Towards Transparent Robot Learning through TDRL-based Emotional Expressions

Joost Broekens, Mohamed Chetouani

Towards Transparent Robot Learning through TDRL-based Emotional Expressions

Joost Broekens, Mohamed Chetouani

***Abstract*—Robots and virtual agents need to adapt existing and learn novel behavior to function autonomously in our society. Robot learning is often in interaction with or in the vicinity of humans. As a result the learning process needs to be transparent to humans. Reinforcement Learning (RL) has been used successfully for robot task learning. However, this learning process is often not transparent to the users. This results in a lack of understanding of what the robot is trying to do and why. The lack of transparency will directly impact robot learning. The expression of emotion is used by humans and other animals to signal information about the internal state of the individual in a language-independent, and even species-independent way, also during learning and exploration. In this article we argue that simulation and subsequent expression of emotion should be used to make the learning process of robots more transparent. We propose that the TDRL Theory of Emotion gives sufficient structure on how to develop such an emotionally expressive learning robot. Finally, we argue that next to such a generic model of RL-based emotion simulation we need personalized emotion interpretation for robots to better cope with individual expressive differences of users.**

***Index Terms*—Robot learning, Transparency, Emotion, Reinforcement Learning, Temporal Difference.**

I. Introduction

Most of the envisioned applications of robotics assume efficient human-robot collaboration mediated by effective exchange of social signals. Models and technologies allowing robots to engage humans in sophisticated forms of social interaction are required. In particular, when humans and robots have to work on common goals that cannot otherwise be achieved by individuals alone, explicit communication transmits overt messages containing information about the task at hand, while implicit communication transmits information about attitudes, coordination, turn taking, feedback and other types of information needed to regulate the dynamics of social interaction. On top of that, the diversity of tasks and settings in which robots (and virtual agents) need to operate prohibits preprogramming all necessary behaviors in advance. Such agents need to learn novel, and adapt existing, behavior.

Reinforcement Learning (RL) [1], [2] is a well-established computational technique enabling agents to learn skills by trial and error, for example learning to walk [3]. Also - given sufficient exploration - RL can cope with large state spaces when coupled to pattern detection using deep-learning [4]. In RL, learning a skill is to a large extent shaped by a feedback signal, called the reward. Through trial and error, a learning robot or virtual character adjusts its behavior to maximize the expected cumulative reward, i.e., to learn the optimal policy.

Robots need to learn autonomously but also in interaction with humans [5], [6]. Robot learning needs a human in the loop. As a consequence, robots must have some degree of awareness of human actions and decisions and must be able to synthesize appropriate verbal and non-verbal behaviors. Human emotion expression is a natural way to communicate social signals, and emotional communication has been shown to be essential in the learning process of infants [7], [8], [9], [10]. Currently, however, there is no clear approach how to generate and interpret such non-verbal behavior in the context of a robot that learns tasks using reinforcement learning.

In this position paper we argue that if robots are to develop task-related skills in similar ways as children do, i.e., by trial and error, and by expressing emotions for help and confirmation and learning from emotions for shaping their behavior, they will need the affective abilities to express and interpret human emotions *in the context of their learning process*. Endowing robots the ability to learn new tasks with humans as tutors or observers will necessarily improve the performance, the acceptability and the adaptability to different preferences. In addition, this approach is essential to engage users lacking programming skills and consequently broaden the set of potential users to include children, elderly people, and other non-expert users.

In this paper we specifically focus on the communicative role of emotion in robots that learn using Reinforcement Learning in interaction with humans. We argue that, even though the RL learning method is powerful as a task learning method, it is not transparent for the average user. We propose that emotions can be used to make this process more transparent, just like in nature. For this we need a computational model of emotion based on reinforcement learning that enables agents to (a) select the appropriate emotional expressions to communicate to humans the state of their learning process, and, (b) interpret detected human emotions in terms of learning signals. We propose that the Temporal Difference Reinforcement Learning (TDRL) Theory of Emotion [11] provides the necessary structure for such a model, and we highlight remaining challenges.

II. Interactive Robot Learning

Interactive Robot Learning deals with models and methodologies allowing a human to guide the learning process of the robot by providing it teaching signals [6]. Interactive Robot Learning schemes are designed with the assumption that teaching signals are provided by experts. Usual teaching signals

Joost Broekens is with the Department of Intelligent Systems of Delft University of Technology, Delft, the Netherlands. E-mail: joost.broekens@gmail.com

M. Chetouani is with the Institute for Intelligent Systems and Robotics, CNRS UMR7222, Sorbonne University, Paris, France. E-mail: mohamed.chetouani@sorbonne-universite.fr

include instructions [12], [13] advice [14], demonstrations [15], guidance [16], [17] and evaluative feedback [5], [17].

The learning schemes could be considered as a transfer learning approach from the human expert to the robot learner. The level of expertise of the human is rarely challenged and mostly considered as ground truth. However, when naive users teach robots, either by demonstration or guidance, this may lead to low quality, or sparse, teaching signals from which it will be hard to learn. For example, in [18], imitation learning performed with children with autism spectrum disorders results in lower performance compared to learning with typical children. In [19], the authors studied models of human feedback and show that these are not independent of the policy the agent is currently following.

Designing effective Interactive Robot Learning mechanisms requires to tackle several challenges. Here, we report the ones that are related to the role of emotion and transparency:

- Developing appropriate learning algorithms. In contrast to the recent trends in machine learning, robots have to learn from little experiences and sometimes from inexperienced users. There is a need to develop new machine learning methods that are able to deal with suboptimal learning situations while ensuring generalization to various users and tasks.
- Designing Human-Robot Interaction. On the one hand the robot needs to correctly interpret the learning signals from the human. This involves detection of the signal and interpretation of that signal in the context of the learning process. On the other hand the human needs to understand the behavior of the robot. Robot's actions influence how humans behave as teacher during teaching. This leads to the need for transparency-based protocols.

III. Transparency in Interactive Robot Learning

To efficiently engage humans in sophisticated forms of teaching/learning interactions, robots should be endowed with the capacity to analyze, model and predict humans' non-verbal behaviors [20]. Computational models of the dynamics of social interaction will allow robots to be effective social partners. At the same time, it is expected by humans that robots will be able to perform tasks with a certain level of autonomy. To fill these requirements, there is a need to develop advanced models of human-robot interaction by exploiting explicit and implicit behaviors, such as pointing and showing interest for an object, that regulate the dynamics of both social interaction and task execution.

Dimensions such as alignment, rhythm, contingency, and feedback are also the focus of Interactive Robot Learning. In particular, in [21], it has been shown that robot learning by interaction (including by demonstration [22], [23]) should go beyond the assumption of unidirectional transfer of knowledge from human tutor to the robot learner by explicitly taking into account complex and rich phenomena of social interactions (e.g., mutual adaptation, nature and role of feedback). Understanding and modeling the dynamics of social interaction and learning is a major challenge of cognitive developmental robotics research [24], [25]. This trend is now explored by the research community. For example, to efficiently perform repetitive joint pick-and-place task, a controller able to explicitly exploit interpersonal synchrony has been designed using phase and event synchronization [26].

Usually, robot learning frameworks need to predefine a social interaction mechanism. For instance, a reinforcement based controller will require to script the interpretation of a head nod as a positive feedback. To achieve personalized interaction, we need to develop robots that learn compact social interaction and task models through repeated interactions and task executions. Although the focus of this article is not the learning of social interaction models, we summarize recent findings to highlight the importance of social interaction in human-robot interactive learning.

