to support social interaction between children and robots by providing a context in which the interaction can take place<sup>5</sup>. Further, in order to enable the development of empathic qualities, the robot's onboard sensing capabilities were augmented through the use of additional devices, namely, a web camera, a Microsoft Kinect sensor, and a Q sensor, as detailed in Section 4.

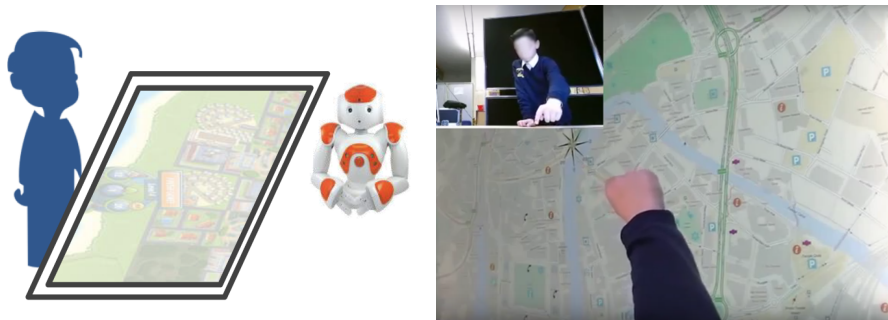


Fig. 1. Scenario (left) and map application (right)

The task is a map reading exercise and consists of following a trail on a local city map by selecting appropriate map symbols (Fig. 1). Each step of instructions

<sup>b</sup><https://www.softbankrobotics.com/emea/en/robots/nao>

in the trail was delivered verbally by the robot while also being visible on the screen until the step was completed. An instruction step always included three map competences (map symbol, cardinal direction and distance), e.g., “Go east 500 meters until you reach a bus stop”. As the task progressed, the difficulty level increased (e.g., including more complex cardinal directions and distances that needed to be transformed). Map reading tools were available within the task in the form of a compass, a map key and a measuring tool, which the robot encouraged the child to use when needed. The robotic tutor’s role was to help the child in their task while playing the scenario based on a pedagogical strategy, which is described next.

### 3.3. Pedagogical Strategy

The pedagogical strategy drew on observations of practicing teachers tutoring and scaffolding children on paper-based mock-ups of the map reading task. As noted in the literature, key functions of scaffolding include “recruitment of the child’s interest in the task, establishing and maintaining an orientation towards task-relevant goals, highlighting critical features of the task that the child might overlook, demonstrating how to achieve goals and helping to control frustration”<sup>86</sup>. Against this background, the robot was equipped with a set of tactics that aimed to facilitate progress in the task. It should be noted that we differentiate between the overall strategy and the tutoring tactics, where the tactics were the constitutive elements within the overall strategy. For example, if the strategy was concerned with encouraging children to utilize the available tools in the task, particular tactics pertaining to those ends were triggered. Whereas the tactics addressed *what* could be said, the strategy dictated *when* they should be said.

First, several tactics were implemented to encourage children to answer themselves, such as pumping the child for more information by asking questions, delivering hints, or providing short elaborations or longer tutorials on hard concepts<sup>31</sup>. Second, there were more assertive ways of guiding the child by focusing their attention on the task (e.g., “Try measuring this again”), or breaking down the task into smaller elements<sup>82</sup>. Third, short verbal cues, or keywords, were used consisting of just one or two important words that conveyed the critical elements of the step (e.g., “50 meters”)<sup>2,63</sup>. The robot could re-question the child by repeating an instruction in a slightly different way<sup>31</sup>. Fourth, if the child had difficulties in progressing in the task, the robot could splice in the correct answer<sup>31</sup>. In addition, the robot also provided feedback to the child. This could be either positive: “Good job”; very positive: “Wow! Way to go”; neutral: “Yes, ok”; or indicative of the child’s answer being almost correct: “That was almost correct. You figured out the right distance and symbol, but you were supposed to go Northwest”<sup>36</sup>. We refrained from implementing negative utterances such as “That was incorrect” as the teachers in our mock-up studies exclusively conveyed the other forms of feedback. Research furthermore suggests that negative feedback may lower intrinsic motivation<sup>24</sup>.

In essence, the pedagogical strategy had three types of tactics: tactics to prompt

reflection/elicite information from the learner; tactics to supply content to the learner; and tactics to form a social bond with the learner. For each tactic, there were around ten different utterances that the robot could randomly choose from.

#### 4. Architecture and System Components

The architecture of the system used to realize the main goals presented was based on the framework architecture of the EMOTE project. Table 1 describes the architecture's modules of the proposed framework and Fig. 2 illustrates the relationship and flow between the different module components of the system. To this end, the overall aim of the EMOTE project and its architecture is to provide a robotic platform with empathic capabilities that supports children's learning via adaptation to their learning progress and affective states.

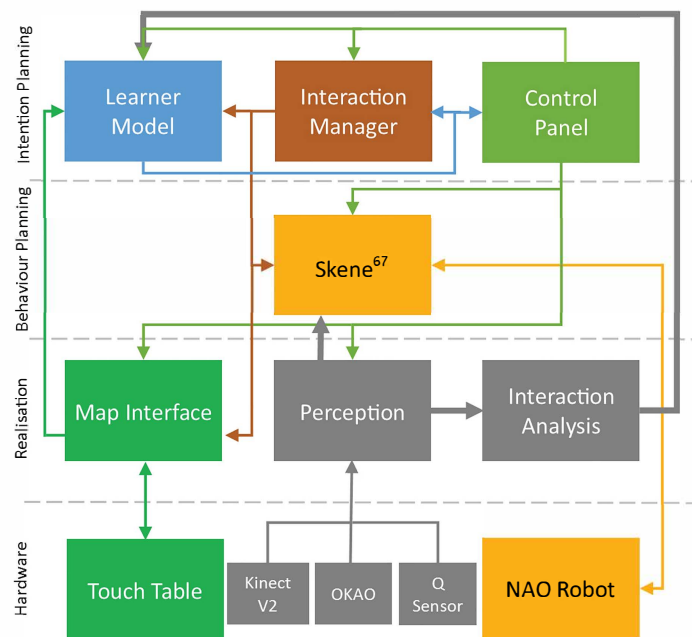


Fig. 2. EMOTE system architecture

Table 1. Architecture’s modules

	Description
Map Interface	Map Reading activity task. The scenario is explained further in Section 5.3.
Perception	Centralizes the input of Kinect, OKAO and Q Sensor and computes hand gestures, head position and gaze estimation of the child. Additionally, it saves all the available data (videos, voice, skeleton data, facial expressions etc.) in a synchronized manner for offline analysis. The perception is explained further in Section 4.2.
Interaction Analysis	Updates the learner model with an estimate of the child’s valence and arousal during the interaction with the learner. It receives regular sensor updates from the Perception Module and sends regular affective updates to the Learner Model.
Learner Model	Creates and stores a representation of the child such as affective state, learned competencies, conducted actions, and history of right or wrong answers. It provides a summary of this information to the Interaction Manager to enable the system to adapt to the learner.
Interaction Manager	This is the central decision making body of the architecture. It is responsible for updating and maintaining the context of the interaction and also for deciding how to respond to the input received. The decision making process and how the system adapts to the learner (and their affective state) is explained in Section 4.3.
Skene (Behaviour)	Skene <sup>67</sup> is a semi-autonomous behaviour planner that translates high-level intentions originated at the decision-making level into a schedule of atomic behaviour actions (e.g. speech, gazing, gesture) to be performed by the lower levels.
Control Panel	The Control Panel is the interface of the operator and allows to start or stop the interaction, input learner’s details, select task scenarios and monitor the interaction.

The system (1) infers children’s affective states in real-time via sensors embedded in the environment; (2) tracks their learning progress; and (3) adapts to the perceived state of the child by delivering appropriate pedagogical and empathic strategies to promote children’s learning and engagement with the robot.

