Emotion and Reinforcement: Affective Facial Expressions Facilitate Robot Learning

Joost Broekens

Leiden Institute of Advanced Computer Science, Leiden University
Niels Bohrwerg 1, 2333CA Leiden, The Netherlands
broekens@liacs.nl

Abstract. Computer models can be used to investigate the role of emotion in learning. Here we present *EARL*, our framework for the systematic study of the relation between *e*motion, *a*daptation and *r*einforcement *l*earning (RL). EARL enables the study of, among other things, communicated affect as reinforcement to the robot; the focus of this chapter. In humans, emotions are crucial to learning. For example, a parent—observing a child—uses emotional expression to encourage or discourage specific behaviors. Emotional expression can therefore be a reinforcement signal to a child. We hypothesize that affective facial expressions facilitate robot learning, and compare a *social* setting with a *non-social* one to test this. The non-social setting consists of a simulated robot that learns to solve a typical RL task in a continuous grid-world environment. The social setting additionally consists of a human (parent) observing the simulated robot (child). The human's emotional expressions are analyzed in real time and converted to an additional reinforcement signal used by the robot; positive expressions result in reward, negative expressions in punishment. We quantitatively show that the "social robot" indeed learns to solve its task significantly faster than its "non-social sibling". We conclude that this presents strong evidence for the potential benefit of affective communication with humans in the reinforcement learning loop.

Keywords: Reinforcement Learning, Affect, Human-in-the-Loop.

1 Introduction

In humans, emotion influences thought and behavior in many ways [16][17][19][38]. For example, emotion influences how humans process information by controlling the broadness versus the narrowness of attention. Also, emotion functions as a social signal that communicates reinforcement of behavior in, e.g., parent-child relations. Computational modeling (including robot modeling) has proven to be a viable method of investigating the relation between emotion and learning [11][24], emotion and problem solving [3][6], emotion and social robots [7] (for review see [20]), and emotion, motivation and behavior selection [2][5][15][46]. Although many approaches exist and much work has been done on computational modeling of emotional influences on thought and behavior, none explicitly targets the study of the relation between emotion and learning using a complete end-to-end framework in a

T.S. Huang et al. (Eds.): Human Computing, LNAI 4451, pp. 113–132, 2007.

reinforcement learning context[1]. By this we mean a framework that enables systematic *quantitative* study of the relation between affect and RL in a large variety of ways, including (a) affect as reinforcement to the robot (both internally generated as well as socially communicated), (b) affect as perceptual feature to the robot (again internally generated and social), (c) affect resulting from reinforced robot behavior, and (d) affect as meta-parameters for the robot's learning mechanism. In this chapter we present such a framework. We call our framework *EARL*, short for the systematic study of the relation between *e*motion, *a*daptation and *r*einforcement *l*earning.

1.1 Affect as Reinforcement

In this chapter we specifically focus on the influence of socially communicated emotion on learning in a reinforcement learning context. This work is strongly related to research into interactive robot learning based on human advice or guidance [35][43]. We briefly review this area of research in Section 2. In the experimental part of this chapter we show, using our framework *EARL*, that human emotional expressions can be effectively used as additional reinforcement signal used by a simulated robot. Our experimental setting is as follows.

The robot's task is to optimize food-finding behavior while navigating through a continuous grid world environment. The grid world is not discrete, nor is an attempt made to define discrete states based on the continuous input. The gridworld contains walls, path and food patches. The robot perceives its direct surroundings as they are, and acts by turning and driving. We have developed an action-based learning mechanism that learns to predict values of actions based on the current perception of the agent (note that in this chapter we use the terms agent and robot interchangeably). Every action has its own Multi-Layer Perceptron (MLP) network (see also [28]) that learns to predict a modified version of the Q-value for that action [41]. The simulated robot does not use a separate training phase; we adopt the so-called certainty equivalence hypothesis [27].

We have used this setup to ensure that our simulation is as close as possible to a real world setting: continuous input directly fed into MLP networks. By doing so, we hope that observed robot behavior can be extrapolated to the real world: in theory, building the actual robot with appropriate sensors and actuators would suffice to replicate the results. We explain our modeling method in more detail in Section 4-6.

As mentioned above, we study the effect of a human's emotional expression on the learning behavior of the robot. As such the simulated robot uses the recognized emotion as a motivator for action, while the human uses its expression to signal the relevance of certain events (see also, [12]). In humans, emotions are crucial to learning. For example, a parent—observing a child—uses emotional expression to encourage or discourage specific behaviors. In this case, the emotional expression is used to setup an *affective communication channel* [36] and is used to communicate a reinforcement signal to a child. In this chapter we take *affect* to mean the positiveness versus the negativeness (*valence*) of a situation, object, etc. (see [11][38] and [39] for a more detailed argumentation of this point of view, and [47] for a detailed discussion

[1] Although the work by Gandanho [24] is a partial exception as it explicitly addresses emotion in the context of RL. However, this work does not address social human input and social robot output.

on the relation between valence and reinforcement learning). In our experiments, a human observes a simulated robot while the robot learns to find food. Affect in the human's facial expression is recognized by the robot in real time. As such, a smile is interpreted as communicating positive affect and therefore converted to a small additional reward (additional to the reinforcement the robot receives from its simulated environment). The expression of fear is interpreted as communicating negative affect and therefore converted to a small additional punishment. We call this the *social* setting. We vary between three types of social settings: one in which affect is a strong reinforcement but only for several learning trials; one in which affect is a moderate reinforcement for a longer period of time; and finally one in which affect is a moderate reinforcement while (in contrast to the first two types) the robot learns a social reward function that maps its perceived state to the social reinforcement. In the latter type, the robot can use its learned social reward function when the human stops giving social reinforcement. Finally, there is a *non-social* control setting to which the results of the social settings are compared. The non-social setting is a standard experimental reinforcement learning setup using the same elements as the social setups but without the social reinforcement.

We hypothesized that robot learning (in a RL context as described above) is facilitated by additional social reinforcement. Our experimental results support this hypothesis. We compared the learning performance of our simulated robot in the social and non-social settings, by analyzing averages of learning curves. The main contribution of this research is that it presents *quantitative evidence of the fact that a human-in-the-loop can boost learning performance in real-time by communicating reinforcement using facial expressions, in a non-trivial learning environment*. We belief this is an important result. It provides a solid base for further study of human mediated robot-learning in the context of real-world applicable reinforcement learning, using the communication protocol nature has provide for that purpose, i.e., emotional expression and recognition. As such, our results add weight to the view that robots can be trained and their behaviors optimized using *natural social cues*. This facilitates human-robot interaction and is relevant to human computing [32], to which we devote more attention in the discussion.