An effective way to handle the interplay between social interaction and task execution is to formulate it as a multi-task learning problem. This will require to simultaneously learn two models, one for the social interaction and one for the task execution. In [27], we have shown that task learning with a reinforcement learning framework is significantly improved by a social model (convergence with minimal human-robot interactions). In the context of imitation, we designed interpersonal models for improving social capabilities of a robot controller learning based on a Perception-Action (PerAc) architecture [28]. The models have been evaluated with two simultaneous tasks: (i) prediction of social traits (i.e., identity and pathology: typical vs. with autism spectrum disorders) and (ii) a traditional posture imitation learning task. This approach allows learning human identity from dynamics of interaction [29].

It is now clear that the analysis and modeling of interpersonal mechanisms is central to the design of robots capable of efficient collaborative task executions. However, the coordination and synchrony between observable behaviors and the roles of these phenomena during interpersonal human-robot interactions for social learning and task learning remain unclear. In this paper, we argue that there is a need to develop transparency-based learning protocols, where the human teacher has access to the current state of the robot's learning process in an intuitive way. We further argue that the expression of emotions simulated based on the temporal difference errors of the robot provides the basis for such an intuitive signal.

In [6], learning schemes allowing transparency have been introduced. Among the schemes, an agent was designed that gives feedback to the human users before performing an action. The feedback proposed is simple: gazing to objects relevant to the future action. This simple signal increases the transparency by reducing uncertainty and explicitly given (or not) the turn to the human teacher for providing guidance and/or feedback. However, most of the approaches in the literature dealing with transparency (i) deal with explicit signals (e.g. gazing), (ii) assume emitter-receiver based interaction and (iii) do not consider emotion.

In this paper, we will show how emotions can be employed to design transparency mechanisms for interactive robot learning frameworks that use Reinforcement Learning as learning method (figure 1).

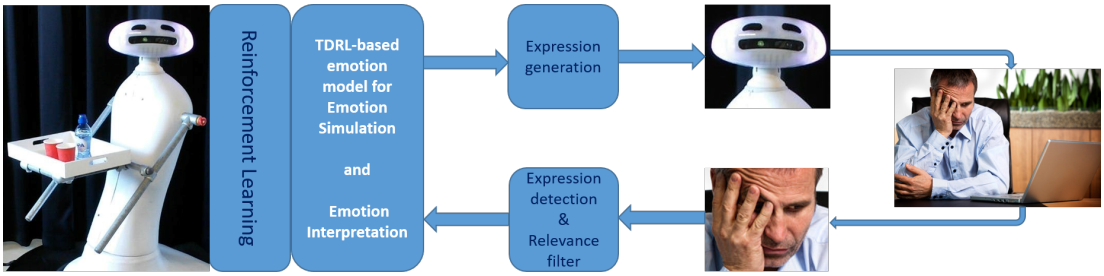

Fig. 1. Interactive robot learning: while a robot learns a new skill, emotions can be used in complex loops such as the expression of a robot's intentions and current states (i.e. to improve transparency), the perception of a human's emotional state, and the computation of a representation that is compatible with the reinforcement learning framework.

IV. Emotion

The vast majority of emotion theories propose that emotions arise from personally meaningful (imagined) changes in the current situation. An emotion occurs when a change is personally meaningful to the agent. In cognitive appraisal theories emotion is often defined as a valenced reaction resulting from the assessment of personal relevance of an event [30], [31], [32]. The assessment is based on what the agent believes to be true and what it aims to achieve as well as its perspective on what is desirable for others. In theories that emphasize biology, behavior, and evolutionary benefit [33], [34], the emotion is more directly related to action selection but the elicitation condition is similar: an assessment of harm versus benefit resulting in a serviceable habit aimed at adapting the behavior of the agent. Also in computational models that simulate emotions based on either cognitive appraisal theories [35], [36], [37], [38] or biological drives and internal motivation [39], [40] emotions always arise due to (internal) changes that are assessed as personally relevant.

We adopt the following definition of emotion: an emotion is a valenced reaction to (mental) events that modify future action, grounded in bodily and homeostatic sensations [11]. This sets emotion apart from mood, which is long term [41] and undirected, as well as attitude, which is a belief with associated affect rather than a valenced assessment involved in modifying future behavior.

In general, emotion plays a key role in shaping human behavior. On an interpersonal level, emotion has a communicative function: the expression of an emotion can be used by others as a social feedback signal as well as a means to empathize with the expresser [42]. On an intra-personal level emotion has a reflective function [43]: emotions shape behavior by providing feedback on past, current and future situations [44].

Emotions are essential in development, which is particularly relevant to our proposal. First, in infant-parent learning settings a childs expression of emotion is critical for an observers understanding of the state of the child in the context of the learning process [8]. Second, emotional expressions of parents are critical feedback signals to children providing them with an evaluative reference of what just happened [7], [10], [45]. Further, emotions are intimately tied to cognitive complexity. Observations from developmental psychology show that children start with undifferentiated distress and joy, growing up to be individuals with emotions including guilt, reproach, pride, and relief, all of which need significant cognitive abilities to be developed. In the first months of infancy, children exhibit a narrow range of emotions, consisting of distress and pleasure [46]. Joy and sadness emerge by 3 months, anger around 4 to 6 months with fear usually reported first at 7 or 8 months [46].

V. Emotion and Transparency in Robotics

As mentioned in III, transparency allows to engage humans in complex interactive scenarios. In most of the ongoing works, verbal and non-verbal signals are employed to develop transparency mechanisms [6]. In [47], the authors show that transparency reduces conflict and improves robustness of the interaction, resulting in better human-machine team performance. In [48], the authors review literature relating the complex relationship between the ideas of utility, transparency and trust. In particular, they discuss the potential effects of transparency on trust and utility depending on the application and purpose of the robot. These research questions are currently addressed to design new computational models of human-robot interaction. In [49], the authors identify nonverbal cues that signal untrustworthy behavior and also demonstrate the human mind's readiness to interpret those cues to assess the trustworthiness of a social robot. Transparency could be considered as an enabling mechanism for successfully fulfilling some ethical principles [50]. The interplay between transparency and ethical principles is of primordial importance in interactive machine learning frameworks, since the machines continuously collect implicit data from users.

Regarding the interplay between transparency and emotion, in [51] the authors proposed to expose users to direct physical interaction with a robot assistant in a safe environment. The aim of the experiment was to explore viable strategies a humanoid robot might employ to counteract the effect of unsatisfactory task performance. The authors compared three sorts of strategies: (i) non-communicative, most efficient, (ii) non-communicative, makes a mistake and attempts to rectify it, and (iii) communicative, expressive, also makes a mistake and attempts to rectify it. The results show that the expressive robot was preferred over a more efficient one. User satisfaction was also significantly increased in the communicative and expressive condition. Similar results have been obtained in a study aimed at investigation of how transparency and task type

influence trustfulness, perceived quality of work, stress level, and co- worker preference during human-autonomous system interaction [52]. Of particular interest, the author showed that a transparency mechanism, feedback about the work of roboworker, increases perceived quality of work and self-trust. These results are moderated significantly by individual differences owing to age and technology readiness.

Taken together, these results show the importance of transparent mechanisms through emotion for users. It challenges human-robot interaction since the communicative and expressive robots were preferred over the more efficient one. We conclude that there is a need to better investigate transparency mechanisms in Human-Robot Interaction. In this paper, we propose to address this challenge by combining machine learning and affective computing.

VI. Temporal Difference Reinforcement Learning

Reinforcement Learning (RL) [1], [2] is a well-established computational technique enabling agents to learn skills by trial and error. In a recent survey [3] a large variety of tasks have been show to be enabled by RL, such as walking, navigation, table tennis, and industrial arm control. The learning of a skill in RL is mainly determined by a feedback signal, called the reward, r. In contrast to the definition of reward in the psychological conditioning literature where a negative "reward" is referred to as punishment, reward in RL can be positive or negative. In RL cumulative reward is also referred to as $return$. Through trial and error, a robot or virtual agent adjusts its behavior to maximize the expected cumulative future reward, i.e., it attempts to learn the optimal policy.