The main system’s components are detailed in Table 1. The system captures the child’s behaviors in real-time using different sensors, including electrodermal activity using a Q-sensor<sup>c</sup>, facial expressions and eye gaze using a web camera with the OKAO SDK<sup>d</sup>, and head direction and body position using a Microsoft Kinect 2 device<sup>e</sup>. Details on the data parameters collected from the sensors are outlined in Table 2. Overall, the sensors’ input is handled by the Perception module, which is responsible for synchronization and processing, before passing it over to the Interaction Analysis module to infer information about the child’s state. The output from the Interaction Analysis module is then passed to the Learner Model module to update the estimated affective state of the child. The Interaction Manager is responsible for the robot’s decision making, and specifically for the selection of the robot’s pedagogical and empathic strategies via the Skene behaviour planner, based on information received from the Learner Model.

Sensor	Data recorded
Microsoft Kinect v.2	- Head position (x, y, z) - Head direction (x, y) - Facial Action Units
Webcam + Omron’s OKAO suite	- Face position (x, y) - Head direction angles - Eye gaze angles - Smile estimation and detection - Face expression: anger, disgust, fear, joy, sadness, surprise, neutral
Stereo Microphone	- Direction of the detected noises (left/right)
Multi-action touch screen	- Screen coordinates relative to the last touch the user did on screen
Q-sensor	- Electrodermal activity and body temperature

Table 2. Sensors’ reading parameters

<sup>c</sup><http://qsensor-support.affectiva.com><sup>d</sup><http://www.omron.com/ecb/products/mobile/><sup>e</sup><https://developer.microsoft.com/en-us/windows/kinect/>

#### 4.1. *Affect Sensing*

The Interaction Analysis module inferred children’s affect, as defined by dimensions of valence and arousal<sup>70,87,45</sup>. The implementation of the affect recognition capabilities of the Interaction Analysis module was informed by the annotation and analysis of a WoZ study as described by Corrigan et al.<sup>18</sup>.

For valence, the system used as an input the data from OKAO (via the Perception Manager). OKAO provides classifications of six discrete emotional states: happiness, surprise, fear, anger, disgust and sadness. However, our aim with regards to the interaction management and use of empathic and pedagogical strategies were more limited and targeted towards the support of learning. We thus aimed for a classification of valence into either positive, neutral or negative states. Across response systems, evidence has supported the idea that measures of emotional responses can be interpreted as dimensions rather than discrete states (e.g., happiness, fear or anger)<sup>58</sup>. Fortunately, data obtained from measures designed on the basis of a discrete view of emotions, can usually be translated into such a dimensional framework with only limited loss of information<sup>45,87,89</sup>. Given the challenges of reliable data recording in schools and the limited duration of the interaction, this choice for a dimensional affect sensing framework therefore appeared to outweigh the potential drawbacks of not having a more categorical distinction between these discrete affective states. We thus recorded OKAO data in memory, and then calculated which pattern, identified as belonging to a basic emotion, occurred most frequently over a five second period. From there we determined if the valence of the child’s emotions was either positive, neutral or negative.

The Interaction Analysis module further utilized the standardized skin conductance data from the Q-Sensor to determine whether the child’s arousal-level was increasing or decreasing over time. As with valence, we recorded the skin conductance data over a period of five seconds. Skin conductance is a widely used indicator of physiological arousal<sup>8</sup> including HCI/HRI<sup>49,52</sup>. Therefore, we classified with a threshold and computed a running average of the skin conductance, using a rule based architecture, wherein the child’s arousal was considered as either high, neutral or low. For both valence and arousal outputs, we further computed a confidence value to describe the accuracy of the module’s affect estimation. The valence confidence value was calculated by averaging the last 5 seconds of OKAO’s internal confidence value, which specified the accuracy of the recognized face extraction. The confidence value was then used by other modules to decide whether to use or reject the generated affect for the child. Because it was essential to provide the most recent data to the Learner Model as soon as it became available, the multi-threaded design of the Interaction Analysis module enabled it to respond almost immediately.

#### 4.2. *Empathic Strategies*

Depending on the robot’s perception of the child’s learning and affective state, the tutoring strategy varied. The robot’s actions were controlled by the Interaction

Manager (IM) component seen in Fig. 2. The IM consisted of a generic Engine that executed an authored script in the form of specific interaction rules. This split supports re-usability since the IM can be applied to other interactive tasks by simply changing the script.

The IM Engine implemented the Information State Update approach<sup>81</sup> through a two-step process: *update context* and *select next action*. Both were driven by rules in the script that were executed when their preconditions were satisfied. If multiple update context rules were matched, then they were all executed, but action-selection rules were addressed through one of two conflict-resolution approaches. Either the first rule that was matched could be executed - putting a premium at authoring time on rule ordering - or one rule of the matched set was picked at random and executed. The second approach could be used to vary equivalent dialogue actions across a number of different utterances. The conflict resolution strategy to be used was specified in the script file<sup>40</sup>. Note that the IM engaged in rule-chaining: When a rule was triggered, it often created the pre-conditions for other rules to fire. Thus the path from the robot's assessment of the current situation through to its chosen response usually involved the firing of many different rules. The IM included 25 pedagogical tactics (see Section 3.3), but given the richness of the context, there were 900 different scenarios in which they could be invoked, producing a very large rule set.

The essence of the empathic tutor was that the actions it took related to the affective state of the child with which it interacted. The IM script therefore contained actions whose preconditions depended on affective state. There were two sources for this information. One was of course data coming from the sensors estimating the valence and arousal combination as discussed above. The IM contained a substantial set of rules to do this. However, as with all sensor inputs, and specifically sensors trying to detect affective state, ambiguity and error were issues. For example, facial expression recognition could be impacted by the child looking down at the table, in which case the sensor would return *neutral* even if in fact the child displayed a positive or negative facial expression. For this reason, in an initial version of the system based solely on OKAO data, a confidence factor was attached to this data as it was dispatched to the IM, and initially the IM would only trigger an affective rule on it if the confidence was 75% or above.

The need for high confidence and the possibility of missing facial expressions meant that relatively few affective states were detected in a particular session. This in turn impacted evaluation of the system since it meant that empathic robot behaviour was rarely displayed.

For these reasons, we added an IM ability to infer affect from other contextual factors, notably the state of the Map Interface and the recent actions taken by the child. This interactional and learning context was already being assessed by the Learner Module (LM) for pedagogical purposes, so that the information was already available to the IM. A first implementation used very simple information

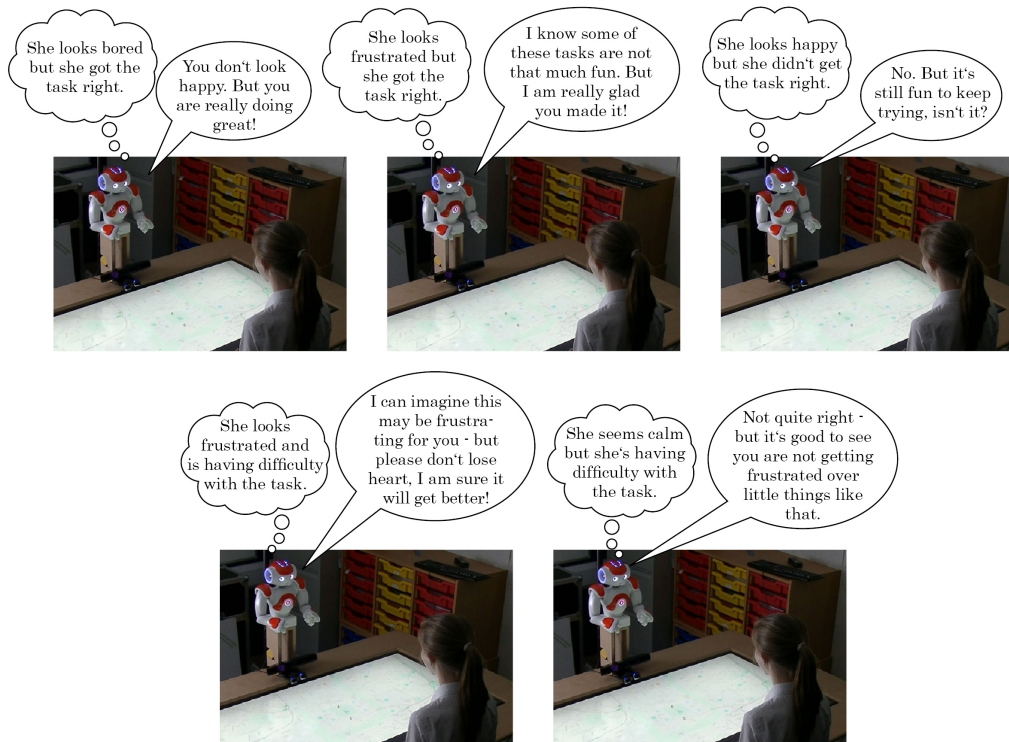


Fig. 3. Examples of possible empathic interactions

```

if overallSkillLevel == high, responseIncorrectCount >= 1, user is frustrated
if overallSkillLevel == medium, responseIncorrectCount >= 2, user is frustrated
if overallSkillLevel == low, responseIncorrectCount >= 3, user is frustrated
if timeout >= 4, user is bored
if overallSkillLevel == high and responseCorrectInARow >= 4, user is bored
Otherwise user is neutral
    
```

Fig. 4. Sample rules for inferred affect

indeed: a large number ( $>3$ ) of incorrect answers given in successive interactions was used to infer frustration, and a large number ( $>4$ ) of timeouts after a user had been asked to carry out a task was used to infer boredom.

