Reinforcement Learning Models of Emotion: Computational Challenges

Joost Broekens[1]

Abstract. In this paper we address the field that computationally studies the relation between adaptive behavior and emotion. This field studies how affective phenomena emerge from simulated adaptive agents and how these agents and their human interaction partners can benefit from this. In particular, we focus on four major challenges when adaptive behavior is operationalized as an agent that learns to solve a task using reinforcement learning (RL) and affect is a signal that is derived from RL primitives and emerges during the interaction of the agent with its environment. For example, learned state utility, $V(s)$, is a signal that resembles fear (negative) and hope (positive), because these emotions signal the anticipation of loss or gain. The four challenges resolve around the following questions: why would a particular signal be labeled as an emotion; is there a generic structure in humans to how mood, emotion and appraisal influence reinforcement learning and action selection; what should benchmark tests look like if we want to investigate the plausibility and effectiveness of an emotional instrumentation of RL; are there other benefits to emotion instrumentation than increased adaptive potential for artificial agents?

1 INTRODUCTION

In this paper we address four major challenges in the field that studies the relation between emotion and Reinforcement Learning (RL). However, we first motivate why it is useful in the first place to study how emotions emerge from a mechanism for adapting behavior. Of course there is a theoretical benefit for psychology and behavioral science to gaining insight into the relation between emotion and adaptive behavior through computational modelling, but there are certainly also applied benefits. First, affective signals can enhance the adaptive potential of an artificial agent by influencing the learning process and action selection [56, 57, 38, 8, 10, 30, 55]. Second, if emotions emerge from RL during interaction with an environment, then any RL-based adaptive agent automatically possesses a computational model of emotion, which reduces the need to design a specific emotion model for that particular agent. Third, a solid grounding of emotion in adaptive behavior makes the expression of that emotion by a Virtual Agent or Robot intrinsically meaningful to humans, because we can relate to why the emotional signal arises. Having emotions emerge from RL variables would solve the grounding problem of emotion in RL-based artificial agents [17, 33], i.e., what does an emotion mean in terms of the functioning of the agent. Solving this problem might seem an abstract and theoretical goal, but this is far from the truth. For example, an adaptive robot that shows fear that is grounded in its learning mechanisms will be much easier to understand for humans, simply because we humans know what it means to have fear when learning to adapt to an environment. So, solving the grounding problem directly helps human-robot and human-agent interaction. This means that for an emotional instrumentation to be useful, adaptive benefit per se is not a requirement. Even if the emotion is purely an epiphenomenon, it is still useful for human-agent interaction and for understanding the fundamentals of how emotion emerges from adaptive behavior.

The underlying hypothesis in this field is that if (a) emotion and feedback-based adaptation of behavior is intimately connected in natural agents, and, (b) RL is a model for feedback-based adaptation of behavior in animals, then (c) this connection should be apparent in computational models of emotion for artificial agents that use RL to adapt their behavior. In this introduction we provide evidence for (a) and (b), which leads us to conclude that indeed it makes sense to computationally study the relation between emotion and reinforcement learning. Then we discuss computational attempts at modeling this relation and present the four major challenges.

We first investigate the premise that emotion and feedback-based adaptation of behavior is intimately connected in natural agents. A broadly agreed-upon function of emotion in humans and other animals is to provide a complex feedback signal for a(n) (synthetic) organism to adapt its behavior [25, 34, 44, 49, 51, 52, 13]. Important for the current discussion is that emotion provides feedback and that this feedback ultimately influences behavior, otherwise we can not talk about the adaptation of behavior. Behavior can be conceptualized as a sequence of actions. So, the generation of behavior eventually boils down to selecting appropriate next actions, a process called action selection [14, 47]. Brain mechanisms have been identified to be responsible for, or at the very least involved in, this process [6, 31]. An important signal that influences action selection in humans is how alternative actions feel. In neuroscience and psychology this signal is often referred to as somatic marker [18], affective value [53] or preference [65]. Another way in which emotion influences action selection is through emotion-specific action tendencies [25], such as the tendency to flee or startle when afraid. Emotion and feedback-based adaption seem to be intimately connected in natural agents via the process called action selection.