1.2 Chapter Layout

The rest of this chapter is structured as follows. In Section 2 we review related work. In Section 3 we discuss, in some detail, affect, emotion and how affect influences learning in humans. In Section 4 we briefly introduce *EARL*, our complete framework. In Section 5 we describe how communicated affect is linked to a social reinforcement signal. In Section 6, we explain our method of study (e.g., the grid-world, the learning mechanism). Section 7 discusses the results and Section 8 discusses these in a broader context and presents concluding remarks and future work.

2 Interactive Robot Learning

One of the main reasons for investigating natural ways of giving feedback to robots in order for them to be able to adapt themselves is *non-expert interaction*; the ability of

persons not familiar with machine learning to change the behavior of robots in a natural way [30][31][43]. Robots can learn from humans in a variety of ways, and humans want to teach robots in many different ways [43]. In general there are three types of robot teaching: *by example*, *by feedback* and *by guidance*. Our paper focuses on a method of interactive robot learning by feedback, but we briefly discuss all three approaches in this section.

2.1 Learning by Example

In the case of learning by example, robots learn behavior by imitating human behavior when that behavior is provided as an example (for review see [9]). The robot either imitates the behavior, or imitates getting towards the goal using the behavior as example. In the first case the behavior is leading, in the second the intention of the behavior is leading. Sometimes, robots can even learn to imitate. This is the case when the robot not only learns to imitate behavior, but also learns to imitate in the first place. The study presented in this chapter is not related to imitative behavior, as the human tutor in our study does not communicate to the robot examples of possible behaviors as to how to find the food (solve the task).

2.2 Learning by Feedback

Our study falls into the category *learning by feedback*. In this case the robot learns by receiving performance feedback from the human in the form of an additional reinforcement signal [10][26][30][31][34[45]. Such signals can come in many forms. For example in the study by Isbell et al. [26], their social chatter bot *Cobot* learns the information preferences of its chat partners, by analyzing the chat messages for explicit and implicit reward signals (e.g., positive or negative words). These signals are then used to adapt its model of providing information to that chat partner. So, Cobot effectively uses social feedback as reward, as does our simulated robot. However, there are several important differences. Cobot does not address the issue of a human observer parenting the robot using affective communication. Instead, it learns based on reinforcement extracted from words used by the user during the chat sessions in which Cobot is participating. Also, Cobot is not a real-time behaving robot, but a chat robot. As a consequence, time constraints related to the exact moment of administering reward or punishment are less important. Finally, Cobot is restricted regarding its action-taking initiative, while our robot is continuously acting, with the observer reacting in real-time.

Thrun et al. [31] describe how their museum tour-guide robot *MINERVA* learns how to attract attention from humans as well as how to optimize tours (in terms of the time available and the exhibits visited). An important task this robot has to learn is how to attract people to its tours. The reward signal for learning this task is based on the amount of people being close to the robot; too close, however, represents a negative reinforcement, so the robot should adapt its attention-attracting behavior to maximize having a reasonably sized group waiting around it. In this study, a non-intrusive measure has been used as basis for the reinforcement signal. This is comparable with the approach used by Mitsunaga et al. [30]. They also use non-intrusive signals (robot-human distance, gaze aversion, and body repositioning) as

reinforcement signal to adapt the robot-human distance to a comfortable one. Obviously, two key differences between these studies and ours exist: we explicitly use the affective channel (facial expression) to communicate reinforcement, and we analyzed whether this reinforcement signal helps the simulated robot to solve a task that is not related to human-robot interaction per se.

Studies that are particularly related to ours are the ones by Papudesi and Huber [34][35]. They investigate if a composite reward function (composed of the normal reinforcement given by the environment and reinforcement based on human advice) enhances robot learning of a navigation problem in a grid-based maze. The human-based part of the reward function is done in quite a clever way. The robot is given a set of advice instructions on where to go first or what choice to make at junctions. This advice is in terms of state-action pairs, so, a certain action in a certain state (representing a location in the maze) is given a slight selection bias. All biases together form a bias function (over the state-action space) that can be translated to a user-based reinforcement function. This user-based reinforcement is the first part of the reward function, and is added to the environment-based reinforcement. Together this forms the composite reward function. This composite function is used for training. The interesting part is that by using this two-step approach of administering user reward, formal analysis of the maximum permissible user-reward values is possible. The authors have shown boundaries for the human advice such that several problems related to additional user rewards can be overcome, problems such as "looping" (due to intermediate user reinforcement, the robot keeps looping through the same state-action pairs). The key difference between their approach and ours is that we use the facial expression to communicate the reward function, that we have a continuous state representation (see Section 6) and that we administer the user's reinforcement directly, without a bias function. It would be interesting to merge both approaches and use facial expression to feed the bias function as defined in [34].

2.3 Learning by Guidance

Learning by guidance is a relatively new approach to interactive human-robot learning [42][43][45]. Guidance can be differentiated from feedback and imitation (example) in the following way. While feedback gives intentional information after the fact, guidance gives intentional information before the fact (*anticipatory reinforcement*; [44]). For example, smiling at a robot after it has taken the right turn towards the food (our study), is quite different from proposing a certain turn to the robot before it has chosen itself [44].

While imitation assumes a sequence of behaviors that lead towards a goal state, guidance is about future-directed learning cues and as such is much broader defined. For example, showing how to tie shoelaces is very different from drawing a child's attention to the two edges when stuck in the beginning. In general, robot guidance is about directing attention, communicating motivational intentions, and proposing actions [43].

For example in the work by Thomaz and Breazeal [43], the authors show an interesting way in which guidance can be added to a standard reinforcement learning mechanism. A human can advice the agent to pay more attention to a specific object in the problem environment (in this case a learning-to-cook environment, called

Sophie's kitchen, with a simulated robot). The guidance is transformed into an action-selection bias, such that the simulated robot selects actions that have to do with the advised object with higher probability. As such, this approach resembles the one by Papudesi and Huber [34]: the behavioral bias is given at the level of state-action pairs, not directly at the level of reward and punishment. The main reason why biasing action-selection as done by [43] can best be seen as guidance and biasing action-states as done by [34] can best be seen as feedback, is the way both studies use the human advice. In the former (T&B), the advice is immediately integrated into the action-selection, and guides the robot's next actions, while in the latter (P&H) the advice is first translated to an additional reward for certain state-action pairs and the resulting composite reward function is used for training the robot.

In a real sense, guidance by biasing action-selection can be seen as narrowing-down attention towards certain objects or features. By biasing action-selection, certain actions have a higher change of being selected than others. This kind of human-robot interaction can help solve exploration-exploitation issues in very large state spaces, as the human can guide de robot into a useful direction, while the robot still has the opportunity to explore [44].

3 Affect Influences Learning

In this chapter we specifically focus on the influence of socially communicated affect on learning, i.e., on affectively communicated feedback. Affect and emotion are concepts that lack a single concise definition, instead there are many [37]. Therefore we first explain our meaning to these concepts.