However more sophisticated inferences were then included by adding LM estimates of the user's skill level along with the type of interactional data just mentioned. Sample rules are shown in Fig. 4. Moreover, the addition of the Q sensor for affect sensing allowed the confidence threshold for using sensed affect in the IM to be lowered to 25% .



Empathic abilities do not reside solely in the agent, but may better be regarded as arising from the interaction between the robotic tutor and the child. In consequence, we expected the child's impression of socio-emotional support and empathy coming from the robot to be influenced also by the presence of *Empathic Behaviors* - see Fig. 3 for some examples taken from a much larger set that combines affective input with pedagogical state and interaction history.

The first example represents a rule triggered by perceived (i.e., from the sensors) or inferred boredom where the task was progressing well: the Dialogue Action triggered was - `acknowledgeEmpathy: boredSuccessfulTaskCompletion`. The second example illustrates the comparable dialogue action for frustration - `acknowledgeEmpathy: frustratedSuccessfulTaskCompletion`. The third example covers the situation where the user was thought to be happy but was unsuccessful in the task - `acknowledgeEmpathy: happyUnsuccessfulAttempt`. The fourth example covers frustration and lack of success - `splice:frustrated`, and the last example the combination of a calm child and lack of success - `splice:calm`. Thus when a student was frustrated, the IM spliced in the correct answer early (i.e., presenting the right answer). However, when the student was calm or happy, it gave them more time and chances to solve the task. If the child was bored, the IM used social actions to re-engage them in the task, ranging from reassurance to the telling of a joke.

These interactions and a large number of others were generated by the relevant IM rules with affective pre-conditions, and were executed via the Skene behavior planner (see Fig. 2 and Table 1). They not only included dialogue Actions such as those shown, but also emotionally expressive sound emblems<sup>46</sup>, Nao gestures (added by Skene), as well as perceived presence and socio-emotional attentiveness expressed by following the child's movements and attention via enabled *Gaze Tracking* and *Head Tracking*.

Certain other standard HRI interactive behaviours, such as *Idle Behaviors* and personalizing *Utterances with Student Names* were enabled in both the empathic and the non-empathic tutor, as these latter behaviors could arguably be assumed to relate only to the robot's perceived intelligence rather than contributing significantly to its perceived empathy.

## 5. Pilot User Study

In order to explore the consequences of endowing our robotic tutor with empathic capabilities, we performed a pilot user study to investigate children's perceived empathy of the fully autonomous empathic robot compared to a non-empathic version of the robot. In this study, we mainly wanted to investigate whether the implemented empathic strategies indeed led to children perceiving the robot as being more empathetic. However, we also aimed to investigate some other related aspects, which will be explained in the subsequent section.

### 5.1. Hypothesis and Areas of Exploration

In line with the main aim of the project, to develop an empathic robotic tutor, we formulated our main hypothesis as follows: **(H) Perceived empathy:** Children interacting with the robotic tutor endowed with empathic qualities will rate it as more empathic than children interacting with the robotic tutor without those qualities. We furthermore explored children's perceptions of the interaction session in terms of enjoyment and the robot's helpfulness, but expected this to be high, regardless of condition because of a novelty effect. Furthermore, building on the findings of Kennedy et. al.<sup>47</sup> we also aimed to explore the perceived relational status of the robotic tutor in the empathic and non-empathic conditions, reasoning that empathic qualities of the robot could influence this status. Finally, although our study was relatively small and short-term, we wanted to explore whether there were any differences in perceived or actual learning effects.

### 5.2. Study Design

We designed a between-subjects experiment with two conditions that manipulated the behavior of a NAO robot torso as Empathic or Non-Empathic in the context of a map reading task. For the empathic condition, all system components described in Section 4 were activated. In addition to affect sensing and robot adaptation via empathic strategies, in the empathic condition the robot tracked the child's head direction and moves on the multitouch table and followed them with its head accordingly, displayed idle behaviours and utilised utterances with the name of the child. As NAO lacks the capacity for facial emotional expressions, the empathic condition further included specifically designed emotional sound emblems consisting of sequences of "bleeps and beeps". These synthetic sounds were selected from a large toolkit of validated robot sounds<sup>46</sup> that can be used for multimodal back-channelling<sup>34</sup>. In this study, we aimed to use the emotional sounds to help create a subtle sense of empathic concern expressed by the robot. For the non-empathic condition only idle behaviors and personalization of the utterances with student names remained enabled. Table 3 shows the attributes that were activated in each of the conditions. Details on affect sensing and empathic strategies used are outlined in Section 4.1 and Section 4.2.

### 5.3. Participants

We recruited 26 participants (13 girls and 13 boys) aged between 10-11 years old ( $M=10.5$ ,  $SD=0.51$ ) from a school in Birmingham, UK (Table 4 shows the details of the participants' age and gender). Informed consent was given by the parents, and the children were also asked for their assent. The study was approved by the University of Birmingham's Ethical Review Committee and followed the University's Code of Practice for Research. We randomly divided the participants into two groups for the two between-subjects conditions (empathic and non-empathic).

Table 3. Attributes of empathic &amp; non-empathic robotic tutor - more details on the properties of the system attributes are provided in Section 4

	Empathic	Non-Empathic
Inferred Affect	✓	
Empathic Strategies	✓	
Robot Sound Emblems <sup>46</sup>	✓	
Following Child's Head and Moves	✓	
Idle Behaviors	✓	✓
Utterances with Student Names	✓	✓

Table 4. Participants' age and gender details

Age	Empathic condition (female, male)	Non-empathic condition (female, male)
10	7 (2f, 5m)	6 (4f, 2m)
11	6 (4f, 2m)	7 (3f, 4m)
Total	13 (6f, 7m)	13 (7f, 6m)

#### 5.4. *Experimental setup*

The experimental setup featured a NAO T14 robot (torso version) attached to a 55" touch-sensitive interactive display from MultiTaction<sup>f</sup> placed within a custom-made aluminum table frame. A web camera, a Microsoft Kinect sensor, and a Q sensor were used to capture children's behaviours and physiological indicators, as illustrated in the example shown in Figure 3 of Section 4.2. The child sits in front of the robot, while the multitouch table is positioned between them. We showed in a previous study an overall user preference and higher engagement rates when the robot is positioned in front of the user, compared to when it sits on the side<sup>62</sup>. The web camera and Kinect were positioned in front of the child in order to fully capture the child's face from a frontal perspective.

#### 5.5. *Procedure*

The study was set up in an office room at the school as in the example illustrated in Fig. 3. The study session started by asking the participant to fill out the pre-study questionnaires. Thereafter, the administrator gave a brief description of the setup with the robot and the Map Reading application (Section 3.2). The participant was then asked to perform a map reading exercise using the multi-taction table with help from the robotic tutor. Each scenario session took 10-20 min to complete, and children were asked to fill out the post-study questionnaires afterwards.