There are several approaches to learning the optimal policy (way of selecting actions) with RL. First, we can separate value-function-based methods, which try to iteratively approximate the return, and policy search methods, which tries to directly optimize some parametrized policy. The first tries to learn a value function that matches expected return (cumulative future reward), and uses the value function to select actions. The typical example is Q-learning. The second tries to learn a policy directly by changing the probabilities of selecting particular actions in particular states based on the estimated return.

In this article we focus on value-function based methods. In this class of RL approaches we can further discriminate model-based and model-free. Model-based RL [53] refers to approaches where the environmental dynamics $T(s, a, s')$ and reward function $r(s, a, s')$ are learned or known. Here T refers to the transition function specifying how a state s' follows from an action a in a state s. This is usually a Markovian probability, so $T(s, a, s') = P(s'|s, a)$. Planning and optimization algorithms such as Monte Carlo methods and Dynamic Programming (DP) are used to calculate the value of states based on the Bellman equation (see [1]) directly. In model-based RL, we thus approximate the transition and reward function from the sampled experience. After acquiring knowledge of the environment, we can mix real sample experience with planning updates.

However, in many applications the environment's dynamics are hard to determine, or the model is simply too big. As an alternative, we can use sampling-based methods to *learn* the policy, known as model-free reinforcement learning. In model-free RL we iteratively approximate the value-function through temporal difference (TD) reinforcement learning (TDRL), thereby avoiding having to learn the transition function (which is usually challenging). Well-known algorithms are Q-learning [54], SARSA [55] and TD(λ) [56]. TDRL approaches share the following: at each value update, the value is changed using the difference between the current estimate of the value and a new estimate of the value. This new estimate is calculated as the current reward and the return of the next state. This difference signal is called the temporal difference error. It reflects the amount of change needed to the current estimate of the value of the state the agent is in. The update equation for Q-learning is given by:

$$Q(s,a)_{new} \leftarrow Q(s,a)_{old} + \alpha\Big[TD\Big] \tag{1}$$

$$TD = r + \gamma \max_{a'} Q(s',a') - Q(s,a)_{old} \tag{2}$$

where α specifies a learning rate, γ the discount factor and r the reward received when executing action a, and $Q(s, a)$ the action value of action a in state s. The TD error in this formulation of Q learning is equal to the update taking the best action into account. In the case of SARSA, where the update is based on the actual action taken, the TD Error would be:

$$TD = r + \gamma Q(s',a') - Q(s,a)_{old} \tag{3}$$

Note that although model-based RL methods typically do not explicitly define the TD error, it still exist and can be calculated by taking the difference between the current value and new value of the state, as follows:

$$TD = Q(s,a)_{new} - Q(s,a)_{old} \tag{4}$$

with $Q(s, a)_{new}$ calculated through the Bellman equation. This is important to keep in mind in the discussion on TDRL emotions, as TDRL-based emotion simulation is also possible for model-based RL.

The problem with RL is that the learning process requires both exploration and exploitation for the task to be learned in an efficient way. As the reward function and exploration process used in RL can be complex, it is generally hard to observe the robot and then understand the learning process, i.e, to understand what is happening to the Q function or policy of the robot and why it makes particular choices during the learning process. For example, exploration can result in very ineffective actions that are completely off-policy (not what the robot would typically do when exploiting the model). This is fine, as long as you know that the robot also knows that there is a better option. Ineffective actions are fine as long as they reflect exploration, not exploitation. In the latter case you need to correct the robot. It is hard for an observer to extract what is going on in the "mind" of a RL robot, because the difference between exploration on the one hand, and exploitation of a bad model on the other is not observable. RL lacks transparency.

VII. Computational models of Emotion in Reinforcement Learning Agents

In the last decades, emotions, in particular the emotions of joy, distress, anger, and fear and the dimensions of valence and arousal, have been used, modeled and studied using Reinforcement Learning in agents and robots (for a recent overview see [57]). Overall, human emotion expressions can be used as *input* for the RL model, either as state or as feedback (reward), emotions can be *elicited by* (i.e. simulated in) the RL model and used to influence that, which we will detail later, and, emotions can be expressed by the robot as social signal *output*. For example, human emotion expressions have been used as additional reward signals for a reinforcement learning agent [58], increasing the learning speed of the agent. Also, emotions have been simulated in adaptive agents based on homeostatic processes and used as internal reward signal or modification thereof [59], as well as modification of learning and action selection meta-parameters [60], [61]. In a similar attempt, cognitive appraisal modeling based on information available during the learning process has been used as additional reward signal [62]. Finally, emotional expressions of the robot have been used as communicative signal already in earlier work on robot learning, although these expressions were not coupled to reinforcement learning [63].

Most relevant to the transparency of a robot's learning process are the different ways in which emotions can be *elicited by* the RL model. This elicitation process defines what emotions occur over time, and is the basis for the emotions to be expressed as social signals. In general there are four ways [57] in which emotion elicitation can occur: 1) homeostasis and extrinsic motivation, 2) appraisal and intrinsic motivation, 3) reward and value function, 4) hard-wired connections from sensations. In homeostasis-based emotion elicitation, emotions result from underlying (biological) drives and needs that are (not) met [64], e.g., "hunger elicits distress". In appraisal-based emotion elicitation, emotions result from evaluation processes that assess the current state of the RL model [65], e.g., "unexpected state transitions elicit surprise". In reward and value-based emotion elicitation, emotions are derived from (changes in) the reward and values of the visited states, e.g., "average reward over time equals valence" [66]. In hard-wired emotion elicitation, emotions are elicited by properties of perceived states, e.g, "bumping into a wall elicits frustration" [67].

In section IV we have shown that in nature emotion plays a key role in development. For a learning robot to be able to express the right emotion at the right intensity at the right time, the emotion elicitation process needs to be grounded in the robot's learning process. RL is a suitable and popular method for robot task learning. If we want people to understand a learning robot's emotional expressions to be used to make that learning process transparent to the user, this means that there is need for *a generic computational approach able to elicit emotions grounded in the RL learning process*. Out of the four emotion elicitation methods listed above, only reward and value based emotion elicitation can be simulated using general RL approaches by which we mean that it does not need additional assumptions about processes that either underly the RL reward signal (such as homeostasis) or are external to the RL model (such as appraisal or hard-wired perception-emotion associations). The emotion can be simulated with the basic RL constructs such as reward, value, and temporal difference. In other words, we would like to bring the computational approach to simulate emotions as closely to the RL method as possible.

We conclude this section with the following requirement for the emotion elicitation process in learning robots: emotion elicitation must be grounded in the RL learning model and based on (derivations of the) reward or value function.

VIII. Transparent RL with TD-RL based Emotions

Emotion is essential in development, as discussed previously. Reinforcement learning is a powerful, but non-transparent model for robot and agent learning, as also discussed. Computational models of emotion based on (derivations of the) value and reward function are most promising for modeling RL-based emotions, when it comes to grounding the emotion in the learning process of the agent. The challenge to be addressed here is therefore *how to develop a computational model that enables RL-grounded emotion elicitation that can be used to enhance the transparency of the learning process*. Two aspects are of major importance here: (1) it should enable the robot to express emotions as well as (2) interpret human emotions, both in the context of the learning process of the robot. Here we argue that a recent theory of emotion, called TDRL Emotion Theory, is a suitable candidate for such a computational model of emotion.