3.1 Emotion and Affect

In general, the term emotion refers to a set of—in social animals—naturally occurring phenomena including facial expression, motivation, emotional actions such as fight or flight behavior, a tendency to act, and—at least in humans—feelings and cognitive appraisal (see, e.g., [40]). An emotional state is the combined activation of instances of a subset of these phenomena, e.g., angry involves a tendency to fight, a typical facial expression, a typical negative feeling, etc. Time is another important aspect in this context. A short term (intense, object directed) emotional state is often called an *emotion*; while a longer term (less intense, non-object directed) emotional state is referred to as *mood*. The direction of the emotional state, either positive or negative, is referred to as *affect* (e.g., [39]). Affect is often differentiated into two orthogonal (independent) variables: *valence*, a.k.a. pleasure, and *arousal* [19][39]. Valence refers to the positive versus negative aspect of an emotional state. Arousal refers to the activity of the organism during that state, i.e., physical readiness. For example, a car that passes you in a dangerous manner on the freeway, immediately (*time*) elicits a strongly negative and highly arousing (*affect*) emotional state that includes the expression of anger and fear, feelings of anger and fear, and intense cognitive appraisal about what could have gone wrong. On the contrary, learning that one has missed the opportunity to meet an old friend involves cognitive appraisal that can negatively influence (*affect*) a person's mood for a whole day (*time*), even though the

associated emotion is not necessarily arousing (*affect*). Eating a piece of pie is a more positive and biochemical example. This is a bodily, emotion-eliciting event resulting in mid-term moderately-positive affect. Eating pie can make a person happy by, e.g., triggering fatty-substance and sugar-receptor cells in the mouth. The resulting positive feeling typically is not of particularly strong intensity and certainly does not involve particularly high or low arousal, but might last for several hours.

3.2 Emotional Influences

Emotion influences thought and behavior in many ways. For example, at the neurological level, malfunction of certain brain areas not only destroys or diminishes the capacity to have (or express) certain emotions, but also has a similar effect on the capacity to make sound decisions [17] as well as on the capacity to learn new behavior [4]. Behavioral evidence suggests that the ability to have sensations of pleasure and pain is strongly connected to basic mechanisms of learning and decision-making [4]. These findings indicate that brain areas important for emotions are also important for "classical" cognition and instrumental learning.

At the level of cognition, a person's belief about something is updated according to the associated emotion: the current emotion is used as information about the perceived object [14][21], and emotion is used to make the belief resistant to change [22]. Ergo, emotions are "at the heart of what beliefs are about" [23].

Emotion plays a role in the regulation of the amount of information processing. For instance, Scherer [40] argues that emotion is related to the continuous checking of the environment for important stimuli. More resources are allocated to further evaluate the implications of an event, only if the stimulus appears important enough. Furthermore, in the work of Forgas [21] the relation between emotion and information processing strategy is made explicit: the influence of mood on thinking depends on the strategy used. In addition to this, it has been found that positive moods favor creative thoughts as well as integrative information processing, while negative moods favor systematic analysis of incoming stimuli (e.g. [1][25]).

Emotion also regulates behavior of others. Obvious in human development, expression (and subsequent recognition) of emotion is important to communicate (dis)approval of the actions of others. This is typically important in parent-child relations. Parents use emotional expression to guide behavior of infants. Emotional interaction is essential for learning. Striking examples are children with an autistic spectrum disorder, typically characterized by a restricted repertoire of behaviors and interests, as well as social and communicative impairments such as difficulty in joint attention, difficulty recognizing and expressing emotion, and lacking of a social smile (for review see [13]). Apparently, children suffering from this disorder have both a difficulty in building up a large set of complex behaviors *and* a difficulty understanding emotional expressions and giving the correct social responses to these. This disorder provides a clear example of the interplay between learning behaviors and being able to process emotional cues.

As argued by Breazeal and Brooks [8], human emotion is crucial to understanding others, as well as ourselves, and this could very well be two equally crucial functions for robot emotions.

3.3 Socially Communicated Affect

To summarize, emotion and mood influence thought and behavior in a variety of ways, e.g., a person's mood influences processing style and attention, emotions influence how one thinks about objects, situations and persons, and emotion is related to learning new behaviors.

In this study we focus on the role of affect in guiding learning in a social human-robot setting. We use affect to denote the positiveness versus negativeness of a situation. We ignore the arousal a certain situation might bring. As such, positive affect characterizes a situation as good, while negative affect characterizes that situation as bad (e.g., [39]). Further, we use affect to refer to the *short term* timescale: i.e., to emotion. We hypothesize that affect communicated by a human observer can enhance robot learning. In our study we assume that the recognition of affect translates into a reinforcement signal. As such, the robot uses a *social reinforcement* in addition to the reinforcement it receives from its environment while it is building a model of the environment using reinforcement learning mechanisms. In the following sections we first explain our framework after which we detail our method and discuss results and further work.

4 EARL: A Computational Framework to Study the Relation Between Emotion, Adaptation and Reinforcement Learning

To study the relation between emotion, adaptation and reinforcement learning, we have developed an end-to-end framework. The framework consists of four parts:

- An emotion recognition module, recognizing emotional facial expression in real time.
- A reinforcement learning agent to which the recognized emotion can be fed as input.
- An artificial emotion module slot; this slot can be used to plug into the learning agent different models of emotion that produce the artificial emotion of the agent as output. The modules can use all of the information that is available to the agent (such as action repertoire, reward history, etc.). This emotion can be used by the agent as intrinsic reward, as metalearning parameter, or as input for the expression module.
- An expression module, consisting of a robot head with the following degrees of freedom: eyes moving up and down, ears moving up and down on the outside, lips moving up and down, eyelids moving up and down on the outside, and RGB eye colors.

Emotion recognition is based on quite a crude mechanism based upon the face tracking abilities of OpenCV [48]. Our mechanism uses 9 points on the face, each defined by a blue sticker: 1 on the tip of the nose, 2 above each eyebrow, 1 at each mouth corner and 1 on the upper and lower lip. The recognition module is configured to store multiple prototype point constellations. The user is prompted to express a certain emotion and press space while doing so. For every emotional expression (in the case of our experiment neutral, happy and afraid), the module records the

positions of the 9 points relative to the nose. This is a prototype point vector. After configuration, to determine the current emotional expression in real time the module calculates a weighted distance from the current point vector (read in real-time from a web-cam mounted on the computer screen) to the prototype vectors. Different points get different weights. This results in an error measure for every prototype expression. This error measure is the basis for a normalized vector of recognized emotion intensities. The recognition module sends this vector to the agent (e.g., neutral 0.3, happy 0.6, fear 0.1). Our choice of weights and features has been inspired by work of others (for review see [33]). Of course the state of the art in emotion recognition is more advanced than our current approach. However, as our focus is affective learning and not the recognition process per se, we contented ourselves with a low fidelity solution (working almost perfectly for neutral, happy and afraid, when the user keeps the head in about the same position).