<sup>f</sup><https://www.multitaction.com/hardware/mt-cell>

## 5.6. Measurements

*Perceived Empathy and Interaction Quality:* This questionnaire examined to what extent the children perceived the robot as empathic, and how they experienced the educational interaction with the robot. It contained seven questions on a Smiley Face Likert scale<sup>33</sup>: (E1) “I enjoyed working with Nao.”, (E2) “Nao knew when I was struggling.”, (E3) “Nao tried to imagine how I was feeling.”, (E4) “Nao tried to help me.”, (E5) “Nao was pleased when I did well.”, (E6) “Did you find the Nao robot empathic?”, and (E7) “Did Nao empathize with you?”. For the final two questions, we provided the following description of empathy: *Empathy is when you are able to understand and care about how someone else is feeling. For example, worrying for a friend who is having a bad day or understanding why the football team is happy when they win a match.*

Questions E1, E2, and E4 were intended to measure interaction quality of the educational situation with a robot, while E3, E5, E6, and E7 focused on empathy. We thus aimed to separate more general helpfulness and pleasantness of the interaction from the items targeting empathy. While overall liking and perceived helpfulness of the interaction with the robot are important indicators of possible future acceptance of robots in this role, these items did not directly concern perceiving a robot as empathic.

*Relational Status:* Kennedy et al.<sup>47</sup> asked children about the relational status of a robot they had been interacting with, using the following question: For me, I think the robot was like a: ‘brother or sister’, ‘classmate’, ‘stranger’, ‘relative (e.g. cousin or aunt)’, ‘friend’, ‘parent’, ‘teacher’, ‘neighbor’. Grouping the responses of the children into either ‘teacher’ or ‘not teacher’ they found that children did not perceive the robot to be a teacher, despite the robot being introduced to the children as such. Since it is possible that the empathic qualities of a robot influence its perceived relational status, we asked the children in our pilot study the same question after the experiment.

*Learning Effects:* In order to explore whether there were any actual learning effects, we gauged children’s geography knowledge before and after the experiment, by designing two tests that were sufficiently challenging for the target group, so that variations could be observed. To avoid children recalling answers from the pre-test at the time of the post-test, these two tests were designed differently, whereby the post-test was much more difficult to complete without mistakes.

*Perceived Learning:* In order to gauge whether the children perceived that they had learned something from the interaction, we asked them three questions about their self-assessed map-reading skills before and after the interaction with the robot. Specifically, these questions asked the children to evaluate their current skill at ‘distance measuring’, ‘compass direction’, and ‘map symbol reading’. These questions were assessed via a 5-point Smiley Face Likert scale as suggested by<sup>33</sup>.

### 5.7. Analysis and Results

*Perceived empathy (H):* Perceived empathy was assessed as a score of the 4 empathy questions: (E3) “Nao tried to imagine how I was feeling”; (E5) “Nao was pleased when I did well”; (E6) “Did you find the Nao robot empathic”; and (E7) “Did Nao empathize with you?” ( $\alpha = .63$ ). An independent samples *t*-test indicated that there was a significant difference for the score of these 4 items ( $p = .028$ ), with the empathic group ( $M = 17.69, SD = 1.70$ ) significantly higher than the non-empathic one ( $M = 15.77, SD = 2.42$ ). This supports our first hypothesis about Perceived Empathy as a function of the robot’s enabled empathic behaviors. Concerning the acceptable yet still lower than expected Cronbach’s  $\alpha$  of this scale, we conducted additional non-parametric tests at the level of the individual items. In this analysis, all 4 individual sub-items pointed in the same direction, suggesting greater perceived empathy in the empathic condition. However, individually, only the most directly phrased item (E7) showed a statistically significant higher level of perceived empathy in the empathic group ( $Mdn = 5$ ) than in the non-empathic one, ( $Mdn = 3$ ),  $U = 37.5, p = .008, r = .52$ .

*Interaction quality:* As expected, the robot was perceived very positively in both conditions, with E1: I enjoyed working with Nao ( $ME1 = 4.69$ ), E2: Nao knew when I was struggling ( $ME2 = 4.23$ ), and E4: Nao tried to help me ( $ME4 = 4.65$ ). There were no significant differences between the two conditions for these measures.

*Relational Status:* Grouping children’s answers about the relational status of the robot into ‘teacher’ or ‘not-teacher’ revealed no difference between the conditions. In both groups only 2 children reported the robot to be a teacher, while the other 11 children in both groups reported it not to be a teacher.

*Learning Effects:* Since the post-test for content knowledge was harder than the pre-test, none of the groups showed any positive learning effects. There were also no significant differences in learning effects between the two conditions.

*Self-assessed map reading skills:* Taken as a single group, children’s self-assessed knowledge increased significantly from before the interaction ( $M = 2.31, SD = 0.75$ ) to after the interaction ( $M = 2.92, SD = 1.32$ ) with the robotic tutor ( $p = 0.01$ ). However, there was no significant difference ( $p = 0.87$ ) between the two groups on the self-assessed knowledge of map reading, either on the pre- or post-test.

### 5.8. Discussion and Limitations

Our pilot study showed that children indeed perceived the empathy-enabled robotic tutor as significantly more empathic than the version without empathic capabilities. All other areas of exploration, such as interaction quality, relational status, or perceived or actual learning effects, did not indicate any significant differences. There are several reasons for this. First of all, our sample was rather small, which was due to the complexity and time required for data recording in this autonomous HRI setup. Unfortunately, while our design and data collection appeared to have sufficient statistical power to detect large effects that could be expected for the

presence of an empathy-enabled social robot<sup>g</sup>, we were unable to recruit the much more substantial sample size that would have been required to reveal more subtle effects with an acceptable level of statistical power<sup>h</sup>. We were confident to at least obtain a strong effect of the direct presence of the empathic social robot because the general effectiveness of physical presence has been well documented<sup>55</sup>, because empathy was the primary focus of our interaction design<sup>44</sup>, and due to the often surprising sensitivity of children in response to the behavior of social robots that has been observed in some studies<sup>53</sup>. Nevertheless, our study is clearly limited with regards to conclusions beyond perceived empathy and interaction quality. It appears plausible that any empathy-induced effects on learning should be weaker than the initial effects of empathy itself, as children are likely to have been sufficiently motivated and engaged in both versions of this short learning task. Based on the present empirical support for the claim that robotic tutors in child social robotics can be endowed with (perceived) empathic capabilities, future studies could examine potential empathy-induced learning gains by means of a statistically more powerful within-subjects design. For this initial study, we preferred a between-subjects design because it provides more control over alternative accounts, e.g., with respect to controlling for possible sequence effects, and the possibility of participants guessing the experimental hypotheses and thus responding in a socially desirable fashion. However, such future work could build upon, and complement, the present work.

A further limitation of the present study concerns the finding that the robot was still perceived as very helpful even without the additional empathic functions. It is possible that already the basic interaction and personalization in this condition may have exceeded the expectations of the children, resulting in a somewhat generalized liking of the robot. This finding may be unsurprising, given that children have previously been shown to respond positively to short interactions with robots<sup>53</sup>. A more long-term or repeated exposure to the empathic tutor would likely result in more pronounced gains in these related domains, as the more empathic tutor should be better equipped to support long-term motivation after initial novelty effects have worn off. This is supported by our finding that the children, overall, perceived the robot to be less empathic in the non-empathy condition - even though they appeared to be willing to be forgiving about it in this study. Finally, despite the fact that our setup had been tested during the WoZ studies and in a pre-study, both groups experienced some technical problems with the multitouch table, such as becoming unresponsive and not sending input to the robot. For example, we found that the table does not respond very smoothly to presses and is sensitive to presses with several parts of the hand at the same time. While these problems occurred with

<sup>g</sup>Post-hoc power-analyses in G-Power<sup>25</sup> (V.3.1.9.2) showed a large effect size ( $d = .92$ ) for the empathy effect reported above, reflecting an estimated power of 74%.

<sup>h</sup>According to G-Power, even a medium-sized effect ( $d = .20$ ) would already have required nearly four times the available sample size (i.e.,  $N = 102$  children) to achieve 80% power in a between-subjects  $t$ -test.

both groups, in future studies it needs to be made sure that this part is infallible, because it may inhibit the robot from responding to children's actions on the table.