We now investigate our second premise; is there any support for RL being a plausible model of natural feedback-based learning? In RL an (artificial) organism learns, through experience, estimated utility of situated actions. It does so by solving the credit assignment problem, i.e., how to assign a value to an action in a particular state so that this value is predictive of the total expected reward (and punishment) that follows this action. After learning, the action selection process of the organism uses these learned situated action

[1] Delft University of Technology, Netherlands, email: joost.broekens@gmail.com

values to select actions that optimize reward (and minimize punishment) over time. Here we refer to situated action value as *utility*. In RL, reward, utility, and utility updates are *the* basic elements based on which action selection is influenced. These basic elements have been identified in the animal brain including the encoding of utility [62], changes in utility [29], and reward and motivational action value [3, 4, 5, 62]. In these studies it is argued that these processes relate to instrumental conditioning, in particular to the more elaborate computational model for instrumental conditioning called, indeed, Reinforcement Learning [19, 43]. It seems RL is a plausible model for feedback-based adaptation of behavior in animals.

So, the hypothesis that emotion and RL are intimately connected in animals is supported by the converging evidence that both RL and emotion seem to influence action selection using a utility-like signal. In neuroscience the connection between emotion, or affective signals in general, and reinforcement learning is confirmed by the large amount of work showing a relation between the orbitofrontal cortex, reward representation, and (subjective) affective value (for review see [53]). This connection can be studied computationally using RL-based adaptive agents. For example, different groups have shown that in some cases a mood-like signal emerging from the interaction between the agent and its environment can be used to optimize search behavior of an adaptive agent [8, 10, 30, 55] by manipulating the amount of randomness in the action selection process. Other groups have shown that explicit relations exist between emotion and the (prediction of) utility according to human subjects [28]. It has also been shown that adaptive agents that use their emotion as a feedback signal for learning, where that emotion itself is emergent from RL variables (such as reward, utility, and utility change) are in some tasks more adaptive then standard RL agents [56, 57] or learn faster [38]. Furthermore, already in 1999 an exhaustive attempt has been made to investigate different ways in which both emotion and RL can jointly influence action selection [27]. However, such studies face 4 major challenges that will be introduced in the next section.

2 CHALLENGES

First, it is a point of debate which and why particular signals coming from the RL variables or the agent's interaction with the environment should be labeled as "emotion", "mood", or "appraisal", which are the three affect-related situation dependent phenomena, let alone why a particular signal should be labeled as, e.g., "fear". For example, why is the amount of control (an appraisal as per [22]) equal to the difference between the utility of an agent's current situation minus the utility of the next situation, which represents the completeness of a learned interaction model [56] but not equal to the difference between the highest and lowest utility of the next situation, which represents the freedom of choice one has. Both instrumentations of control are equally plausible. Another example is the modeling of joy. 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 [61] on the one hand, and a correlation between dopamine signaling and euphoria exists on the other [21]. Joy reactions habituate upon repeated exposure to jokes [16] and computationally the temporal difference signal for a particular situation habituates upon repeated exposure [63], and both joy and the temporal difference signal are modulated by expectation of the reward. Does that mean that joy equals the temporal difference signal? The challenge is thus to come up with testable hypotheses about how emotion, mood and appraisal emerge from or are even equal to RL-based adaption-related signals. These hypotheses should be based on cognitive and behavioral theories describing eliciting conditions for emotions such as OCC [44] and [52].