The core of the TDRL Theory of Emotion is that all emotions are manifestations of temporal difference errors [11]. This idea is building on initial ideas by Brown and Wagner [68] and Redish [69], [70], and extending ideas of Baumeister [44], Rolls [71] and the work on intrinsic motivation [72]. Evidence for this view is found along three lines. First, the elicitation of emotion and the TD error is similar: emotions as well as the temporal difference error are feedback signals resulting from the evaluation of a particular current state or (imagined/remembered) event. Second, the functional effect of emotion and the TD error is similar: emotion and the temporal difference error impact future behavior by influencing action motivation. The evolutionary purpose of emotion and the TD error is similar: both emotion and the TD error aim at long-term survival of the agent and the optimization of well-being. An elaborate discussion of the support for and ramifications of TDRL Emotion Theory is discussed in [11], for example showing that TDRL Emotion Theory is consistent with important cognitive theories of emotion [73], [32], [31]. In this article we focus on what it proposes and how this is relevant for transparency of the robot's learning process.

It is very important to highlight that what the TDRL Theory of emotion proposes is that emotions are manifestations of TD errors. From a computational perspective this means that TDRL *is* the computational model of emotion. So it is not the case that RL is used to learn or construct a model. Another way to put it, is that TDRL Emotion Theory proposes to redefine

emotion as TD error assessment, with different emotions being manifestations of different types of TD error assessments.

We now summarize how the TDRL Theory of emotion defines joy, distress, hope and fear. In TDRL Emotion Theory these labels are used to refer to specific underlying temporal difference assessment processes and should not be taken literally or as a complete reference to all possible emotions that might exist in the categories that these labels typically refer to in psychology. We will use TDRL-based joy, distress, hope and fear throughout the article as running examples of how a robot could make transparent the current state of the learning process.

Joy (distress) is the manifestation of a positive (negative) temporal difference error. Based on an assessment of the values of different actions, an agent selects and executes one of those actions and arrives at a new situation. If this situation is such that the action deserves to be selected more (less) often - either because the reward, r, was higher (lower) than expected and/or the value estimate of the resulting situation, $Q(s', a')$, is better (worse) than expected resulting in a positive (negative) temporal difference signal - the action value is updated with a TD error,e.g. using equation (1). TDRL Emotion Theory proposes that this update manifests itself as joy (distress). Joy and distress are therefore emotions that refer to the now, to actual situations and events. For Q-learning this would mean that Joy and Distress are defined as follows:

$$if(TD > 0) \Rightarrow Joy = TD \tag{5}$$

$$if(TD < 0) \Rightarrow Distress = TD \tag{6}$$

With the TD error defined in the standard way for Q-learning:

$$TD = r + \gamma \max_{a'} Q(s', a') - Q(s, a)_{old} \tag{7}$$

We discuss the link between the psychology of emotion and TDRL in detail in [11], and we show joy and distress are simulated in RL agents in [74]. To give some insight into why the TD error models joy and distress consider the following. Joy and distress typically habituate over time, while the forming of a behavioral habit is taking place. The typical TD Error also "habituates" over time while the agent is learning the task. Consider a standard Q-learning agent that learns to navigate a 10x10 discrete gridworld maze with $\gamma = 0.7$ and terminal state $r = 5$. In Figure 2 the typical gradual decline of the TD error over the course of learning the task is plotted. The TD error declines because over time all Q values converge to the actual Bellman equation for a Max_a policy. TD errors therefore become smaller and more rare over time. The TDRL emotion interpretation is that the TD error simulates the joy experienced by this agent. The agent gets used to repeated rewarded encounters, resulting in habituation and thus in less joy.

The joy/distress signal - or in RL terms the temporal difference error - is also the building block for hope/fear. Hope (fear) refers to the anticipation of a positive (negative) temporal difference error. To explain fear and hope in the TDRL Theory of Emotion, the agent needs to have a mental model of the agent-environment interactions that is able to represent uncertainty and is used by the agent to anticipate.

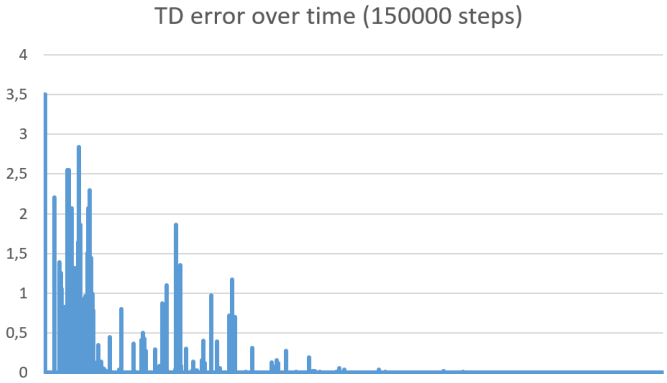

Fig. 2. Typical TD Error plot over time for a Q-learning agent learning a 10x10 discrete gridworld maze with $\gamma = 0.7$ and terminal state $r = 5$. TD error on the Y-axis, and steps on the X-axis. This effect on the TD error of learning to solve a static problem is general and not specific to this particular gridworld.

In RL this is referred to as model-based RL. In this case, the agent not only learns action value estimates (Model-free RL, such as SARSA) but also learns the probabilities associated with action-environment transitions $P(s'|s, a)$, i.e., what next state s' to expect as a result of action a in state s. Fear and hope emotions result from the agents mental simulation of such transitions to potential future states. If the agent simulates transitions to next possible next states and at some point a positive (negative) temporal difference error is triggered, then that agent knows that for this particular future there is also a positive (negative) adjustment needed for the current state/action. This process can be simulated using for example Monte Carlo Tree Search procedures such as UCT [75]. As this positive adjustment refers to a potential future transition, it doesn't feel exactly like joy (distress). It is similar, but not referring to the actual action that has been selected. The point is that fear (hope) shares a basic ingredient with joy (distress), i.e., the temporal difference error assessment. There is a qualitative difference though, which has to do with the cognitive processing involved in generating the signal: while joy and distress are about the now, hope and fear are about anticipated futures.

The TDRL view on emotion thus proposes a way to naturally ground the emotion elicitation process in the learning process of the agent or robot. Joy, distress, hope and fear can be computationally simulated in the proposed manner [75]. Such simulations have shown to be in line with natural affective phenomenon including habituation, fear extinction and effects of the agent's policy on fear intensity [75], [76], [74].

This opens up the next important step towards making the learning process more transparent to the user using emotional expressions of the robot or agent: expressing the elicited emotions in a plausible way. Artificial expression of distress, joy, fear and hope is relatively well-studied and easy to express using, e.g., the Facial Action Coding System proposed by Ekman [77]. Distress can be expressed as sadness, joy as joy, fear as fear, and hope as positive surprise. As the TDRL view

explicitly excludes "neutral" surprise as an emotion [11], there is no potential confusion between surprise and fear, which is relatively common in robot, agent and human emotion expression recognition. Hope and fear in the TDRL view refer to positive anticipation and negative anticipation and can therefore be distinguished, e.g. on the face, by the presence of surprise/arousal-like feature such as open eyes and mouth, while distinguishing between positive and negative using the mouth curvature (corner pullers).

Expressing joy, distress, hope and fear enables the robot to communicate its learning process to the user in a continuous and grounded manner. The expression is a social signal towards the user showing whether or not the robot is getting better or worse (joy / distress), and whether or not the robot is predicting good or bad things to happen in the future (hope / fear). This gives the user important information to assess the difference between how the robot thinks it is doing (expression of joy and distress) compared to how the user thinks the robot is doing (based on the behavior of the robot). Second, it enables the user to assess whether or not to interfere with the learning process either though feedback or through guidance [6].