## 6. Conclusions

In this paper we have described our approach to develop a robotic tutor with empathic qualities. Our study shows that by using inferred affect, Skene empathic behaviors, robot sound emblems, gaze tracking and head tracking, we have been able to indeed implement a robotic tutor that children perceive as significantly more empathic than a robotic tutor that only displays idle behaviors and addresses children by their name. However, given the limitations of our pilot study, we are unable to say that this will eventually lead to significantly better perceived or actual learning gains or to a differently perceived relationship to the robot. Future work should include longer-term within-subjects designs to achieve reasonable statistical power to detect more subtle learning effects. However, developing enough interesting, differentiated, and yet closely comparable educational material for a robotic tutor then becomes one of the challenges. At present, it further remains unclear which of the components are essential for creating this empathic behavior. Future work may focus on studying inferred affect in combination with separate components, such as empathic behaviors, or robot sound emblems, and combinations thereof. For the sound emblems, we speculate that the presence of an appropriate type of social robot may be necessary, as simply adding synthetic beeps to speech may result in reduced perceived naturalness<sup>34</sup>. More subtle expressions informed by affect sensing, such as the robot adapting its tone of voice to the emotional state of the child, could also be investigated. Finally, in some situations, online affect sensing might perhaps not be needed to create a sense of the robot having empathy. This could be the case if the child's affect in the situation is highly predictable, such as in the study by Kory Westlund et al.<sup>85</sup>, where the robot is reading a story to children. Future work in educational social robotics will thus not only need to improve affect sensing as such, but also to anticipate likely affect, as well as how to respond empathically without simply mimicking expressions.

## Acknowledgments

Thanks to the students and teachers, from Sutton Park Community Primary School, for their support and participation. Also, thanks to Linus Serholt for his valuable input. Finally, thanks to all of the EMOTE project's team. This work was supported by the European Commission (EC) and was funded by the EU FP7 ICT-317923 project EMOTE. The work was partly supported by the Swedish Research Council (grant n. 2015-04378) and the COIN project (RIT15-0133) funded by the Swedish Foundation for Strategic Research. The authors are solely responsible for the content of this publication. It does not represent the opinion of the EC, and the EC is not responsible for any use that might be made of data appearing therein.

## References

1. Patrícia Alves-Oliveira, Dennis Küster, Arvid Kappas, and Ana Paiva. Psychological science in hri: Striving for a more integrated field of research. In *2016 AAAI Fall Symposium Series*, 2016.
2. Julia Anghileri. Scaffolding practices that enhance mathematics learning. *Journal of Mathematics Teacher Education*, 9(1):33–52, 2006.
3. Christoph Bartneck. Interacting with an embodied emotional character. In *Proc. of the 2003 Int. Conference on Designing Pleasurable Products and Interfaces*, DPPI '03, pages 55–60, New York, NY, USA, 2003. ACM.
4. C Daniel Batson. These things called empathy: eight related but distinct phenomena. *MIT press*, 2009.
5. Paul Baxter, Rachel Wood, and Tony Belpaeme. A touchscreen-based 'sandtray' to facilitate, mediate and contextualise human-robot social interaction. In *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, HRI '12, pages 105–106, New York, NY, USA, 2012. ACM.
6. Tony Belpaeme, Paul Baxter, Joachim De Greeff, James Kennedy, Robin Read, Rosemarijn Looije, Mark Neerincx, Ilaria Baroni, and Mattia Coti Zelati. Child-robot interaction: Perspectives and challenges. In *The Int. Conference on Social Robotics*, pages 452–459. Springer, 2013.
7. Fabiane Barreto Vavassori Benitti. Exploring the educational potential of robotics in schools: A systematic review. *Computers Education*, 58(3):978 – 988, 2012.
8. Wolfram Boucsein. *Electrodermal activity*. Springer Science & Business Media, 2012.
9. John T Cacioppo and William Patrick. *Loneliness: Human nature and the need for social connection*. WW Norton & Company, 2008.
10. Ginevra Castellano, Iolanda Leite, and Ana Paiva. Detecting perceived quality of interaction with a robot using contextual features. *Autonomous Robots*, pages 1–17, 2016.
11. Ginevra Castellano, Iolanda Leite, André Pereira, Carlos Martinho, Ana Paiva, and Peter W. McOwan. Affect recognition for interactive companions: challenges and design in real world scenarios. *Journal on Multimodal User Interfaces*, 3(1):89–98, 2010.
12. Ginevra Castellano, Iolanda Leite, André Pereira, Carlos Martinho, Ana Paiva, and Peter W Mcowan. Multimodal affect modeling and recognition for empathic robot companions. *Int. Journal of Humanoid Robotics*, 10(01):1350010, 2013.
13. Ginevra Castellano, Iolanda Leite, André Pereira, Carlos Martinho, Ana Paiva, and Peter W. Mcowan. Context-sensitive affect recognition for a robotic game companion. *ACM Trans. Interact. Intell. Syst.*, 4(2):10:1–10:25, June 2014.
14. Ginevra Castellano, Ana Paiva, Arvid Kappas, Ruth Aylett, Helen Hastie, Wolmet Barendregt, Fernando Nabais, and Susan Bull. Towards empathic virtual and robotic tutors. In *The Int. Conference on Artificial Intelligence in Education*, pages 733–736. Springer, 2013.
15. Ginevra Castellano and Christopher Peters. Socially perceptive robots: Challenges and concerns. *Interaction Studies*, 11(2):201, 2010.
16. Caitlyn Clabaugh, Gisele Ragusa, Fei Sha, and Maja Matarić. Designing a socially assistive robot for personalized number concepts learning in preschool children. In *The 2015 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, pages 314–319. IEEE, 2015.
17. Mark Coeckelbergh. Artificial companions: Empathy and vulnerability mirroring in human-robot relations. *Studies in ethics, law, and technology*, 4(3), 2010.
18. Lee J. Corrigan, Christopher Peters, Dennis Küster, and Ginevra Castellano. *Engagement Perception and Generation for Social Robots and Virtual Agents*, pages 29–51.



22 *Obaid et al.*

Springer International Publishing, Cham, 2016.

19. Sandra Costa, Alberto Brunete, Byung-Chull Bae, and Nikolaos Mavridis. Emotional storytelling using virtual and robotic agents. *International Journal of Humanoid Robotics*, 15(03):1850006, 2018.
20. Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. Wizard of oz studies - why and how. *Knowledge-based systems*, 6(4):258–266, 1993.
21. Mark H Davis. Measuring individual differences in empathy: Evidence for a multi-dimensional approach. *Journal of personality and social psychology*, 44(1):113–126, 1983.
22. Mark H Davis. *Empathy: A social psychological approach*. Westview Press, 1994.
23. Jean Decety and Meghan Meyer. From emotion resonance to empathic understanding: A social developmental neuroscience account. *Development and Psychopathology*, 20(4):10531080, Oct 2008.
24. Edward L Deci, Robert J Vallerand, Luc G Pelletier, and Richard M Ryan. Motivation and education: The self-determination perspective. *Educational psychologist*, 26(3-4):325–346, 1991.
25. Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. G\*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. 39(2):175–191.
26. Norma Deitch Feshbach and Seymour Feshbach. Empathy and education. *The social neuroscience of empathy*, pages 85–98, 2009.
27. Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. A survey of socially interactive robots. *Robotics and autonomous systems*, 42(3):143–166, 2003.
28. Marina Fridin. Storytelling by a kindergarten social assistive robot: A tool for constructive learning in preschool education. *Computers Education*, 70:53 – 64, 2014.
29. Jairo N Fuertes, Thomas I Stracuzzi, Jennifer Bennett, Jennifer Scheinholtz, A Mislouack, Mindy Hersh, and David Cheng. Therapist multicultural competency: A study of therapy dyads. *Psychotherapy: Theory, Research, Practice, Training*, 43(4):480, 2006.
30. Goren Gordon, Samuel Spaulding, Jacqueline Kory Westlund, Jin Joo Lee, Luke Plummer, Marayna Martinez, Madhurima Das, and Cynthia Breazeal. Affective personalization of a social robot tutor for children’s second language skills. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI’16, pages 3951–3957. AAAI Press, 2016.
31. Arthur C Graesser, Katja Wiemer-Hastings, Peter Wiemer-Hastings, and Roger Kreuz. Autotutor: A simulation of a human tutor. *Cognitive Systems Research*, 1(1):35 – 51, 1999.
32. Leslie S Greenberg, Jeanne C Watson, Robert Elliot, and Arthur C Bohart. Empathy. *Psychotherapy: Theory, research, practice, training*, 38(4):380, 2001.
33. Lynne Hall, Colette Hume, and Sarah Tazzyman. Five degrees of happiness: Effective smiley face likert scales for evaluating with children. In *Proc. of the 15th International Conference on Interaction Design and Children*, pages 311–321. ACM, 2016.
34. Helen Hastie, Pasquale Dente, Dennis Küster, and Arvid Kappas. Sound emblems for affective multimodal output of a robotic tutor: a perception study. In *Proc. of the 18th ACM Int. Conference on Multimodal Interaction*, pages 256–260. ACM, 2016.
35. Elaine Hatfield, John T Cacioppo, and Richard L Rapson. Emotional contagion. *Current directions in psychological science*, 2(3):96–100, 1993.
36. John Hattie and Helen Timperley. The power of feedback. *Review of Educational Research*, 77(1):81–112, 2007.
37. Ursula Hess and Agneta Fischer. Emotional mimicry as social regulation. *Personality*