Second, in order to test the validity of these hypotheses, we need benchmark scenarios that specify the emergence of affective phenomena during adaptation of behavior. Such benchmark scenarios do not exist. This is a serious issue because (a) researchers re-invent similar scenarios with different names [8, 27, 56] without replicating results of others first, and, (b) small differences in scenarios can have major influences [30, 8, 55]. The underlying issue is that we do not know how individual elements of the scenario such as the extrinsic reward function (stochastic or not), the problem to be learned (Markovian or not, stochastic or not), the learning mechanism (e.g., Q-learning versus TD(1)), and policy dependency (on-policy versus off-policy) influence the emergence of emotion for a particular RL instrumentation. This means that it is unclear if emergence of an emotional phenomenon may be generalized. To give an example, fear extinction is a phenomenon any fear instrumentation should be able to show. However, what kind of learning task (e.g., foraging task, partially observable, stochastic reward function) is needed to test for fear extinction after an initial negative encounter? The challenge is therefore to device benchmark scenarios that can be used to test for the replication of affective phenomena *in computo*, such as habituation and fear extinction. The goal for these scenarios is to test the plausibility of hypotheses about emergence of emotion, mood and appraisal, e.g., to test the validity of a claim such as "the utility of the current state equals hope/fear".

Third, it is unclear how emotion, mood and appraisal influence action selection in the RL architecture. Action selection can be influenced in roughly three ways [56]: directly by changing action values, indirectly by changing action-selection parameters, or indirectly by influencing what is called the intrinsic reward function (the intrinsic reward is the signal used for learning, while the extrinsic reward is the feedback signal from the environment). Mood, appraisal and emotion can therefore influence action selection each in three different ways. The question is why one way of influencing is more plausible or useful than another. For example, why would mood influence action selection meta parameters [8] but not action values? Is this computationally easier? Is this more effective in terms of adaption benefit? Is it biologically more plausible? The challenge is to formulate testable hypotheses about how affective phenomena influence action selection. These hypotheses should, again, be based on existing research on affective influences on cognition and behavior.

Fourth, it is unclear what kind of adaptive benefits and 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. In the majority of the cases the average (or final) fitness is the outcome measure to optimize. However, this is a one-dimensional approach to the role of emotion in adaptation. For example, fear can be used to influence intrinsic reward or randomness in action selection, and then it might serve the agent on average over the course of its lifetime. This instrumentation of fear forgoes the biological function of the fear response, which is to influence action selection immediately: fear makes an organism flee or wait (startle), and, for example, waiting simply stalls action selection to gain more time to absorb information. In scenarios where waiting makes sense, the effect of the latter fear instrumentation will thus be different compared to scenarios where waiting has no function (e.g., when the agent is the only actor that triggers state changes). In addition to adaptive benefits there are human-agent interaction benefits, as mentioned in the introduction. The challenge is thus to define benchmark scenarios in which particular adaptive or interaction benefits can appear. These scenarios can be different from the ones aimed for in challenge two, and need not be psychologically or biologically inspired per se.

3 APPROACH

We propose to tackle these challenges as follows. First, as a community we need an overview of specific adaptive and interaction benefits that can be expected from emotion, mood and appraisal in an RL setting. Examples include increased overall fitness (higher average reward), faster learning convergence, quicker recovery from changes in the environment, and increased human understanding of the robot's current state in the learning process. For more information on this topic see the review [40]. Second, we need a publicly available set of benchmark scenarios to test these benefits, as explained above. Third, to test individual hypotheses about how (or if) emotion, mood and appraisal emerge from RL, benchmark scenarios need to be developed by behavioral scientists, emotion psychologists and RL researchers interested in the role of emotion in adaptive behavior. Each scenario should involve a RL task, a prediction of which affective phenomenon emerges when over the course of learning to adapt to the task, and finally a set of experimental variables of which variation has a, in psychology and behavioral science, well-known effect on the affective phenomena. The last step to address involves the development of a shared simulation environment, analogous to the agent negotiation community, which successfully launched a shared test bed for agent-agent negotiation [36]. One important benefit of this is that differences in results between groups of researchers cannot be attributed to implementation differences. This is the last step, as the requirements for this simulation environment should be derived from the benchmark scenarios and expected adaptive and interaction benefit. These steps require collaboration, coordination and the building of a community, and funding is necessary for this to happen effectively and efficiently.