Note that we do not aim at generically recognizing detailed emotional expressions. Instead, we tune the recognition module to the individual observer to accommodate his/her personal and natural facial expressions. The detail with which this is done reflects our experimental needs: extract positive and negative reward signals from the observer's face. In a real-world scenario with observers and robots autonomously acting next to each other, a more sophisticated mechanism is needed to correctly read reward signals from the observers. Such a mechanism needs to be multi-modal.

The reinforcement learning agent receives this recognized emotion and can use this in multiple ways: as reinforcement, as information (additional state input), as metaparameter (e.g., to control learning rate), and as social input directly into its emotion model. In this chapter we focus on social reinforcement, and as such focus on the recognized emotion being used as additional reward or punishment. The agent, its learning mechanism and how it uses the recognized emotion as reinforcement are detailed in Sections 5 and 6.

The artificial emotion model slot enables us to plug in different emotion models based on different theories to study their behavior in the context of reinforcement learning. For example, we have developed a model based on the theory by Rolls [38], who argues that many emotions can be related to reward and punishment and the lack thereof. This model enables us to see if the agent's situation results in a plausible (e.g., scored by a set of human observers) emotion emerging from the model. By scoring the plausibility of the resulting emotion, we can learn about the compatibility of, e.g., Rolls' emotion theory with reinforcement learning. However, in the current study we have not used this module, as we focus on affective input as social reinforcement.

The emotion expression part is a physical robot head. The head can express an arbitrary emotion by mapping it to its facial features, again according to a certain theory. Currently our head expresses emotions according to the Pleasure Arousal Dominance (PAD) model by Mehrabian [29]. We have a continuous mapping from the 3-dimensional PAD space to the features of the robot face. As such we do not need to explicitly work with emotional categories or intensities of the categories. The mapping appears to work quite well, but is in need of validation (again using human observers). We have not used the robot head for the studies reported upon in this chapter.

We now describe in detail how we coupled the recognized human emotion to the social reinforcement signal for the robot. Then we explain in detail our adapted reinforcement learning mechanism (such that it enabled learning in continuous environments), and our method of study as well as our results.

5 Emotional Expressions as Reinforcement Signal

As mentioned earlier, emotional expressions and facial expressions in particular can be used as social cues for the desirability of a certain action. In other words, an emotional expression can express reward and punishment if directed at an individual. We focus on communicated affect, i.e., the positiveness versus negativeness of the expression. If the human expresses a smile (happy face) this is interpreted as positive affect. If the human expresses fear, this is interpreted as negative affect. We interpret a neutral face as affectless.

We have studied the mechanism of communicated affective feedback in a human-robot interaction setup. The human's face is analyzed (as explained above) and a vector of emotional expression intensities is fed to the learning agent. The agent takes the expression with the highest intensity as dominant, and equates this with a *social reinforcement* of, e.g., 2 (happy), −2 (fear) and 0 (neutral). It is important to realize that this is a simplified setup, as the human face communicates much more subtle affective messages and at the very least is able to communicate the degree of reward and punishment. For example, fear and anger are two distinct negative emotions that have different meaning and different action-tendencies. Fear involves a tendency to avoid and is not directed at an individual (although it can be caused by an individual), while anger involves a tendency to approach and is outwardly directed at someone else. A little bit of anger might be interpreted as a little bit of punishment, while a lot of anger better be interpreted as "don't ever do this again". However, to investigate our hypothesis (affective human feedback increases robot learning performance) the just described mechanism is sufficient. For the sake of simplicity, in this experiment we take fear as a "prototype for negative affective facial expression" and happiness as a "prototype for positive affective facial expression".

The social reinforcement, called r_{social}, is simply added to the "normal" reinforcement the agent receives from its environment (together forming a composite reinforcement). So, if the agent walks on a path somewhere in the gridworld, it receives a reinforcement (say 0), but when the user smiles, the resulting actual reinforcement becomes 2, while if the user looks afraid, the resulting reinforcement becomes −2.

6 Method

To study the impact of social reinforcement on robot learning, we have used our framework in a simulated continuous gridworld. In this section we explain our experimental setup.

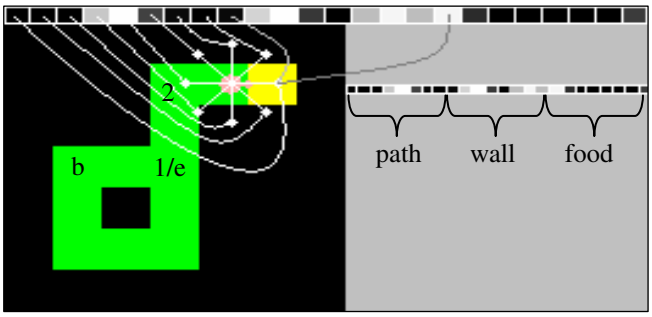

Fig. 1. The experimental gridworld. The agent is the "circle with nose" in the top right of the maze, where the nose denotes its direction. The 8 white dots denote the points perceived by the agent. These points are connected to the elements of state *s* (neural input to the MLPs used by the agent) as depicted. This is repeated for all possible features, in our case: path (gray), wall (black), and food (light gray), in that order. The *e* denotes the cell in which social reinforcement can be administered through smiling or expression of fear, the *1* and *2* denote key locations at which the agent has to learn to differentiate its behavior, i.e., either turn left (*1*) or right (*2*). The agent starts at *b*. The task enforces a non-reactive best solution (by which we mean that there is no direct mapping from reinforcement to action that enables the agent to find the shortest path to the food). If the agent would learn that turning right is good, it would keep walking in circles. If the agent learns that turning left is good, it would not get to the food.

6.1 Continuous Gridworld as Test Environment

A simulated robot (agent) "lives" in a continuous gridworld environment consisting of wall, food and path patches (Figure 1). These are the features of the world observable by the agent. The agent cannot walk on walls, but can walk on path and food. Walls and path are neutral (have a reinforcement of 0.0), while food has a reinforcement of 10. One cell in the grid is assumed to be a 20 by 20 object. Even though wall, path and food are placed on a grid, the world is continuous in the following sense: the agent has real-valued coordinates, moves by turning or walking in a certain direction using an arbitrary speed (in our experiments set at 3), and perceives its direct surroundings (within a radius of 20) according to its looking direction (one out of 16 possible directions). The agent uses a "relative eight neighbor metric" meaning that it perceives features of the world at 8 points around it, with each point at a distance of 20 from the center point of the agent and each point at an interval of 1/4 PI radians, with the first point always being exactly in front of it (Figure 1). The state perceived by the agent (its percept) is a real-valued vector of inputs between 0 and 1; each input is defined by the relative contribution of a certain feature in the agent-relative direction corresponding to the input. For example, if the agent sees a wall just in front of it (i.e., the center point of a wall object is exactly at a distance of 20 as measured from the current agent location in its looking direction) the first value in its perceived state would be equal to 1. This value can be anywhere between 0 and 1 depending on the distance of that point to the feature. For the three types of features, the agent thus has 3x8=24 real-valued inputs between 0 and 1 as its perceived world state *s* (Figure 1). As such the agent can approach objects (e.g., a wall) from a large number of possible angles and positions, with every intermediate position being possible. For

all practical purposes, the learning environment can be considered continuous. We did not define discrete states based on the continuous input to facilitate learning. Instead we chose to use the perceived state as is, to maximize potential transferability of our experimental results to real-world robot learning.