- and *Social Psychology Review*, 17(2):142–157, 2013.
38. Steve Higgins, Emma Mercier, Liz Burd, and Andrew Joyce-Gibbons. Multi-touch tables and collaborative learning. *British Journal of Educational Technology*, 43(6):1041–1054, 2012.
  39. Steven E. Higgins, Emma Mercier, Elizabeth Burd, and Andrew Hatch. Multi-touch tables and the relationship with collaborative classroom pedagogies: A synthetic review. *International Journal of Computer-Supported Collaborative Learning*, 6(4):515–538, Dec 2011.
  40. Srinivasan Janarthanam, Helen Hastie, Amol Deshmukh, Ruth Aylett, and Mary Ellen Foster. A reusable interaction management module: Use case for empathic robotic tutoring. In *Proceedings of Sem-DIAL 2015*, SEM-Dial.
  41. W. Lewis Johnson, Jeff W. Rickel, and James C. Lester. Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *Int. Journal of Artificial intelligence in education*, 11(1):47–78, 2000.
  42. W. Lewis Johnson, Paola Rizzo, Wauter Bosma, Sander Kole, Mattijs Ghijsen, and Herwin van Welbergen. *Generating Socially Appropriate Tutorial Dialog*, pages 254–264. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.
  43. Aidan Jones, Susan Bull, and Ginevra Castellano. Personalising robot tutors’ self-regulated learning scaffolding with an open learner model. In *Proceedings of WONDER (International Workshop on Educational Robots) workshop, International Conference on Social Robotics 2015 (ICSR15), Paris, France, October 2015*, 2015.
  44. Aidan Jones, Dennis Küster, Christina A. Basedow, Patricia Alves-Oliveira, Sofia Serholt, Helen Hastie, Lee J. Corrigan, Wolmet Barendregt, Arvid Kappas, Ana Paiva, and Ginevra Castellano. Empathic robotic tutors for personalized learning: a multi-disciplinary approach. 2015.
  45. Arvid Kappas, Eva Krumhuber, and Dennis Küster. Facial behavior. *Handbook of communication science: nonverbal communication*, pages 131–166, 2013.
  46. Arvid Kappas, Dennis Küster, Pasquale Dente, and Christina Basedow. Simply the best! creation and validation of the bremen emotional sounds toolkit. In *Poster presented at the 1st Int. Convention of Psychological Science, Amsterdam, the Netherlands*, 2015.
  47. James Kennedy, Paul Baxter, and Tony Belpaeme. The robot who tried too hard: Social behaviour of a robot tutor can negatively affect child learning. In *Proc. of the Tenth Annual ACM/IEEE Int. Conference on Human-Robot Interaction*, pages 67–74. ACM, 2015.
  48. James Kennedy, Paul Baxter, Emmanuel Senft, and Tony Belpaeme. Social robot tutoring for child second language learning. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction, HRI '16*, pages 231–238, Piscataway, NJ, USA, 2016. IEEE Press.
  49. Dennis Küster and Arvid Kappas. Measuring emotions online: Expression and physiology. In *Cyberemotions*, pages 71–93. Springer, 2017.
  50. Jan Kdzierski, Pawe Kaczmarek, Micha Dziergwa, and Krzysztof Tcho. Design for a robotic companion. *International Journal of Humanoid Robotics*, 12(01):1550007, 2015.
  51. Iolanda Leite, Ginevra Castellano, André Pereira, Carlos Martinho, and Ana Paiva. Long-term interactions with empathic robots: Evaluating perceived support in children. In *Proc. of the Int. Conference on Social Robotics*, pages 298–307. Springer, 2012.
  52. Iolanda Leite, Rui Henriques, Carlos Martinho, and Ana Paiva. Sensors in the wild: Exploring electrodermal activity in child-robot interaction. In *Proc. of the 8th*

24 *Obaid et al.*

- ACM/IEEE Int. Conference on Human-robot Interaction, HRI '13*, pages 41–48, Piscataway, NJ, USA, 2013. IEEE Press.
53. Iolanda Leite, Carlos Martinho, and Ana Paiva. Social robots for long-term interaction: A survey. *Int. Journal of Social Robotics*, 5(2):291–308, 2013.
  54. Daniel Leyzberg, Samuel Spaulding, and Brian Scassellati. Personalizing robot tutors to individuals' learning differences. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 423–430. ACM, 2014.
  55. Jamy Li. The benefit of being physically present: a survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *Int. Journal of Human-Computer Studies*, 77:23–37, 2015.
  56. Changchun Liu, Karla Conn, Nilanjan Sarkar, and Wendy Stone. Online affect detection and robot behavior adaptation for intervention of children with autism. *IEEE Transactions on Robotics*, 24(4):883–896, 2008.
  57. Mary Margaret Lusk and Robert K Atkinson. Animated pedagogical agents: Does their degree of embodiment impact learning from static or animated worked examples? *Applied cognitive psychology*, 21(6):747–764, 2007.
  58. Iris B Mauss and Michael D Robinson. Measures of emotion: A review. *Cognition and emotion*, 23(2):209–237, 2009.
  59. Ali Meghdari, Azadeh Shariati, Minoo Alemi, Ali Amoozandeh Nobaveh, Mobin Khamooshi, and Behrad Mozaffari. Design performance characteristics of a social robot companion arash for pediatric hospitals. *International Journal of Humanoid Robotics*, 0(0):1850019, 0.
  60. Roxana Moreno, Richard E. Mayer, Hiller A. Spires, and James C. Lester. The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction*, 19(2):177–213, 2001.
  61. Omar Mubin, Catherine J Stevens, Suleman Shahid, Abdullah Al Mahmud, and Jian-Jie Dong. A review of the applicability of robots in education. *Journal of Technology in Education and Learning*, 1:209–0015, 2013.
  62. F. Papadopoulos, D. Kster, L.J. Corrigan, A. Kappas, and G. Castellano. Do relative positions and proxemics affect the engagement in a human-robot collaborative scenario? *Interaction Studies*, 17(3).
  63. Monica L Parson. Focus student attention with verbal cues. *Strategies*, 11(3):30–33, 1998.
  64. André Pereira, Carlos Martinho, Iolanda Leite, and Ana Paiva. iCat, the chess player: The influence of embodiment in the enjoyment of a game. In *Proc. of the 7th Int. Joint Conference on Autonomous Agents and Multiagent Systems - Volume 3, AAMAS '08*, pages 1253–1256, Richland, SC, 2008.
  65. Aaron Powers, Sara Kiesler, Susan Fussell, and Cristen Torrey. Comparing a computer agent with a humanoid robot. In *Proc. of the ACM/IEEE Int. Conference on Human-robot Interaction, HRI '07*, pages 145–152, New York, NY, USA, 2007. ACM.
  66. Aditi Ramachandran, Chien-Ming Huang, and Brian Scassellati. Give me a break!: Personalized timing strategies to promote learning in robot-child tutoring. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, HRI '17*, pages 146–155, New York, NY, USA, 2017. ACM.
  67. Tiago Ribeiro, André Pereira, Eugenio Di Tullio, Patricia Alves-Oliveira, and Ana Paiva. From thalamus to skene: High-level behaviour planning and managing for mixed-reality characters. In *Proceedings of the IVA 2014 Workshop on Architectures and Standards for IVAs*, 2014.
  68. Charles Rich, Brett Ponsler, Aaron Holroyd, and Candace L Sidner. Recognizing en-