The final step involves goodwill and patience. If we accept the validity of our benchmark scenarios and simulation environment, then we should accept results that (do not) support a hypothesis that a particular RL signal should indeed be labeled as a particular emotion. Only then can we, in a structured way, investigate these hypotheses and build support for them, or reject them after careful consideration. If others do not agree with the affective label we gave to a particular RL-based signal, then our first reaction should be to investigate why this research passed the benchmark tests. As argued in [40] a more solid approach is essential for better shared definitions of affective signals in RL agents.

4 FIRST STEPS

Preliminary investigation [9, 32] suggests three computational (or cognitive) requirements to model emotions in terms of adaptation-related signals. The first concerns the element of novelty. Novelty covers concepts such as predictability, familiarity and suddenness, which are important factors in many emotions [2, 45, 37, 48, 54], in particular fear and surprise. Adaptive systems can only represent novelty if some form of prediction is present in the system. For example, if a model contains likelihoods of next states then this allows the derivation of novelty where less likely states are assumed novel. The second requirement concerns situated adaptation intensity. By this we mean that the intensity of an adaptation-related signal varies based on the desirability of the situation that causes the adaptation-related signal [26]. For example, humans differentiate between important and less important events, resulting in strong and weak emotions respectively. The third requirement is the existence of some order of development of signals. In early childhood, humans develop their emotions in order, from simple to complex, where the more complex emotions seem to depend on the existence of simple emotions [35, 34]. An adaptive system inherently lacking such an order of development in its adaptation-related signals cannot have these mapped correctly (time-wise) onto complex emotions. For example, fear (the worrying form, not the startle form) in humans occurs later in development than distress, and for an RL algorithm to model this there must be some adaption-related signal that only becomes relevant after some initial learning has occurred. This set of requirements strongly limits the number of available RL algorithms, but there are still many that meet them all (e.g., all model-based RL).

There are are also emotion-theoretical requirements for RL-based models of emotion. With regards to the order of development of emotions in infants, humans start with a small number of distinguishable emotions that increases during development. In the first months of infancy, children exhibit a narrow range of emotions consisting of distress and pleasure [59]. Joy and sadness emerge by 3 months, anger around 4 to 6 months with fear usually reported first at 7 or 8 months [59]. We propose to start by modeling the elicitation of the emotions of joy, distress, hope, fear, relief, disappointment and boredom. This set of emotions is meaningful with respect to learning [46], represents actual feedback, anticipated feedback and reflective feedback, and is realistically modeled with RL because these emotions depend mainly on novelty detection and goal congruence [50] (note that agent-directed emotions pride, shame and anger, are more difficult because this needs the concept of agency, something not inherently present in RL-agents). In line with this effort we have recently proposed that joy and distress can be modeled by the temporal difference signal [9] and that RL agents that learn to walk a slippery cliff best simulate the emotions of hope and fear by taking their model-based *predictions* of future temporal difference signals [41]. These findings are in congruence with the cognitive emotion theory by Reisenzein [49].

Important characteristics of emotions that should be present in benchmark scenarios include emotion intensity, habituation and extinction. Habituation is the decrease in intensity of the response to a reinforced stimulus resulting from that stimulus-reinforcer pair being repeatedly received, while extinction is the decrease in intensity of a response when a previously conditioned stimulus is no longer reinforced [42, 7, 23, 64]. A model of emotion based on adaption-related signals should be consistent with habituation and extinction, and in particular fear extinction as this is a well-studied phenomenon [42]. For example, learning to walk a slippery slope should always involve some fear, though the intensity should gradually decrease after each successful passage.