In a in a recent paper [75], in which we report on a small scale simulation study, we showed that hope and fear emerge from anticipation in model-based RL. We trained an agent to learn the optimal policy walking on a slippery slope along a cliff. We simulated hope and fear by simulating forward traces using UCT planning (a Monte Carlo tree search technique). We were able to show that more simulation resulted in more fear, just like closeness of the threat (agent being close to the cliff) as well as more random mental action selection policies. In earlier work we already showed that TD error-based emotions produce plausible joy and distress intensity dynamics [74].

A more futuristic case highlighting how TDRL emotions can play a role in human-robot interaction is the following. You bought a new humanoid service robot and just unpacked it. It is moving around in your house for the first time, learning where what is and learning how to perform a simple fetch an carry task. You observe the robot as it seems to randomly move around your bedroom. You interpret this as it being lost, and actively guide the robot back to your living room. The robot goes back to the bedroom and keeps moving around randomly. You call the dealer and the dealer explains to you that it is exploring and that this is normal. However you still do not know how long it will go on and whether or not it is exploring the next time it wanders around your bedroom. Perhaps it is really lost the next time?.

If the robot is equipped with TDRL Emotions the scenario looks very different. You bought a new TDRL Emotion-enabled humanoid service robot and just unpacked it. It is moving around in your house for the first time, learning where what is and learning how to perform a simple fetch an carry task. You see the robot expressing joy while it seems to randomly move around your bedroom. You interpret this as it learning novel useful things. When you try to actively guide the robot back to your living room, it starts to first express distress and when you keep on doing that it expresses fear. You let go of the robot and the robot goes back to the bedroom and keeps moving around randomly looking happy. You decide that the robot is not ready exploring and that this is normal. When you ask the robot to fetch a drink, it starts to express hope and moves out of the bedroom. You decide that it was never lost in the bedroom.

Now let's do this a last time from the RL perspective of the robot. I just got turned on in a new environment. I do not know how the dynamics of the environment, so I enter curiosity-driven learning mode [78] to learn to navigate the environment. This launches curiosity-driven intrinsic motivation to target exploration. I have already reduced uncertainty about this big space I am in (for humans: the living), so I move to that small space over there. I am assessing positive TD errors while reducing the uncertainty in this small room (for humans: the bedroom). I express this positive TD error as joy. I notice that at some point my user user influences the execution of my actions such that the expected reduction in uncertainty is less than expected, this generates negative TD errors, which I express as distress. The user continues to push me away from areas of uncertainty reduction (highly intrinsically motivating areas), which generates negative TD error predictions, which I express as fear. The users lets go of me. I immediately move back to the intrinsically motivating area and express joy. The user asks me for a drink. I switch to task-mode and predict that with about 50 percent chance I can achieve a very high reward. This generates a predicted positive TD error, which I express as hope. I move out of the bedroom.

A second way in which TDRL-based emotions help in making the interactive learning process more transparent is that it can be used to interpret the user's emotional expressions in a way that relates that emotion to the learning process. Consider the following human-in-the-loop robot learning scenario (Figure 1). A robot learns a new service task using reinforcement learning as its controller. The computational model of emotion production (emotion elicitation) decides how the current state of the learning process should be mapped to an emotion. The emotion is expressed by the robot. The robot owner interprets the signal, and reacts emotionally to the robot. The reaction is detected and filtered for relevance to the task (e.g., the owners frown could also be due to reading a difficult email). The computational model of emotion interpretation maps the owners reaction to a learning signal usable by the reinforcement learning controller. In this scenario it becomes clear that the same computational model based on TDRL for emotion expression can be guiding in the interpretation of the user's emotional expression. To make this concrete, consider the following example. The expression of anger or frustration by a user directed at the robot should be interpreted as a negative TD error to be incorporated in the current behavior of the robot, as anger in this view is a social signal aimed at changing the behavior of the other [11]. The expression of encouragement should be incorporated as an anticipated joy signal, i.e., expressing encouragement should be interpreted as "go on in the line you are thinking of". It is a form of externally communicated hope.

IX. Discussion

We have argued that emotions are important social signals to take into account when robots need to learn novel - and adapt existing - behavior in the vicinity of - or in interaction with - humans. Reinforcement learning is a promising approach for robot task learning, however it lacks transparency. Often it is not clear from the robot's behavior how the learning process is doing, especially when exploring novel behavior. We proposed to make use of a TDRL-based emotion model to simulate appropriate emotions during the robot's learning process for expression and indicated that the same model could be a basis for interpreting human emotions in a natural way during the robot's learning process. However, there are several important challenges that remain open.

First, it is unclear if the labels used in TDRL Emotion Theory to denote particular manifestations of TD error processing can be used for expression when used *in interaction with a robot for the transparency of the learning process*. The proposed labels might not be the most appropriate ones and hence the emotional expression that follows might not be the most appropriate one either. For example consider the model for joy in the TDRL emotion view. Relations between specific emotions and RL-related signals seem to exist, e.g., the relation between joy and the temporal difference signal in RL. The temporal difference error is correlated with dopamine signaling in the brain [79] on the one hand, and a correlation between dopamine signaling and euphoria exists on the other [80]. Joy reactions habituate upon repeated exposure to jokes [81]. The temporal difference signal for a particular situation habituates upon repeated exposure [2]. Joy and the temporal difference signal are modulated by expectation of the reward. However, does that mean that *expression of joy* is the most appropriate social signal to express when the learning process is going well? The case of distress is even more interesting. Continued negative TD errors would manifest itself as distress. However, for the purpose of transparency of robot learning continued distress is perhaps better expressed as frustration.

Second, it is not clear if emotions should be expressed with intensity and dynamics as simulated, or, if there are thresholds or other mapping functions involved that modulate expression of the simulated emotion. The TDRL Emotion Theory proposes a straightforward start for joy, distress hope and fear directly based on the TD error assessment, but robot expression and "robot feeling" are perhaps different things. For example in earlier work of one of the authors (JB, unpublished but available upon request) the TD error was first normalized and then expressed as valence on an iconic face, so that reward magnitude did not influence the expression intensity. Another issue is related to timing. Expressions of emotions involve some delay with respect to the event and have an onset, hold and decay. How to computationally map the TD error to a - for human observers perceived as natural - expression is an open question. There are many challenges related to the dynamics of emotion-related expressions of TD error assessments. Novel computational models and subsequent human-robot interaction studies are needed to address how expressed affective signals relate to "felt" learning-related emotions by the robot.

Third, it is unclear how, and if, simulated robot emotion and human expressed emotion should influence the RL learning process. The function of simulated emotion has been linked to five major parts in the RL loop [57]: 1) reward modification, such as providing additional rewards, 2) state modification, such as providing additional input to the state used by the RL method, 3) meta-learning, such as changing learning parameters gamma or alpha, 4) action selection, such as changing the greediness of the selection, and, 5) epiphenomenon, the emotion is only expressed. The issue with respect to transparency of the learning process is that a user might expect some form of mixed influence on the robot's process when it expresses the emotion. So, even if the emotion that is "felt" by the robot is correctly simulated by a TDRL model, the user might expect something functional to happen when (a) the robot expresses an emotion, and, (b) the user expresses an emotion. For example, if the robot expresses fear, the user might expect the robot to also incorporate that in its action selection, such as proposed in [11]. Fear in nature is a manifestation of anticipated distress but at the same time a manifestation of a change in behavior to steer away from that distress. If a robot simulates fear, the user probably assumes it is also going to do something about it. Similarly, if the user expresses anger, he or she might expect the robot to respond with guilt to show that the communicated TD error has been effective used. If this signal is not given by the robot, then that might limit transparency. On top of that, expression of emotion by a user towards a robot is probably depending on that user's personal preferences. As such, the interpretation of a user's emotional expression might need to be personalized.