- agement in human-robot interaction. In *Proc. of the 5th ACM/IEEE Int. Conference on Human-Robot Interaction (HRI)*, pages 375–382. IEEE, 2010.
69. Jennifer L Robison, Scott W Mcquiggan, and James C Lester. Modeling task-based vs. affect-based feedback behavior in pedagogical agents: An inductive approach. In *AIED*, pages 25–32, 2009.
  70. James A Russell. Emotion, core affect, and psychological construction. *Cognition and Emotion*, 23(7):1259–1283, 2009.
  71. Selma Šabanović, S Reeder, and Bobak Kechavarzi. Designing robots in the wild: In situ prototype evaluation for a break management robot. *Journal of Human-Robot Interaction*, 3(1):70–88, 2014.
  72. Martin Saerbeck, Tom Schut, Christoph Bartneck, and Maddy Janse. Expressive robots in education - varying the degree of social supportive behavior of a robotic tutor. In *28th ACM Conference on Human Factors in Computing Systems (CHI2010)*, pages 1613–1622, Atlanta, 2010. ACM.
  73. Sofia Serholt, Wolmet Barendregt, Asimina Vasalou, Patrícia Alves-Oliveira, Aidan Jones, Sofia Petisca, and Ana Paiva. The case of classroom robots: teachers’ deliberations on the ethical tensions. *AI & SOCIETY*, Jun 2016.
  74. Sofia Serholt, Christina Basedow, Wolmet Barendregt, and Mohammad Obaid. Comparing a humanoid tutor to a human tutor delivering an instructional task to children. *2014 IEEE-RAS International Conference on Humanoid Robots*, pages 1134–1141, 2014.
  75. Amanda J. C. Sharkey. Should we welcome robot teachers? *Ethics and Information Technology*, 18(4):283–297, Dec 2016.
  76. Noel Sharkey and Amanda Sharkey. The crying shame of robot nannies: An ethical appraisal. *Interaction Studies*, 11(2):161–190, 2010.
  77. Tania Singer and Claus Lamm. The social neuroscience of empathy. *Annals of the New York Academy of Sciences*, 1156(1):81–96, 2009.
  78. Karsten Stueber. Empathy. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2017 edition, 2017.
  79. Fumihide Tanaka and Shizuko Matsuzoe. Children teach a care-receiving robot to promote their learning: Field experiments in a classroom for vocabulary learning. *Journal of Human-Robot Interaction*, 1(1), 2012.
  80. Serge Thill, Cristina A Pop, Tony Belpaeme, Tom Ziemke, and Bram Vanderborght. Robot-assisted therapy for autism spectrum disorders with (partially) autonomous control: Challenges and outlook. *Paladyn*, 3(4):209–217, 2012.
  81. David Traum and Staffan Larsson. The Information State Approach to Dialogue Management. In *Current and New Directions in Discourse and Dialogue*, pages 325–353. 2003.
  82. Irina Verenikina. Scaffolding and learning: Its role in nurturing new learners. In *Learning and the learner: Exploring Learning for New Times (Kell, P, Vialle, W, Konza, D Vogl, G , Eds)*, pages 161–180. University of Wollongong, 2008.
  83. Anna-Lisa Vollmer, Robin Read, Dries Trippas, and Tony Belpaeme. Children conform, adults resist: A robot group induced peer pressure on normative social conformity. 3(21).
  84. Chezare A Warren. *Empathic interaction: White female teachers and their Black male students*. PhD thesis, The Ohio State University, 2012.
  85. Kory Westlund, M Jacqueline, Sooyeon Jeong, Hae W Park, Samuel Ronfard, Aradhana Adhikari, Paul L Harris, David DeSteno, and Cynthia L Breazeal. Flat vs. expressive storytelling: Young childrens learning and retention of a social robots narrative.

26 *Obaid et al.*

*Frontiers in Human Neuroscience*, 11:295, 2017.

86. David Wood and Heather Wood. Vygotsky, tutoring and learning. *Oxford review of Education*, 22(1):5–16, 1996.
87. Michelle Yik, James A Russell, and James H Steiger. A 12-point circumplex structure of core affect. *Emotion*, 11(4):705–731, 2011.
88. Ramazan Yilmaz and Ebru Kılıç-Çakmak. Educational interface agents as social models to influence learner achievement, attitude and retention of learning. *Computers & Education*, 59(2):828–838, 2012.
89. Zhihong Zeng, Maja Pantic, Glenn I Roisman, and Thomas S Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE transactions on pattern analysis and machine intelligence*, 31(1):39–58, 2009.

## Appendix A

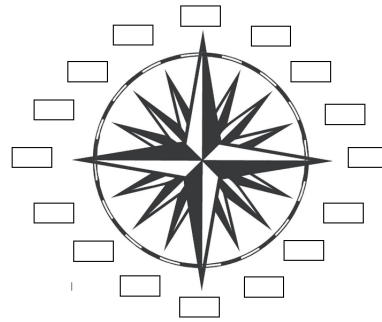


Fig. 5. Geography post-test cardinal directions 'fill in the blanks'

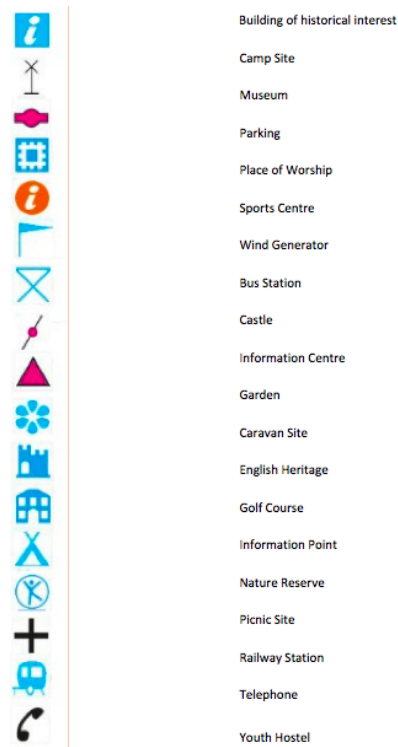
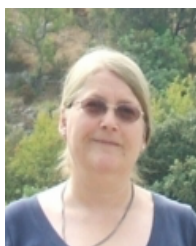


Fig. 6. Map reading post-test map symbols 'match symbol to right label'



**Mohammad Obaid** received the BSc., MSc. (First Class Honors) and Ph.D. degrees in Computer Science from the University of Canterbury, Christchurch, New Zealand, in 2004, 2007 and 2011, respectively. He worked at several international research centers including the Human Centered Multimedia Lab (HCM Lab), University of Augsburg (Germany), the Human Interface Technology Lab New Zealand (HITLab NZ), University of Canterbury (New Zealand), and the t2i Lab, Chalmers University of Technology (Sweden). In 2016, he joined the the Social Robotics Lab, Department of Information Technology, Uppsala University (Uppsala, Sweden) as a researcher. Dr. Obaid is one of the founders of the Applied Robotics Group at the Interaction Design Division at Chalmers University of Technology (established in 2015). He is the (co-)author of over 55 publications within the areas of his research interests on Human-Robot Interaction and Human-Computer Interaction. In recent years, he has served in organizing committee and program committee of main HCI/HRI related conferences such as HRI, HAI, VRST, NordiCHI, and MUM.