Affect and cognition are also related on a deeper more implicit level. Mood influences cognition. For example, a positive versus a slightly negative mood is known to make humans focus on the big picture or on detail respectively [24]. By artificially biasing the update signal (TD error) toward the positive or negative side, we can investigate if this effect can be replicated, e.g., by showing that the agent tends to perform respectively big-picture behavior in the form of exploration in the first case (and feel more joy) versus detail behavior in the form of exploitation in the second (and feel less joy). Cognition also influences mood. For example, humans employ a system that determines the balance between evaluating immediate versus delayed rewards, which is usually attributed to particular dopamine neurons [39]. That is, long term thinking involves a larger influence of future rewards, meaning the intensity of emotions linked to expectation (such as hope) can be expected to be higher for long-term thinkers. This could be modeled with the lambda in Reinforcement Learning. Lambda is a learning parameter that influences the effect of future rewards on the current state. Thus, by varying learning parameters, we should be able to replicate effects on behavior and emotion resulting from deeper relations between cognition and affect. Such simulations have been performed by us [11, 12] and others [15, 20, 55], but never in relation to the occurring emotions.

5 OUTLOOK

We have argued for the need to systematically study emotions in relation to reinforcement learning. We have listed 4 main challenges and taken first steps at how to attack these. However, we feel these challenges should be dealt with by the affective computing, adaptive behavior, psychology and neuroscience communities together. We also feel that the fields of reinforcement learning and emotion research can benefit from each other. Both fields essentially study animal adaptation mechanisms, but from a complementary perspective. We think that after 17 years of sporadic research in this field, sometimes explicitly referred to as investigating the link between emotion and adaptation, sometimes implicitly, the field needs a shared experimental practice and agenda. Without this, debates will be about the details, not about the big questions, and this will (and has) hinder(ed) progress. It is strange that two areas of study that deal so explicitly with adaptation and action selection in natural and artificial agents, that are both grounded in what can easily be called the fundamentals of animal adaption (feedback, reward, conditioning, anticipation, motivation, curiosity, stress, policy learning), and that are both mature areas in and of itself, would not come up with a program to merge insights. We feel, given recent publication activity in the area of affect, reinforcement learning, intrinsic motivation and neuroscience [1, 56, 58, 60, 53, 34, 9, 41], that the time to do this is now.

REFERENCES

[1] Joscha Bach, 'Modeling motivation and the emergence of affect in a cognitive agent', *Theoretical Foundations of Artificial General Intelligence*, 241–262, (2012).

[2] C. Bartneck, 'Integrating the occ model of emotions in embodied characters', in *Proceedings of the Workshop on Virtual Conversational Characters*, p. 3948, (2002).

[3] Kent C. Berridge, 'Pleasures of the brain', *Brain and Cognition*, **52**(1), 106–128, (2003).

[4] Kent C. Berridge and Terry E. Robinson, 'Parsing reward', *Trends in Neurosciences*, **26**(9), 507–513, (2003).

[5] Kent C. Berridge, Terry E. Robinson, and J. Wayne Aldridge, 'Dissecting components of reward: liking, wanting, and learning', *Current Opinion in Pharmacology*, **9**(1), 65–73, (2009).

[6] Rafal Bogacz and Kevin Gurney, 'The basal ganglia and cortex implement optimal decision making between alternative actions', *Neural Computation*, **19**(2), 442–477, (2007).

[7] Nicolas Luis Bottan and Ricardo Perez Truglia, 'Deconstructing the hedonic treadmill: Is happiness autoregressive?', *The Journal of Socio-Economics*, **40**(3), 224–236, (2011).

[8] Joost Broekens, *Affect and Learning: a computational analysis*, Phd thesis, Leiden University, 2007.

[9] Joost Broekens, Elmer Jacobs, and Catholijn M. Jonker, 'A reinforcement learning model of joy, distress, hope and fear', *Connection Science*, 1–19, (2015).

[10] Joost Broekens, Walter Kosters, and Fons Verbeek, 'On affect and self-adaptation: Potential benefits of valence-controlled action-selection', in *Bio-inspired Modeling of Cognitive Tasks*, pp. 357–366, (2007).