6.2 Reinforcement Learning in Continuous Environments

Reinforcement learning in continuous environments introduces several important problems for standard RL techniques, such as Q learning, mainly because a large number of potentially similar states exist as well as a very long path length between start and goal states making value propagation difficult. We now briefly explain our adapted RL mechanism. As RL in continuous environments is not specifically the topic of the chapter we have left out some of the rational for our choices.

The agent learns to find the path to the food, and optimizes this path. At every step the agent takes, the agent updates its model of the expected benefit of a certain action as follows. It learns to predict the value of actions in a certain perceived state s, using an adapted form of Q learning. The value function, $Q_a(s)$, is approximated using a Multi-Layer Perceptron (MLP), with 3x8=24 input, 24 hidden, and one output neuron(s), with s being the real-valued input to the MLP, a the action to which the network belongs, and the output neuron converging to $Q_a(s)$. As such, every action of the agent (5 in total: forward, left, right, left and forward, right and forward) has its own network. The output of the action networks are used as action values in a standard Boltzmann action-selection function [41]. An action network is trained on the Q value—i.e., $Q_a(s) \leftarrow Q_a(s)+\alpha(r+\gamma Q(s')-Q_a(s))$ —where r is the reward resulting from action a in state s, s' is the resulting next state, $Q(s')$ the value of state s', α is the learning rate and γ the discount factor [41]. The learning rate equals 1 in our experiments (because the learning rate of the MLP is used to control speed of learning, not α), and the discount factor equals 0.99. To cope with a continuous gridworld, we adapted standard Q learning in the following way:

First, the value $Q_a(s)$ used to train the MLP network for action a is topped such that $min(r, Q_a(s')) <= Q_a(s) <= max(r, Q(s'))$. As a result, individual $Q_a(s)$ values can never be larger or smaller than any of the rewards encountered in the world. This enables a discount factor close to or equal to 1, needed to efficiently propagate back the food's reward through a long sequence of steps. In continuous, cyclic, worlds, training the MLP on normal Q values using a discount factor close to 1 can result in several problems not further discussed here.

Second, per step of the agent, we train the action-state networks not only on $Q_a(s) \leftarrow Q_a(s)+\alpha(r+\gamma Q(s')-Q_a(s))$ but also on $Q_a(s') \leftarrow Q_a(s')$. The latter seems unnecessary but is quite important. RL assumes that values are propagated *back*, but MLPs generalize while trained. As a result, training an MLP on $Q_a(s)$ also influences its value prediction for s' in the same direction, just because the inputs are very close. In effect, part of the value is actually propagated *forward*; credit is partly assigned to what comes next. This violates the RL assumption just mentioned. Note that the value $Q(s')$ is predicted using another MLP, called the value network, that is trained in the same way as the action networks using the topped-off value and forward propagation compensation.

Third, for the agent to better discriminate between situations that are perceptually similar, such as position "1" and "2" in Figure 1, for each action-network the agent also uses a second network trained on the value of *not* taking the action. This network is trained when other actions are taken but not when the action to which the "negation" network belongs is taken. In effect, the agent has two MLPs per action. This enables the agent to better learn that, e.g., "right" is good in situation "2" but *not* in situation "1". Without this "negation" network, the agent learns much less efficient (results not shown). To summarize, our agent has 5 actions, it has 11 MLPs in total: one to train $Q(s)$, 5 to train $Q_a(s)$ and 5 to train $-Q_a(s)$. All networks use forward propagation compensation and a topped-off value to train upon. The MLP predictions for $Q_a(s)$ and $-Q_a(s)$ are simply added, and the result is used for action-selection.

6.3 Social vs. Non-social Learning

To study the effect of communicated affect as social reinforcement, we created the following setup. First an agent is trained without social reinforcement. The agent repeatedly tries to find the food for 200 trials, i.e., one *run*. The agent continuously learns and acts during these trials. To facilitate learning, we use a common method to vary the MLP learning rate and the Boltzmann action selection β derived from simulated annealing. The Boltzmann β equals to $3+(trial/200)*(6-3)$, effectively varying from 3 in the first trial to 6 in the last. The MLP learning rate equals to $0.1-(trial/200)*(0.1-0.001)$ effectively varying from 0.1 in the first trial to 0.001 in the last. We repeated the experiment 200 times, resulting in 200 runs. Average learning curves are plotted for these 200 runs using a linear smoothing factor equal to 6 (Figure 2).

Second, a new agent is trained *with* social reinforcement, i.e., a human observer looking at the agent with his/her face analyzed by the agent, translating a smile to a social reward and a fearful expression to a social punishment. Again, average learning curves are plotted using a linear smoothing factor equal to 6, but now based on the average per trial over 15 runs (Figure 2). We experimented with three different social settings: (a) a moderate social reinforcement, r_{human}, from trial 20 to 30, where the social reinforcement is either −0.5 or 0.5 (happy vs. fearful, respectively); (b) a strong social reinforcement, r_{human}, from trial 20 to 25 where social reinforcement is either −2 or 2, i.e., more extreme social reinforcement but for a shorter period; (c) a social reinforcement, r_{human}, from trial 29 to 45 where social reinforcement is either −2 or 2 while (in addition to settings *a* and *b*) the agent trains an additional MLP to predict the direct social reinforcement, r_{human}, based on the current state s. The MLP is trained to learn $R_{social}(s)$ as given by the human reinforcement r_{human}. After trial 45, the direct social reinforcement from the observer, r_{human}, is replaced by the learned social reinforcement $R_{social}(s)$. So, during the critical period (the trial intervals mentioned) of social setting *a*, *b* and *c*, the total reinforcement is a composite reward equal to $R(s)+r_{human}$. Only in setting *c*, and only after the critical period until the end of the run, the composite reward equals $R(s)+R_{social}(s)$. In all other periods, the reinforcement is as usual, i.e., $R(s)$. As a result, in setting *c* the agent can continue using an additional social reinforcement signal that has been learned based on what its human tutor thinks about certain situations.

The process of giving affective feedback to a reinforcement learning agent appeared to be quite a long, intensive and attention absorbing experience. As a result, it was physically impossible to observe the agent during all trials in the entire gridworld (after 2 hours of smiling to a computer screen one is exhausted *and* has burning eyes and painful facial muscles). To be able to test our hypothesis, we restricted direct social input to (I) a critical learning period defined in terms of a start and end trail (as discussed above), and (II) the cell indicated by *e* (Figure 1). Only when the agent moves around in this cell *and* is in a social input trial, the simulation speed of the experiment is set to one action per second enabling affective feedback.