Fourth, it is unclear what kind of human-agent interaction benefits can be expected and in what tasks these benefits should be observed. Usually research that investigates the role of emotion in RL-based adaptive agents focuses on increasing adaptive potential. However, the challenge for transparency of the learning process is to create benchmark scenarios in which particular interaction benefits can be studied. One can think about the naturalness of the learning process, willingness of the user to engage in long-term HRI, perceived controllability of the robot, perceived adaptiveness of the robot, etc.. These scenarios most likely are different from scenarios aimed at investigating the adaptive potential of emotion.

Fifth, deploying robots around humans in our society is a difficult enterprise that depends not only on emotion-enabled transparency of the robot's learning process. There are many challenges that are important and urgent for such deployment. For example, robustness of the interaction and human-robot dialog, understanding of context, personalization of behavior, learning from sparse samples and rewards, properly designed HRI, management of expectations of the user, hardware robustness and battery life, the development of an ethical framework for humanoid robots, and price, all play a major role in the practical usefulness and acceptance of robots around us. In this article we argue that emotions are important for transparency of the robot's learning process and that such emotions can be simulated using TDRL. Future work in this direction should confirm or disconfirm whether this idea contributes to the transparency of, and acceptance of robots around us.

X. Outlook

As robots seem to be here to stay, and pre-programming all behavior is unrealistic, robots need to learn in our vicinity and in interaction with us. As emotions play an important role in learning-related interaction, such robots need emotions to express their internal state. However, for this to be transparent, robots should do this in a consistent manner. In other words, the consistency of simulated emotions between different robots for real-world learning robotics is important for their transparency. With regards to the simulation and expression of emotions in learning robots and agents, it would be a great long term challenge to converge to a standard framework for the simulation and expression of emotions based on RL-like learning processes. Such a standard framework, in the same spirit as cognitive appraisal frameworks such as OCC [31], can help to generate and express emotions for learning robots in a unified way. This will help the transparency of the learning process. We proposed TDRL Emotion Theory as a way to do this, however, as a community it is essential to critically assess different models of emotion grounded in RL. In a recent survey [57] we address the different approaches towards emotion simulation in RL agents and robots. One of the key findings is the large diversity of emotion models, but several key challenges remain including the lack of integration and replication of results of others, and, lack of a common method for critically examining such models.

On the other hand, humans have individual ways of giving learning-related feedback to other humans, and this most likely is the case with feedback to robots as well. For example, some people express frustration when a learner repeatedly fails to achieve a task, while other's express encouragement. Regardless of what is the best motivational strategy for human learners, a learning robot needs to be able to personalize their emotion interpretation towards an individual user. So, while on the one hand we need a standardized computational approach to express robot emotions during the learning process, on the other hand we need a personalized computational model to interpret emotions from individual users. The latter can be bootstrapped by taking TDRL-based emotions, as argued in the section on TDRL emotions for transparency, but such as model needs to adapt to better fit individual users. Investigating the extend to which personalization of human feedback interpretation is needed is therefore an important challenge.

XI. Conclusion

We conclude that emotions are important social signals to consider when aiming for transparency of the reinforcement learning process of robots and agents. We have highlighted current challenges in interactive robot learning. We have shown that the TDRL Theory of emotion provides sufficient structure to simulate emotions based on TD error signals. This simulation grounds emotion elicitation in the learning process of the agent and provides a start to also interpret the emotions expressed by the user in the context of learning. We argued that this is a significant step towards making the learning process more transparent. We have highlighted important challenges to address, especially in light of the functional effects of emotion *in interaction* between between robots and people.

Acknowledgment

This work partly received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 765955 (ANIMATAS Innovative Training Network).

References

[1] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press Cambridge, 1998.

[2] G. Tesauro, "Temporal difference learning and td-gammon," *Communications of the ACM*, vol. 38, no. 3, pp. 58–68, 1995.

[3] J. Kober, J. A. Bagnell, and J. Peters, "Reinforcement learning in robotics: A survey," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1238–1274, 2013.

[4] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," *arXiv preprint arXiv:1312.5602*, 2013.

[5] W. B. Knox, P. Stone, and C. Breazeal, *Training a Robot via Human Feedback: A Case Study*, ser. Lecture Notes in Computer Science. Springer International Publishing, 2013, vol. 8239, book section 46, pp. 460–470.

[6] A. L. Thomaz and C. Breazeal, "Teachable characters: User studies, design principles, and learning performance," in *Intelligent Virtual Agents*. Springer, 2006, pp. 395–406.

[7] S. Chong, J. F. Werker, J. A. Russell, and J. M. Carroll, "Three facial expressions mothers direct to their infants," *Infant and Child Development*, vol. 12, no. 3, pp. 211–232, 2003.

[8] K. A. Buss and E. J. Kiel, "Comparison of sadness, anger, and fear facial expressions when toddlers look at their mothers," *Child Development*, vol. 75, no. 6, pp. 1761–1773, 2004.

[9] C. Trevarthen, "Facial expressions of emotion in mother-infant interaction," *Human neurobiology*, vol. 4, no. 1, pp. 21–32, 1984.

[10] M. D. Klinnert, "The regulation of infant behavior by maternal facial expression," *Infant Behavior and Development*, vol. 7, no. 4, pp. 447–465, 1984. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0163638384800053

[11] J. Broekens, "A temporal difference reinforcement learning theory of emotion: A unified view on emotion, cognition and adaptive behavior," *Emotion Review*, submitted.

[12] J. Grizou, M. Lopes, and P. Y. Oudeyer, "Robot learning simultaneously a task and how to interpret human instructions," in *2013 IEEE Third Joint International Conference on Development and Learning and Epigenetic Robotics (ICDL)*, Aug 2013, pp. 1–8.

[13] V. Palologue, J. Martin, A. K. Pandey, A. Coninx, and M. Chetouani, "Semantic-based interaction for teaching robot behavior compositions," in *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Aug 2017, pp. 50–55.

[14] S. Griffith, K. Subramanian, J. Scholz, C. L. Isbell, and A. L. Thomaz, "Policy shaping: Integrating human feedback with reinforcement learning," in *Advances in Neural Information Processing Systems 26*, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2013, pp. 2625–2633. [Online]. Available: http://papers.nips.cc/paper/5187-policy-shaping-integrating-human-feedback-with-reinforcement-learning.pdf

[15] B. D. Argall, S. Chernova, M. Veloso, and B. Browning, "A survey of robot learning from demonstration," *Robotics and autonomous systems*, vol. 57, no. 5, pp. 469–483, 2009.

[16] H. B. Suay and S. Chernova, "Effect of human guidance and state space size on interactive reinforcement learning," in *2011 RO-MAN*, July 2011, pp. 1–6.

[17] A. Najar, O. Sigaud, and M. Chetouani, "Training a robot with evaluative feedback and unlabeled guidance signals," in *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Aug 2016, pp. 261–266.

[18] S. Boucenna, S. Anzalone, E. Tilmont, D. Cohen, and M. Chetouani, "Learning of social signatures through imitation game between a robot and a human partner." *IEEE Transactions on Autonomous Mental Development*, 2014.

[19] J. MacGlashan, M. K. Ho, R. Loftin, B. Peng, G. Wang, D. L. Roberts, M. E. Taylor, and M. L. Littman, "Interactive learning from policy-dependent human feedback," in *Proceedings of the 34th International Conference on Machine Learning*, vol. 70. PMLR, 06–11 Aug 2017, pp. 2285–2294.

[20] K. Dautenhahn, "Socially intelligent robots: dimensions of human–robot interaction," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 362, no. 1480, pp. 679–704, 04 2007.

[21] A. Vollmer, M. Muhlig, J. J. Steil, K. Pitsch, J. Fritsch, K. J. Rohlfing, and B. Wrede, "Robots show us how to teach them: Feedback from robots shapes tutoring behavior during action learning," *PLoS ONE*, vol. 9, no. 3, p. e91349, 03 2014.