[11] Joost Broekens, Walter Kosters, and Fons Verbeek, 'On affect and self-adaptation: Potential benefits of valence-controlled action-selection', in *Bio-inspired Modeling of Cognitive Tasks*, eds., J. Mira and J. A lvarez, pp. 357–366. Springer Berlin Heidelberg, (2007).

[12] Joost Broekens, Walter A. Kosters, and Fons J. Verbeek, 'Affect, anticipation, and adaptation: Affect-controlled selection of anticipatory simulation in artificial adaptive agents', *Adaptive Behavior*, **15**(4), 397–422, (2007).

[13] Joost Broekens, Stacy Marsella, and Tibor Bosse, 'Challenges in computational modeling of affective processes', *IEEE Transactions on Affective Computing*, **4**(3), (2013).

[14] Joanna J. Bryson, 'Mechanisms of action selection: Introduction to the special issue', *Adaptive Behavior - Animals, Animats, Software Agents, Robots, Adaptive Systems*, **15**(1), 5–8, (2007).

[15] Joanna J. Bryson and Emmanuel Tanguy, 'Simplifying the design of human-like behaviour: Emotions as durative dynamic state for action selection', *Int. J. Synth. Emot.*, **1**(1), 30–50, (2010).

[16] Troy Campbell, Ed OBrien, Leaf Van Boven, Norbert Schwarz, and Peter Ubel, 'Too much experience: A desensitization bias in emotional perspective taking', *Journal of Personality and Social Psychology*, **106**(2), 272, (2014).

[17] L. Canamero, 'Emotion understanding from the perspective of autonomous robots research', *Neural networks*, **18**(4), 445–455, (2005).

[18] A. R. Damasio, *Descartes' Error: emotion reason and the human brain*, Penguin Putnam, 1996.

[19] Peter Dayan and Bernard W. Balleine, 'Reward, motivation, and reinforcement learning', *Neuron*, **36**(2), 285–298, (2002).

[20] Kenji Doya, 'Metalearning and neuromodulation', *Neural networks*, **15**(46), 495–506, (2002).

[21] Wayne C. Drevets, Clara Gautier, Julie C. Price, David J. Kupfer, Paul E. Kinahan, Anthony A. Grace, Joseph L. Price, and Chester A. Mathis, 'Amphetamine-induced dopamine release in human ventral striatum correlates with euphoria', *Biological Psychiatry*, **49**(2), 81–96, (2001).

[22] P. C. Ellsworth and K.R Scherer, 'Appraisal processes in emotion', in *Handbook of the affective sciences*, eds., R.J. Davidson, H.H. Goldsmith, and K.R. Scherer, p. 572595, New York, (2003). Oxford University Press.

[23] Edna B Foa and Michael J Kozak, 'Emotional processing of fear: exposure to corrective information', *Psychological bulletin*, **99**(1), 20, (1986).

[24] B. L. Fredrickson, 'What good are positive emotions?', *Review of general psychology*, **2**(3), 300, (1998).

[25] N. H. Frijda, 'Emotions and action', in *Feelings and Emotions: the amsterdam symposium*, eds., A. S. R. Manstead and Nico H. Frijda, p. 158173. Cambridge University Press, (2004).

[26] Nico H Frijda, Peter Kuipers, and Elisabeth Ter Schure, 'Relations among emotion, appraisal, and emotional action readiness', *Journal of Personality and Social Psychology*, **57**(2), 212, (1989).

[27] Sandra Clara Gadanho, *Reinforcement Learning in Autonomous Robots: An Empirical Investigation of the Role of Emotions.*, Ph.D. dissertation, Edinburgh, 1999.

[28] Jonathan Gratch, Stacy Marsella, Ning Wang, and Brooke Stankovic, 'Assessing the validity of appraisal-based models of emotion', in *3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009.*, (2009).