7 Results

The results clearly show that learning is facilitated by social reinforcement. In all three social settings (Figure 2a, b and c) the agent needs fewer steps to find the food during the trials in which the observer provides assistance to the agent by expressing positive or negative affect. Interestingly, at the moment the observer stops reinforcing, the agent gradually looses the learning benefit it had accumulated. This is independent of the size of the social reinforcement (both social learning curves in Figure 2a and b show dips that eventually return to the non-social learning curve).

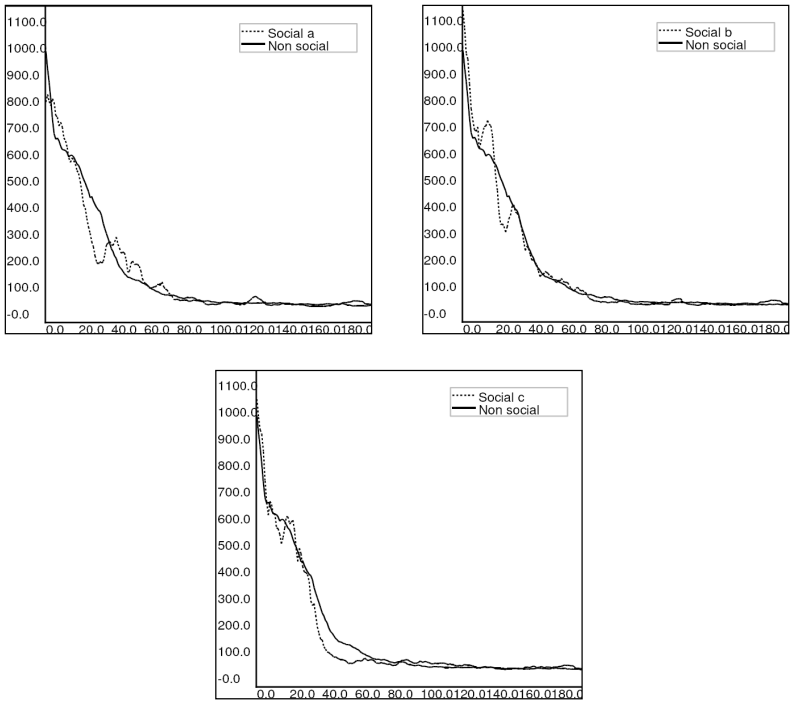

Fig. 2. Results of the learning experiments. From top to bottom showing the difference between the non-social setting and social setting *a*, *b*, and *c* respectively.

Loss of learning benefit as observed in social settings *a* and *b* (learning curve moving up, starting at trial 30 and 25 respectively) can be easily explained. The social reinforcement was not given long enough for the agent to internalize the path to the food (i.e., propagate back the food's reward to the beginning of the path). As soon as the observer stops reinforcing, the agent starts to forget these rewards, i.e., the MLPs are again trained to predict values as they are without social reinforcement. So, either the observer should continue to reinforce until the agent has internalized the solution, or the agent needs to be able to build a representation of the social reward function and use it when direct social reinforcement is not available. We have experimented with the second (social setting *c*): we enabled the agent to learn the social reward function. Now the agent uses direct social reinforcement at the emotional input spot (*e*, Figure 1) during the critical period, and uses its social reward prediction, $R_{social}(s)$, when direct social reinforcement stops. Results clearly show that the agent is now able to keep the benefit it had accumulated from using social reinforcement (Figure 2c). These results show that a combination of using social reinforcement and learning a social reward function facilitates robot learning, by enabling the robot to quicker learn the optimal solution to the food due to the direct social reinforcement as well as keep that solution by using its learned social reward function when social reinforcement stops.

8 Conclusion, Discussion and Future Work

Our results show that affective interaction in human-in-the-loop learning can provide a significant benefit to the efficiency of a reinforcement learning robot in a continuous grid world. We believe our results are particularly important to human-robot interaction for the following reasons. First, advanced robots such as robot companions, robot workers, etc., will need to be able to adapt their behavior according to human feedback. For humans it is important to be able to give such feedback in a natural way, e.g., using emotional expression. Second, humans will not want to give feedback all the time, it is therefore important to be able to define critical learning periods as well as have an efficient social reward system. We have shown the feasibility of both. Social input during the critical learning periods was enough to show a learning benefit, and the relatively easy step of adding an MLP to learn the social reward function enabled the robot to use the social reward when the observer is away.

We have specifically used an experimental setup that is compatible with a real-world robot: we have used continuous inputs and MLP-based training of which it is known that it can cope with noise and generalize over training examples. As such we believe our results can be generalized to real-world robotics. However, this most certainly needs to be experimented with.

A related issue is that the training time needed to learn our (arguably) simple task is quite long. This is also due to the representational format of the environment resulting in long state-action sequences to the goal state with states that resemble each other quite a lot. A discrete world with less, more discriminative, states can use a standard form of reinforcement learning and will show a more marked effect of intermediate social reinforcement.

Future work includes a broader evaluation of the EARL framework including its ability to express emotions generated by an emotional model plugged into the RL agent. Further, we envision to experiment with controlling metaparameters (such as exploration/exploitation and learning rate) based on the agent's internal emotional state or social rewards [3][11][18]. Currently we use simulated annealing-like mechanisms to control these parameters. Further, the agent could try to learn what an emotional expression predicts. In this case, the agent would use the emotional expression of the human in a more pure form (e.g., as a real-valued vector of facial feature intensities as part of its perceived state *s*). This might enable the agent to learn what the emotional expression means for itself instead of simply using it as reward.

As mentioned earlier, no distinction is made between different facial expressions that portray positive emotions or negative emotions. For example, no difference is made between the meaning of sadness versus anger. Thus, the current setup is highly simplified regarding the type of information that can be communicated through the affective channel. Future work includes a coupling between other reinforcement learning parameters and other aspects of facial expressions. For example, fear portrays a future danger and as such could be used by the agent to reconsider its current actions used for action-selection. Anger communicates a form of blame: the agent should have known better. This could be used to reevaluate (and perhaps internally simulate) a stored sequence of recent interactions in order to come up with an alterative, more positive, outcome than the current one.

Another way to extend this work is proposed by Thomaz, Hoffman and Breazeal [45]. Currently, the simulated robot is influenced by the human observer only at a certain spot in the maze. This is quite limited. However, it has been proposed [7][45] that human tutors could very well use a robot's behavioral cues as a signal to intervene with the learning process. For example, agent-to-teacher signals such as gaze, gesture and hesitation could be used by the tutor to, for example, propose actions to the agent, give motivational feedback etc. [45]. In our setup, we have often observed behavior that can be characterized as hesitation (e.g., the simulated robot switching between turning left and right but not deciding on really making the complete turn to take a branch in the maze). It would be interesting to allow the human tutor to influence the robot more freely, and to investigate if (1) humans tend to recognize hesitation behavior in our setup and (2) if affective feedback can still be used in these circumstances or whether a more guidance-based approach is need at these hesitation moments.