**Ruth Aylett** is Professor of Computer Sciences in the School of Maths and Computer Science at Heriot-Watt University. She researches Affective Systems, Social Agents in both graphical and robotic embodiments, and Human-Robot Interaction, as well as Interactive Narrative. She led three EU projects (VICTEC, eCIRCUS and eCUTE) in the period 2001-2012 applying empathic graphical characters to education against bullying (FearNot!) and in cultural sensitivity (ORIENT, Traveller, MIXER). She also worked as a PI in the projects LIREC (investigating long-lived robot companions) and EMOTE (an empathic robot tutor). She led the EPSRC-funded network of excellence in interactive narrative, RIDERS. She is currently PI of the project SoCoRo (Socially Competent Robots) which is investigating the use of a mobile robot to train high-functioning adults with an Autism Spectrum Disorder in social interaction. She has authored more than 250 referred publications in conferences, journals and book chapters, and has been an invited speaker at various events, most recently AAMAS 2016.



**Wolmet Barendregt** is an associate professor at the Department of Applied IT, Gothenburg University, Sweden. She specializes in research on educational computer games, motivation, learning, and usability. Her research interest is in how users, especially children, interact with ICT (such as robotic tutors and computer games) and learn from and with them, both inside and outside formal schooling.

She has worked in research projects, including the EU projects EMOTE and iRead. Currently she is involved in a project funded by the Swedish Research Council on Social learning for personalized human-robot interactions (2016-2020). She has (co)-authored over 60 publications in refereed journals, conferences and books.



**Christina A. Basedow** received her PhD in Psychology from Jacobs University, Bremen, Germany. She has a strong passion for working in the addictions field and teaching. Christina works as an instructor at Vancouver Island University and also is the Continuing Care Clinical Manager at Edgewood Treatment Center. She has been working with individuals, couples, and families struggling with substance misuse issues, mental health disorders, and co-dependency issues.

Christina is skilled in a variety of areas including family counselling, eating disorders, sexual compulsions, and CBT/DBT. Christina brings her personal experience, compassion, and a wealth of knowledge in psychology, research, and teaching, to her clients, and to EHN. Christina is registered with the BC Association of Registered Clinical Counsellors and currently a Certified Sex Addiction Therapist Candidate.



**Lee J. Corrigan** is a PhD candidate of the School of Electronic, Electrical and Computer Engineering at the University of Birmingham. Prior to that, he earned a BSc with First Class Honours in Games Technology from Coventry University. His academic research focuses on Human Behaviour Analysis for Social Robotics, more specifically, the phenomenon of social and task engagement found during human-robot interactions (HRI). In addition to his research interests, he also has the

knowledge, skills and experience required to design, develop and deploy practical user interfaces, autonomous systems and graphical simulations to represent scientific data and mathematical models.





**Lynne Hall** received her M.Sc. and Ph.D. degrees from the University of Manchester, in 1989 and 1993 respectively. From 1993 to 1997, she worked as an HCI Specialist at Banc Sabadell in Spain. From 1997 to 2002, she worked as a Senior Lecturer in Computing at the University of Northumbria. Since 2002, she has worked at the University of Sunderland, becoming a Professor of Computer Science in 2016. She has worked in research projects including the EU projects eCIRCUS, eCute and EMOTE;

and on outreach projects such as Digital Media Network and Sunderland Software City. She was Chair of British HCI in 2017 and is the secretary to the BCS interactions SIG. She is the author of around 140 publications. She is currently working in the AHRC Creative Fuse North East project, focusing on interdisciplinary fusion with her main research interests in evaluating interactive, creative experiences with children, adults and audiences.



**Aidan Jones** is a PhD Student at the University of Birmingham. His research interests are in the areas of social robotics, self-regulated learning, and deep-learning. He worked on the EU FP7 EMOTE (EMbodied-perceptive Tutors for Empathy-based learning) project (20122016). His research has been presented at the International Conference on Human-Robot Interaction and the International Conference on Social Robotics.



**Arvid Kappas** is professor of psychology and dean at Jacobs University Bremen. He has been conducting research on emotions for over three decades. After obtaining his Ph.D. at Dartmouth College in the USA in 1989, he has lived and worked in Switzerland, Canada, the UK, and Germany, as well as holding visiting positions in Italy and Austria. He is a specialist in the scientific study of how people interact and communicate, specifically nonverbal communication and psychophysiology.

Recently he has become interested in the interaction of humans and machines, particularly affective computing and social robotics. He was the president of the International Society for Research on Emotion (ISRE) from 2013-2018 and is a fellow and member of several scientific organizations. One of his current research activities is with the EU-funded project ANIMATAS that focuses on advancing intuitive human-machine interaction with human-like social capabilities for education in schools.



**Dennis Küster** received his diploma degree in Psychology from University of Kiel, Germany, in 2003, and his PhD from Jacobs University Bremen, Germany, in 2008, where he studied the relationship between emotional experience and social context using facial electromyography and other psychophysiological measures. Between 2007 and 2009, he was at the Universit catholique de Louvain, Belgium, as a post-doc. He is presently a senior researcher at the University of Bremen, department of Computer Science. Throughout his career, his interests have focused on emotions, their measurement, elicitation, expression, and their social functions, including interpersonal interactions mediated by computers, avatars, or robots.



**Ana Paiva** is a Full Professor in the Department of Computer Engineering, IST ("Instituto Superior Técnico") from the University of Lisbon and coordinator of GAIPS "Intelligent and Social Agents Group" at INESC-ID. She investigates the creation of AI and complex systems using an agent-based approach, with a particular focus on social agents. Her primary research interests are in the fields of Autonomous Agents and Multi-Agent Systems, Affective Computing, Virtual Agents and Human-Robot Interaction. Her most recent work focuses on the affective and social elements present in the interactions between humans and robots more specifically on how to engineer natural and positive relationships with social robots. For the past twenty years she has served in the organization of numerous international conference and workshops (including, AAMAS, HRI, ACII), and (co-)authored over 200 publications in refereed journals, conferences and books.



**Fotis Papadopoulos** received his MSc degree in Advanced Computer Science from the University of Lancaster and his Ph.D. degree in Robotics from the University of Hertfordshire. While finishing his Ph.D., he worked as Research assistant on the LIREC EU project. He worked as Research Fellow in the University of Birmingham for the EMOTE project between 2013 and 2016. Now he holds a Research Fellow position in the University of Plymouth for the EU project L2TOR. His research interests include Human-Robot Interaction, tele-operation, robot tutoring and interactive and collaborative games.



**Sofia Serholt** is a Postdoctoral Researcher in the Department of Computer Science and Engineering, Chalmers University of Technology, Sweden. She wrote her dissertation *Child-Robot Interaction in Education* and received her Ph.D. degree in Applied Information Technology with specialization in Educational Sciences in 2017 from the University of Gothenburg, Sweden. She also holds an M.S. degree in Education. Her research interests include child-robot interaction, educational robotics and ethics.

She is an active member of the research group Applied Robotics in Gothenburg, and has previously worked on the EU FP7 EMOTE (EMbOdied-perceptive Tutors for Empathy-based learning) project (2012-2016). Currently, she is a collaborator on the EU project PLACED (Place- and Activity-Centric Digital Library Services) and the Wallenberg-funded project START (Student Tutor and Robot Tutee).



**Ginevra Castellano** is a Senior Lecturer at the Department of Information Technology, Uppsala University, where she leads the Social Robotics Lab. Her research interests are in the areas of social robotics and affective computing, and include social learning, personalized adaptive robots, multimodal behaviours and uncanny valley effect in robots and virtual agents. She was the coordinator of the EU FP7 EMOTE (EMbOdied-perceptive Tutors for Empathy-based learning) project (2012-2016).

She is the recipient of a Swedish Research Council starting grant (2016-2019) and PI for Uppsala University of the EU Horizon 2020 ANIMATAS (Advancing intuitive human-machine interaction with human-like social capabilities for education in schools; 2018-2021) project and of the COIN (Co-adaptive human-robot interactive systems) project, funded by the Swedish Foundation for Strategic Research (2016-2021). She was an invited speaker at the Summer School on Social Human-Robot Interaction 2013; member of the management board of the Association for the Advancement of Affective Computing; Program Committee member of ACM/IEEE HRI, IEEE/RSJ IROS, ACM ICMI, AAMAS, ACII and many others. Castellano was a general co-chair of IVA 2017.