[29] Masahiko Haruno and Mitsuo Kawato, 'Different neural correlates of reward expectation and reward expectation error in the putamen and caudate nucleus during stimulus-action-reward association learning', *Journal of neurophysiology*, **95**(2), 948–959, (2006).

[30] Eric Hogewoning, Joost Broekens, Jeroen Eggermont, and Ernst Bovenkamp, 'Strategies for affect-controlled action-selection in soar-rl', in *IWINAC 2007, Part II, LNCS 4528*, eds., J. Mira and J.R. lvarez, pp. 501–510, Berlin, (2007). Springer.

[31] J.C Houk, C Bastianen, D Fansler, A Fishbach, D Fraser, P.J Reber, S.A Roy, and L.S Simo, 'Action selection and refinement in subcortical loops through basal ganglia and cerebellum', *Philosophical Transactions of the Royal Society B: Biological Sciences*, **362**(1485), 1573–1583, (2007).

[32] Elmer Jacobs, Joost Broekens, and Catholijn Jonker, 'Joy, distress, hope and fear in reinforcement learning', in *Proceedings of the 2014 inter-*

national conference on Autonomous agents and multi-agent systems, eds., Ana Bazzan, Michael Huhns, Alessio Lomuscio, and Paul Scerri, pp. 1615–1616. International Foundation for Autonomous Agents and Multiagent Systems, (2014).
[33] Kiril Kiryazov, Robert Lowe, Christian Becker-Asano, and Tom Ziemke. Modelling embodied appraisal in humanoids: Grounding pad space for augmented autonomy, 2011.
[34] Marc D. Lewis, 'Bridging emotion theory and neurobiology through dynamic systems modeling', *Behavioral and Brain Sciences*, **28**(02), 169–194, (2005).
[35] Michael Lewis, Jeannette M. Haviland-Jones, and Lisa Feldman Barrett, *Handbook of Emotions*, The Guilford Press, third edn., 2008.
[36] Raz Lin, Sarit Kraus, Tim Baarslag, Dmytro Tykhonov, Koen Hindriks, and Catholijn M Jonker, 'Genius: An integrated environment for supporting the design of generic automated negotiators', *Computational Intelligence*, (2012).
[37] Luis Macedo, Amilcar Cardoso, Rainer Reisenzein, Emiliano Lorini, and C Castelfranchi, 'Artificial surprise', *Handbook of Research on Synthetic Emotions and Sociable Robotics: New Applications in Affective Computing and Artificial Intelligence. IGI Global, Hershey: USA*, (2009).
[38] RP Marinier and John E Laird, 'Emotion-driven reinforcement learning', in *Cognitive Science*, pp. 115–120, (2008).
[39] Samuel M. McClure, David I. Laibson, George Loewenstein, and Jonathan D. Cohen, 'Separate neural systems value immediate and delayed monetary rewards', *Science*, **306**(5695), 503–507, (2004).
[40] Thomas Moerland, J. Broekens, and Catholijn M. Jonker, 'Computational models of emotion in reinforcement learning agents: a survey', *Machine Learning*, (in press).
[41] Thomas Moerland, Joost Broekens, and Catholijn. M. Jonker, 'Fear and hope emerge from anticipation in model-based reinforcement learning', in *Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16)*, pp. 848–854, (2016).
[42] K. M. Myers and M. Davis, 'Mechanisms of fear extinction', *Mol Psychiatry*, **12**(2), 120–150, (2006).
[43] John P. O'Doherty, 'Reward representations and reward-related learning in the human brain: insights from neuroimaging', *Current opinion in neurobiology*, **14**(6), 769–776, (2004).
[44] Andrew Ortony, Gerald L. Clore, and Allan Collins, *The Cognitive Structure of Emotions*, Cambridge University Press, 1988.
[45] Andrew Ortony, Gerald L. Clore, and Allan Collins, *The Cognitive Structure of Emotions*, Cambridge University Press, 1988.