With regards to human computing [32], our work shows two things: *real-time* natural feedback (in our case, facial expressions) is feasible and desirable for robot learning, and, a personalized reward function can be learned based on this real-time interaction. This is relevant to two issues in human computing: *dynamics* and *learning/education*. Our work quantitatively shows that dynamic interaction with a simulated learning robot, using natural means of input (face) instead of traditional means (keyboard, mouse) enhances learning of robot behavior. Our interpretation of *dynamics* in this paper is somewhat different from the *dynamics* as meant in [32], i.e., the dynamics of the behavioral cue itself and the problem of deciphering these dynamics. Nevertheless, dynamic interaction is an important issue to human computing: one would not want to have to stop the robot before feedback can be communicated. Regarding the second issue, i.e., how to learn the user specific

meaning to an interactive pattern, we have shown that it is feasible to learn in real-time a personalized social reward function that the robot can use to train itself. A straightforward addition to this work would be to learn multiple social reward functions depending on, e.g., user and task context, such that the robot can select which reward function to use in what context. This would help the robot adapt to new contexts in a lazy and unsupervised way [32]. As we have already mentioned earlier, one could use a strategy in which the robot does not directly couple reward to facial expressions, but instead *learns* to couple facial features to reward. Now, a robot could first learn what different expressions mean to different users, and subsequently use the appropriate reward function to adapt its behavior.

Finally, a somewhat futuristic possibility is actually quite close: affective Robot-Robot interaction. Using our setup, it is quite easy to train one robot in a certain environment (parent), make it observe an untrained robot in that same environment (child), and enable it to express its emotion as generated by its emotion model using its robot head, an expression recognized and translated into social rewards by the child robot. Apart from the fact that it is somewhat dubious if such a setup is actually useful (why not send the social reward as a value through a wireless connection to the child), it would enable robots to use the same communication protocol as humans.

Regarding the "usefulness" argument just put forward, it seems to apply to our experiment as well. Why didn't we just simulate affective feedback by pushing a button for positive reward and pushing another for negative reward (or even worse, by simulating a button press)? From the point of view of the robot this is entirely true, however, from the point of view of the human—and therefore the point of view of the human-robot interaction—not at all. Humans naturally communicate social signals using there face, not by pushing buttons. The process of expressing an emotion is quite different from the process of pushing a button, even if it was only for the fact that it takes more time and effort to initiate the expression and that the perception of an expression is the perception of a process not a discrete event (like a button press). In a real-world scenario with a mobile robot in front of you it would be quite awkward to have to push buttons instead of just smile when you are happy about its behavior. Further it would be quite useful if the robot could recognize you being happy or sad and gradually learn to adapt its behavior even when you did not intentionally give it a reward or punishment. Abstracting away from the actual affective interaction patterns between the human and the robot in our experiment would have rendered the experiment almost completely trivial. Nobody would be surprised to see that the robot learns better if an intermediate reward is given halfway its route towards food. Our aim was to investigate if affective communication can enhance learning in a reinforcement learning setting. Taking out the affective part would have been quite strange indeed.

Acknowledgments. We would like to sincerely thank Pascal Haazebroek and all the students who helped us develop the *EARL* system, thereby making this research possible. Joris Slob, Chris Detweiler, Sylvain Vriens, Koen de Geringel, Hugo Scheepens, Remco Waal, Arthur de Vries, Pieter Jordaan, Michiel Helvensteijn, Rogier Heijligers, Willem van Vliet, you were great!

References

1. Ashby, F. G., Isen, A. M., Turken, U.: A Neuro-psychological theory of positive affect and its influence on cognition. Psychological Review 106 (3) (1999) 529-550
2. Avila-Garcia, O., Cañamero, L.: Using hormonal feedback to modulate action selection in a competitive scenario. In: From Animals to Animats 8: Proc. 8th Intl. Conf. on Simulation of Adaptive Behavior. MIT Press, Cambridge MA (2004) 243-252
3. Belavkin, R. V.: On relation between emotion and entropy. In: Proc. of the AISB'04 Symposium on Emotion, Cognition and Affective Computing. AISB Press (2004) 1-8
4. Berridge, K. C.: Pleasures of the brain. Brain and Cognition 52 (2003) 106-128
5. Blanchard, A. J., Cañamero, L.: Modulation of exploratory behavior for adaptation to the context. In: Proc. of the AISB'06 Symposium on Biologically Inspired Robotics (Biro-net). AISB Press (2006) 131-137
6. Botelho, L. M., Coelho, H.: Information processing, motivation and decision making. In: Proc. 4th International Workshop on Artificial Intelligence in Economics and Management. (1998)
7. Breazeal, C.: Affective interaction between humans and robots. In: J. Keleman, P. Sosik (eds): Proc. of the ECAL 2001. LNAI, Vol. 2159. Springer-Verlag, Berlin Heidelberg New York (2001) 582-591
8. Breazeal, C., Brooks R.: Robot emotion: A functional perspective. In: J.-M. Fellous, M. Arbib (eds.): Who needs emotions: The brain meets the robot. Oxford University Press USA (2004) 271-310
9. Breazeal, C., Scassellati, B.: Robots that imitate humans. Trends in Cognitive Sciences 6(11) (2002) 481-487
10. Breazeal, C., Velasquez, J.: Toward teaching a robot `infant' using emotive communication acts. In: Edmonds, B., Dautenhahn, K. (eds.): Socially Situated Intelligence: a workshop held at SAB'98, Zürich. University of Zürich Technical Report (1998) 25-40
11. Broekens, J., Kosters, W. A., Verbeek, F. J.: On emotion, anticipation and adaptation: Investigating the potential of affect-controlled selection of anticipatory simulation in artificial adaptive agents. In press (2007)
12. Cañamero, D.: Designing emotions for activity selection. Dept. of Computer Science Technical Report DAIMI PB 545. University of Aarhus Denmark (2000)
13. Charman, T., Baird, G.: Practitioner review: Diagnosis of autism spectrum disorder in 2- and 3-year-old children. Journal of Child Psychology and Psychiatry 43(3) (2002) 289-305
14. Clore, G. L. Gasper, K.: Feeling is believing: Some affective influences on belief. In: Frijda, N., Man-stead A. S. R., Bem, S. (eds.): Emotions and Beliefs. Cambridge Univ. Press, Cambridge UK (2000) 10-44
15. Cos-Aguilera, I., Cañamero, L., Hayes, G. M., Gillies, A.: Ecological integration of affordances and drives for behaviour selection. In: Proc. of the Workshop on Modeling Natural Action Selection. AISB Press (2005) 225-228
16. Custers, R., Aarts, H.: Positive affect as implicit motivator: On the nonconscious operation of behavioral goals. Journal of Personality and Social Psychology 89(2) (2005) 129-142
17. Damasio, A. R.: Descartes' error. Penguin Putnam, New York NY (1994)
18. Doya, K.: Metalearning and neuromodulation. Neural Networks 15 (4) (2002) 495-506
19. Dreisbach, G., Goschke, K.: How positive affect modulates cognitive control: Reduced perseveration at the cost of increased distractibility. Journal of Experimental Psychology: Learning, Memory, and Cognition 30(2) (2004) 343-353