[22] B. Argall, S. Chernova, M. Veloso, and B. Browning, "A survey of robot learning from demonstration," *Robotics and Autonomous Systems*, vol. 67, pp. 469–483, 2009.

[23] A. Billard, S. Callinon, R. Dillmann, and S. Schaal, "Robot programming by demonstration," in *Robot programming by demonstration*, B. Siciliano and O. E. Khatib, Eds. Springer, New York, 2008, ch. 59.

[24] A. Cangelosi, G. Metta, G. Sagerer, S. Nolfi, C. Nehaniv, K. Fischer, J. Tani, T. Belpaeme, G. Sandini, F. Nori, L. Fadiga, B. Wrede, K. J. Rohlfing, E. Tuci, K. Dautenhahn, J. Saunders, and A. Zeschel, "Integration of action and language knowledge: A roadmap for developmental robotics," *IEEE Transactions on Autonomous Mental Development*, pp. 167–195, 2010.

[25] A. Sciutti, A. Bisio, F. Nori, G. Metta, L. Fadiga, T. Pozzo, and G. Sandini, "Measuring human-robot interaction through motor resonance," *International Journal of Social Robotics*, vol. 4, no. 3, pp. 223–234, 2012.

[26] A. Mortl, T. Lorenz, and S. Hirche, "Rhythm patterns interaction - synchronization behavior for human-robot joint action," *PlosOne*, vol. 9, no. 4, p. e95195, 2014.

[27] A. Najar, O. Sigaud, and M. Chetouani, "Social-task learning for hri," vol. 9388, pp. 472–481, 2015.

[28] P. Gaussier, S. Moga, M. Quoy, and J. P. Banquet, "From perception-action loops to imitation processes: A bottom-up approach of learning by imitation," *Applied Artificial Intelligence*, vol. 12, no. 7-8, pp. 701–727, 10 1998.

[29] S. Boucenna, C. D., P. Gaussier, A. N. Meltzoff, and M. Chetouani, "Robots learn to recognize individuals from imitative encounters with people and avatars," *Scientific Reports (Nature Publishing Group)*, vol. srep19908, 2016.

[30] A. Moors, P. C. Ellsworth, K. R. Scherer, and N. H. Frijda, "Appraisal theories of emotion: State of the art and future development," *Emotion Review*, vol. 5, no. 2, pp. 119–124, 2013.

[31] A. Ortony, G. L. Clore, and A. Collins, *The Cognitive Structure of Emotions*. Cambridge University Press, 1988.

[32] K. R. Scherer, A. Schorr, and T. Johnstone, *Appraisal processes in emotion: Theory, methods, research*. Oxford University Press, 2001.

[33] N. H. Frijda, "Emotions and action," in *Feelings and Emotions: the amsterdam symposium*, A. S. R. Manstead and N. H. Frijda, Eds. Cambridge University Press, 2004, p. 158173.

[34] J. Panksepp, *Affective Neuroscience: the foundations of human and animal emotions*. Oxford University Press, 1998.

[35] J. Dias and A. Paiva, *Feeling and reasoning: A computational model for emotional characters*. Springer, 2005, pp. 127–140.

[36] S. C. Marsella and J. Gratch, "EMA: A process model of appraisal dynamics," *Cognitive Systems Research*, vol. 10, no. 1, pp. 70–90, 2009.

[37] A. Popescu, J. Broekens, and M. v. Someren, "Gamygdala: An emotion engine for games," *IEEE Transactions on Affective Computing*, vol. 5, no. 1, pp. 32–44, 2014.

[38] B. R. Steunebrink, M. Dastani, and J.-J. C. Meyer, *A Formal Model of Emotions: Integrating Qualitative and Quatitative Aspects*. IOS Press, 2008, pp. 256–260.

[39] D. Cañamero, "Designing emotions for activity selection in autonomous agents," *Emotions in humans and artifacts*, vol. 115, p. 148, 2003.

[40] I. Cos, L. Cañamero, G. M. Hayes, and A. Gillies, "Hedonic value: enhancing adaptation for motivated agents," *Adaptive Behavior*, p. 1059712313486817, 2013.

[41] C. Beedie, P. Terry, and A. Lane, "Distinctions between emotion and mood," *Cognition and Emotion*, vol. 19, no. 6, pp. 847–878, 2005. [Online]. Available: https://doi.org/10.1080/02699930541000057

[42] A. H. Fischer and A. Manstead, *Social Functions of Emotion*. Guilford Press, 2008, pp. 456–468.

[43] K. Oatley, "Two movements in emotions: Communication and reflection," *Emotion Review*, vol. 2, no. 1, pp. 29–35, 2010. [Online]. Available: http://emr.sagepub.com/content/2/1/29.abstract

[44] R. F. Baumeister, K. D. Vohs, C. N. DeWall, and L. Zhang, "How emotion shapes behavior: Feedback, anticipation, and reflection, rather than direct causation," *Personality and Social Psychology Review*, vol. 11, no. 2, pp. 167–203, 2007.

[45] C. Saint-georges, M. Chetouani, R. Cassel, F. Apicella, A. Mahdhaoui, F. Muratori, M. Laznik, and D. Cohen, "Motherese in interaction: at the cross-road of emotion and cognition? (a systematic review)," *PLoS ONE*, vol. 8, no. 10, p. e78103, 2013.

[46] L. A. Sroufe, *Emotional development: The organization of emotional life in the early years*. Cambridge University Press, 1997.

[47] C. Breazeal, C. D. Kidd, A. L. Thomaz, G. Hoffman, and M. Berlin, "Effects of nonverbal communication on efficiency and robustness in human-robot teamwork," in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Aug 2005, pp. 708–713.

[48] R. H. Wortham and A. Theodorou, "Robot transparency, trust and utility," *Connection Science*, vol. 29, no. 3, pp. 242–248, 2017.

[49] J. J. Lee, B. Knox, J. Baumann, C. Breazeal, and D. DeSteno, "Computationally modeling interpersonal trust," *Frontiers in Psychology*, vol. 4, 2013.

[50] A. Spagnolli, L. E. Frank, P. Haselager, and D. Kirsh, "Transparency as an ethical safeguard," in *Symbiotic Interaction*, J. Ham, A. Spagnolli, B. Blankertz, L. Gamberini, and G. Jacucci, Eds. Springer International Publishing, 2018, pp. 1–6.

[51] A. Hamacher, N. Bianchi-Berthouze, A. G. Pipe, and K. Eder, "Believing in bert: Using expressive communication to enhance trust and counteract operational error in physical human-robot interaction," in *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Aug 2016, pp. 493–500.

[52] K. Kallinen, "The effects of transparency and task type on trust, stress, quality of work, and co-worker preference during human-autonomous system collaborative work," in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '17. New York, NY, USA: ACM, 2017, pp. 153–154.

[53] T. Hester and P. Stone, "Learning and using models," in *Reinforcement Learning*. Springer, 2012, pp. 111–141.

[54] C. J. C. H. Watkins, "Learning from delayed rewards," Ph.D. dissertation, University of Cambridge England, 1989.

[55] G. A. Rummery and M. Niranjan, "On-line Q-learning using connectionist systems," University of Cambridge, Department of Engineering, Tech. Rep., 1994.

[56] R. S. Sutton, "Learning to predict by the methods of temporal differences," *Machine learning*, vol. 3, no. 1, pp. 9–44, 1988.

[57] T. Moerland, J. Broekens, and C. M. Jonker, "Emotion in reinforcement learning agents and robots: A survey." *Machine Learning*, vol. 107, no. 2, p. 443480, 2018.