[46] Reinhard Pekrun, Thomas Goetz, Wolfram Titz, and Raymond P Perry, 'Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research', *Educational psychologist*, **37**(2), 91–105, (2002).
[47] Tony J Prescott, Joanna J Bryson, and Anil K Seth, 'Introduction. modelling natural action selection', *Philosophical Transactions of the Royal Society B: Biological Sciences*, **362**(1485), 1521–1529, (2007).
[48] Rainer Reisenzein, 'Pleasure-arousal theory and the intensity of emotions', *Journal of Personality and Social Psychology*, **67**(3), 525–539, (1994).
[49] Rainer Reisenzein, 'Emotional experience in the computational belief-desire theory of emotion', *Emotion Review*, **1**(3), 214–222, (2009).
[50] Rainer Reisenzein, 'Emotions as metarepresentational states of mind: Naturalizing the belief-desire theory of emotion', *Cognitive Systems Research*, **10**(1), 6–20, (2009).
[51] Peter Robinson and Rana el Kaliouby, 'Computation of emotions in man and machines', *Philosophical Transactions of the Royal Society B: Biological Sciences*, **364**(1535), 3441–3447, (2009).
[52] Edmund T. Rolls, 'Precis of the brain and emotion', *Behavioral and Brain Sciences*, **20**, 177–234, (2000).
[53] Edmund T. Rolls and Fabian Grabenhorst, 'The orbitofrontal cortex and beyond: From affect to decision-making', *Progress in Neurobiology*, **86**(3), 216–244, (2008).
[54] Klaus R. Scherer, Angela Schorr, and Tom Johnstone, *Appraisal Processes in Emotion: Theory, Methods, Research*, Oxford University Press, 2001.
[55] Nicolas Schweighofer and Kenji Doya, 'Meta-learning in reinforcement learning', *Neural networks*, **16**(1), 5–9, (2003).
[56] Pedro Sequeira, *Socio-Emotional Reward Design for Intrinsically Motivated Learning Agents*, Ph.D. dissertation, Inesc-ID Lisbon, 2013.
[57] Pedro Sequeira, FranciscoS Melo, and Ana Paiva, 'Emotion-based intrinsic motivation for reinforcement learning agents', in *Affective Computing and Intelligent Interaction*, eds., Sidney DMello, Arthur Graesser, Bjrn Schuller, and Jean-Claude Martin, volume 6974 of *Lecture Notes in Computer Science*, pp. 326–336. Springer Berlin Heidelberg, (2011).
[58] S. Singh, R. L. Lewis, A. G. Barto, and J. Sorg, 'Intrinsically motivated reinforcement learning: An evolutionary perspective', *Autonomous Mental Development, IEEE Transactions on*, **2**(2), 70–82, (2010).
[59] L Alan Sroufe, *Emotional development: The organization of emotional life in the early years*, Cambridge University Press, 1997.
[60] John E. Steephen, 'Hed: A computational model of affective adaptation and emotion dynamics', *IEEE Transactions on Affective Computing*, **4**(2), 197–210, (2013).
[61] Roland E. Suri, 'Td models of reward predictive responses in dopamine neurons', *Neural networks*, **15**(46), 523–533, (2002).
[62] Saori C. Tanaka, Kenji Doya, Go Okada, Kazutaka Ueda, Yasumasa Okamoto, and Shigeto Yamawaki, 'Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops', *Nat Neurosci*, **7**(8), 887–893, (2004).
[63] G. Tesauro, 'Temporal difference learning and td-gammon', *Communications of the ACM*, **38**(3), 58–68, (1995).
[64] Ruut Veenhoven, 'Is happiness relative?', *Social Indicators Research*, **24**(1), 1–34, (1991).
[65] Robert B. Zajonc and Hazel Markus, 'Affective and cognitive factors in preferences', *Journal of Consumer Research*, **9**(2), 123, (1982).