20. Fong, T., Nourbakhsh, I., Dautenhahn, K.: A survey of socially interactive robots. Robots and Autonomous Systems 42 (2003) 143-166
21. Forgas, J. P.: Feeling is believing? The role of processing strategies in mediating affective influences in beliefs. In: Frijda, N., Manstead A. S. R., Bem, S. (eds.): Emotions and Beliefs. Cambridge University Press, Cambridge UK (2000) 108-143
22. Frijda, N. H., Mesquita, B.: Beliefs through Emotions. In: Frijda, N., Manstead A. S. R., Bem, S. (eds.): Emotions and Beliefs. Cambridge University Press, Cambridge UK (2000) 45-77
23. Frijda, N. H., Manstead, A. S. R., Bem. S.: The influence of emotions on beliefs. In: Frijda, N., Man-stead A. S. R., Bem, S. (eds.): Emotions and Beliefs. Cambridge University Press, Cambridge UK (2000) 1-9
24. Gandanho, S. C.: Learning behavior-selection by emotions and cognition in a multi-goal robot task. Journal of Machine Learning Research 4 (2003) 385-412
25. Gasper, K., Clore, L. G.: Attending to the big picture: Mood and global versus local processing of visual information. Psychological Science 13(1) (2002) 34-40
26. Isbell, C. L. Jr., Shelton, C. R., Kearns, M., Singh, S., Stone, P.: A social reinforcement learning agent. In: Proceedings of the fifth international conference on Autonomous agents. ACM (2001) 377-384
27. Kaelbling, L. P., Littman, M. L., Moore, A. W.: Reinforcement learning: A survey. Journal of Artificial Intelligence Research 4 (1996) 237-285
28. Lin, L. J.: Reinforcement learning for robots using neural networks. Doctoral dissertation. Carnegie Mellon University, Pittsburgh (1993)
29. Mehrabian, A.: Basic Dimensions for a General Psychological Theory. OG&H Publishers, Cambridge Massachusetts (1980)
30. Mitsunaga, N., Smith, C., Kanda, T., Ishiguro, H., Hagita, N.: Robot behavior adaptation for human-robot interaction based on policy gradient reinforcement learning. In: Proc. Of the International Conference on Intelligent Robots and Systems (IROS). IEEE Press (2005) 218-225
31. Thrun, S., Bennewitz, M., Burgard, W., Cremers, A. B., Dellaert, F., Fox, D., Hähnel, D., Rosenberg, C. R., Roy, N., Schulte, J., Schulz, D.: A tour-guide robot that learns. In: Burgard, W., Christaller, T., Cremers, A.B. (eds.): Proc. of the 23rd Annual German Conference on Artificial Intelligence: Advances in Artificial Intelligence. LNAI, Vol. 1701. Springer-Verlag, London UK (1999) 14-26
32. Pantic, M., Pentland, A., Nijholt, A., Huang, T. S.: Human computing and machine understanding of human behavior: A Survey. Proc. ACM Int'l Conf. Multimodal Interfaces (2006) 239-248
33. Pantic, M., Rothkranz, L. J. M.: Automatic analysis of facial expressions: The state of the art. IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (12) (2000) 1424-1445
34. Papudesi, V. N., Huber, M.: Learning from reinforcement and advice using composite reward functions. In: Proc. Of the 16th International FLAIR Conference. AAAI (2003) 361-365
35. Papudesi, V. N., Huber, M.: Interactive refinement of control policies for autonomous robots. In: Proc. Of the 10th IASTED International Conference on Robotics and Applications, Honolulu HI. IASTED (2004)
36. Picard, R. W.: Affective Computing. MIT Press, Cambridge MA (1997)
37. Picard, R. W., Papert, S., Bender, W., Blumberg, B., Breazeal, C. Cavallo, D., Machover, T., Resnick, M., Roy, D., Strohecker, C.: Affective learning — A manifesto. BT Technology Journal 22(4) (2004) 253-269

38. Rolls, E. T.: Précis of The brain and emotion. Behavioral and Brain Sciences 23 (2000) 177-191
39. Russell, J. A.: Core affect and the psychological construction of emotion. Psychological Review 110(1) (2003) 145-72
40. Scherer, K. R.: Appraisal considered as a process of multilevel sequential checking. In: K. R. Scherer, A. Schorr, T. Johnstone (eds.): Appraisal processes in emotion: Theory, Methods, Research. Oxford Univ. Press, New York NY (2001) 92-120
41. Sutton, R., Barto, A.: Reinforcement learning: An introduction. MIT Press, Cambridge MA (1998)
42. Ogata, T., Sugano, S., Tani, J.: Open-end human robot interaction from the dynamical systems perspective: Mutual adaptation and incremental learning. In: Orchard, R., Yang, C., Ali, M. (eds.): 17th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems. LNCS, Vol. 3029. Springer (2004) 435-444
43. Thomaz, A.L., Breazeal, C.: Teachable characters: User studies, design principles, and learning performance. In: Gratch, J., Young, M., Aylett, R., Ballin, D., Olivier, P. (eds.): Proc. of the 6th International Conference on Intelligent Virtual Agents (IVA 2006). LNCS, Vol. 4133, Springer (2006) 395-406
44. Thomaz, A. L., Breazeal, C.: Reinforcement learning with human teachers: Evidence of feedback and guidance with implications for learning performance. In: Proc. of the 21st National Conference on Artificial Intelligence. AAAI Press (2006b)
45. Thomaz, A.L., Hoffman, G., Breazeal, C.: Real-time interactive reinforcement learning for robots. In: Proc. of AAAI Workshop on Human Comprehensible Machine Learning. Pittsburgh, PA (2005)
46. Velasquez, J. D.: A computational framework for emotion-based control. In: SAB'98 Workshop on Grounding Emotions in Adaptive Systems (1998)
47. Wright, I.: Reinforcement learning and animat emotions. In: From Animals to Animats 4: Proc. of the 4th International Conference on the Simulation of Adaptive Behavior (SAB): MIT Press, Cambridge MA (1996) 272-284
48. OpenCV: http://www.intel.com/technology/computing/opencv/index.htm