[58] J. Broekens, "Emotion and reinforcement: affective facial expressions facilitate robot learning," in *Artifical Intelligence for Human Computing*. Springer, 2007, pp. 113–132.

[59] S. C. Gadanho, "Learning behavior-selection by emotions and cognition in a multi-goal robot task," *The journal of machine learning research*, vol. 4, pp. 385–412, 2003.

[60] E. Hogewoning, J. Broekens, J. Eggermont, and E. G. Bovenkamp, "Strategies for affect-controlled action-selection in Soar-RL," in *Nature Inspired Problem-Solving Methods in Knowledge Engineering*. Springer, 2007, pp. 501–510.

[61] A. J. Blanchard and L. Canamero, "From imprinting to adaptation: Building a history of affective interaction," in *Proceedings of the 5th International Workshop on Epigenetic Robotics*. Lund University Cognitive Studies, 2005, pp. 23–30.

[62] P. Sequeira, F. S. Melo, and A. Paiva, "Learning by appraising: an emotion-based approach to intrinsic reward design," *Adaptive Behavior*, p. 1059712314543837, 2014.

[63] C. Breazeal, "Emotion and sociable humanoid robots," *International Journal of Human-Computer Studies*, vol. 59, no. 1, pp. 119–155, 2003.

[64] D. Cañamero, "A hormonal model of emotions for behavior control," *VUB AI-Lab Memo*, vol. 2006, 1997.

[65] P. Sequeira, F. S. Melo, and A. Paiva, "Emotion-based intrinsic motivation for reinforcement learning agents," in *Affective computing and intelligent interaction*. Springer, 2011, pp. 326–336.

[66] J. Broekens, W. A. Kosters, and F. J. Verbeek, "On affect and self-adaptation: Potential benefits of valence-controlled action-selection," in *Bio-inspired modeling of cognitive tasks*. Springer, 2007, pp. 357–366.

[67] D. D. Tsankova, "Emotionally influenced coordination of behaviors for autonomous mobile robots," in *Intelligent Systems, 2002. Proceedings. 2002 First International IEEE Symposium*, vol. 1. IEEE, 2002, pp. 92–97.

[68] R. T. Brown and A. R. Wagner, "Resistance to punishment and extinction following training with shock or nonreinforcement," *Journal of Experimental Psychology*, vol. 68, no. 5, pp. 503–507, 1964.

[69] A. D. Redish, "Addiction as a computational process gone awry," *Science*, vol. 306, no. 5703, pp. 1944–1947, 2004. [Online]. Available: http://www.sciencemag.org/content/306/5703/1944.abstract
[70] A. D. Redish, S. Jensen, A. Johnson, and Z. Kurth-Nelson, "Reconciling reinforcement learning models with behavioral extinction and renewal: Implications for addiction, relapse, and problem gambling," *Psychological Review*, vol. 114, no. 3, pp. 784–805, 2007.
[71] E. T. Rolls, "Precis of the brain and emotion," *Behavioral and Brain Sciences*, vol. 20, pp. 177–234, 2000.
[72] S. Singh, R. L. Lewis, A. G. Barto, and J. Sorg, "Intrinsically motivated reinforcement learning: An evolutionary perspective," *Autonomous Mental Development, IEEE Transactions on*, vol. 2, no. 2, pp. 70–82, 2010.
[73] R. Reisenzein, "Emotional experience in the computational belief–desire theory of emotion," *Emotion Review*, vol. 1, no. 3, pp. 214–222, 2009.
[74] J. Broekens, E. Jacobs, and C. M. Jonker, "A reinforcement learning model of joy, distress, hope and fear," *Connection Science*, pp. 1–19, 2015. [Online]. Available: http://dx.doi.org/10.1080/09540091.2015.1031081
[75] T. M. Moerland, J. Broekens, and C. M. Jonker, "Fear and Hope Emerge from Anticipation in Model-based Reinforcement Learning," in *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2016, pp. 848–854.
[76] E. Jacobs, J. Broekens, and C. M. Jonker, "Emergent dynamics of joy, distress, hope and fear in reinforcement learning agents," in *Adaptive Learning Agents workshop at AAMAS2014*, 2014.
[77] P. Ekman, W. V. Friesen, M. O'Sullivan, A. Chan, I. Diacoyanni-Tarlatzis, K. Heider, R. Krause, W. A. LeCompte, T. Pitcairn, P. E. Ricci-Bitti *et al.*, "Universals and cultural differences in the judgments of facial expressions of emotion." *Journal of personality and social psychology*, vol. 53, no. 4, p. 712, 1987.
[78] P.-Y. Oudeyer and F. Kaplan, "What is intrinsic motivation? a typology of computational approaches," *Frontiers in neurorobotics*, vol. 1, 2007.
[79] R. E. Suri, "Td models of reward predictive responses in dopamine neurons," *Neural networks*, vol. 15, no. 46, pp. 523–533, 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0893608002000461
[80] W. C. Drevets, C. Gautier, J. C. Price, D. J. Kupfer, P. E. Kinahan, A. A. Grace, J. L. Price, and C. A. Mathis, "Amphetamine-induced dopamine release in human ventral striatum correlates with euphoria," *Biological Psychiatry*, vol. 49, no. 2, pp. 81–96, 2001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0006322300010386
[81] T. Campbell, E. OBrien, L. Van Boven, N. Schwarz, and P. Ubel, "Too much experience: A desensitization bias in emotional perspective taking," *Journal of Personality and Social Psychology*, vol. 106, no. 2, p. 272, 2014.

Joost Broekens Joost Broekens is assistant professor of Affective Computing at the LIACS of Leiden University and the Intelligent Systems Department of the TU Delft (NL), and co-founder and CTO of Interactive Robotics. He received a PhD in computer science at the University of Leiden (NL, 2007). He research activities cover computational models of emotion (applied in games, robots, agents, and theoretical), as well as computational modeling of mood (ranging from self-report to expression through robotic gestures in human-robot interaction). He is member of the executive board of the Association for the Advancement of Affective Computing (AAAC), associate editor of the Adaptive Behavior journal, and member of the steering committee of the IEEE Affective Computing and Intelligent Interaction Conference. He has organized multiple interdisciplinary workshops on topics including computational modeling of emotion (Lorentz, Leiden, 2011), grounding emotion in adaptation (IROS, 2016), and emotion as feedback signals (Lorentz, Leiden, 2016). He edited several special issues on these topics in, e.g., Springer LNAI, IEEE Transactions on Affective Computing, and Adaptive Behavior. His research interests include emotions in reinforcement learning, computational models of cognitive appraisal, emotions in games, human perception and effects of emotions expressed by virtual agents and robots, and emotional and affective self-report.

Mohamed CHETOUANI Prof. Mohamed Chetouani is the head of the IMI2S (Interaction, Multimodal Integration and Social Signal) research group at the Institute for Intelligent Systems and Robotics (CNRS UMR 7222), University Pierre and Marie Curie-Paris 6. He is currently a Full Professor in Signal Processing, Pattern Recognition and Machine Learning at the UPMC. His research activities, carried out at the Institute for Intelligent Systems and Robotics, cover the areas of social signal processing and personal robotics through non-linear signal processing, feature extraction, pattern classification and machine learning. He is also the co-chairman of the French Working Group on Human-Robots/Systems Interaction (GDR Robotique CNRS) and a Deputy Coordinator of the Topic Group on Natural Interaction with Social Robots (euRobotics). He is the Deputy Director of the Laboratory of Excellence SMART Human/Machine/Human Interactions In The Digital Society. In 2016, he was a Visiting Professor at the Human Media Interaction group of University of Twente. He is the coordinator of the ANIMATAS H2020 Marie Sklodowska Curie European Training Network.