The Role of Simulated Emotions in Reinforcement Learning
Insights from a Human-Robot Interaction Experiment.

Nijeholt, Floortje Lycklama À.; Broekens, Joost

DOI
10.1109/ACIIW59127.2023.10388167

Publication date
2023

Document Version
Final published version

Published in
2023 11th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos, ACIIW 2023

2023 11th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW) | 979-8-3503-2745-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ACIIW59127.2023.10388167

The Role of Simulated Emotions in Reinforcement Learning: Insights from a Human-Robot Interaction Experiment.

1st Floortje Lycklama à Nijeholt
Delft University of Technology
Delft, The Netherlands
floorlycklama@gmail.com

2nd Joost Broekens
Leiden Institute of Advanced Computer Science
Leiden University
Leiden, The Netherlands
joost.broekens@gmail.com

***Abstract*—Transparency of behavior is important for robots that work with humans. If such robots need to adapt to a variety of users and tasks, they need to learn to optimize their behavior, and Reinforcement Learning (RL) is a promising learning method for this purpose. However, the behavior generated by RL is not inherently transparent due to the exploration/exploitation tradeoff that is needed to optimize a policy.**

Emotions are -for humans- a natural way of communicating intent and situational appraisal. In this study, we implemented emotional expressions based on Temporal Differences as a means to increase the transparency of a robot's learning process. We analysed the effect on the human teacher's behavior and experience, and on the robot's learning result and learning process.

A between-subject experiment with 61 participants and three robot conditions was performed: no emotions, simulated emotions, and simulated emotions with matching attribution. The learning task was one where a human teacher had to help a humanoid robot to learn the meaning of three colors.

Our results demonstrate minimal differences between these three conditions. This means that for simple tasks, emotional expressions grounded in RL do not help nor hurt. We discuss our findings and propose three important criteria for interactive learning tasks when investigating the effect of emotional expressions grounded in RL. Such tasks need to be sufficiently complex, afford robot autonomy, and the emotion must be informative about how the user could influence the robot's actions.

***Index Terms*—Reinforcement Learning, Temporal Difference, Human Teacher, Emotions, Human-Robot Interaction**

I. Introduction and motivation

Transparency of behavior is important for intelligent systems that work with humans [1], especially when such systems become increasingly more complex. When a system is transparent, users are able to better understand what these intelligent systems are doing and why.

This is no different for robots that work with humans. Transparency can help users better understand the reasoning behind the robot's behavior, enabling them to better assess the robot's capabilities [2]. Transparency also reduces conflict and improves the robustness of an interaction, particularly in team performance between robots and humans. [3].

If robots need to adapt to a variety of users and tasks, they need to learn to optimize their behavior. Reinforcement Learning (RL) is a promising learning method for this purpose [4]. By repeatedly interacting with the environment, an RL agent learns to adapt the values of actions to achieve the optimal state transition policy that maximizes the rewards over time.

However, the behavior generated by RL is not inherently transparent due to the exploration/exploitation tradeoff that is needed to optimize a policy for a specific task [5]. During exploration, the robot will perform actions that are not the best known actions at that moment in the learning process, which might be confusing for the users. During exploitation, especially premature exploitation (early convergence), the agent may select actions that are suboptimal, again potentially confusing the human.

Emotions are -for humans- a natural way of communicating intent and situational appraisal. Most emotion theories propose that emotions arise when a change in the situation is personally meaningful to the agent. In cognitive appraisal theories, emotion is often defined as a valenced reaction resulting from the cognitive assessment of personal relevance of an event [6]–[8]. The assessment is based on what the agent believes to be true and what it aims to achieve as well. In theories that emphasize biology, behavior, and evolutionary benefit [9], [10], the emotion is more directly related to action selection but still based on an assessment of harm versus benefit. Importantly, evidence suggests that the expression of the emotion mirrors the assessment [11]. Thus, if this assessment is grounded in RL, expression of the resulting emotion may help transparancy of the learning process.

The Temporal Difference Reinforcement Learning (TDRL) theory of emotion [12] proposes that emotions are manifestations of reward processing in Reinforcement Learning, in particular manifestations of temporal difference assessment. This Temporal Difference (TD) is the agent's perception of gain or loss of utility (well-being), resulting from new evidence. This is a reinterpretation of cognitive appraisal in terms of the reinforcement learning process.

This suggests that agents and robots that use RL to learn can also simulate and express emotions grounded in their learning process. Indeed, evidence suggests that the simulated emotions

are plausible [12]–[15]. However, experimental evidence that these emotions are useful, and plausible *for human teachers* in a robot-human interaction setting is lacking.

In this paper, we investigate the hypothesis that emotional expressions towards a human teacher based on Temporal Difference learning can be used as a means to enhance the transparency of a robot's learning process [5]. Specifically, the study examines the effect of robot emotional expressions and the explanation of the source of the emotion (attribution) on the teacher's behavior and experience, and on the robot's learning result and learning process.

II. Related work on emotion simulation based on Reinforcement Learning

While there are many studies that investigate the effect of robot emotional expression on human observers or teachers, in this paper we specifically focus on the effect of robot emotional expression based on Temporal-Differences in RL on the teacher and robot. Therefore in this paper we summarize related work on simulating emotions based on Reinforcement Learing.

Simulating emotions for RL agents and robots is not new (for review see [16]). Here we focus on previous work that simulates emotion elicitation based on the RL process, and studies the plausibility/impact of the simulated emotions from a human perspective. For example, Broekens et al. [13], propose a computational model of joy, distress, hope, and fear as mappings between RL primitives (reward, value, update signal, etc...) and emotion labels. Joy/distress is derived from the positive/negative temporal difference (TD) signal for the current state, and hope/fear is derived from the learned value of the current state. Results showed that emotions simulated in this way are consistent with emotion elicitation, emotion development, emotion habituation and fear extinction theory. Later work using the same framework showed plausible simulations of fear and hope [17], and regret [15], again based on a theoretical analysis of the simulated emotion intensities.

Moussa and Magnenat-Thalmann [18] included emotions, attachment and learning in a decision-making Q-learning architecture for a virtual agent. Their framework was evaluated by interacting with users in different scenarios. Preliminary results showed that the virtual agent showed appropriate emotional responses to different user behaviours.

A recent study simulated emotions based on temporal difference signals and presented participants with videos of the robot expressing these emotions [19]. The study was inconclusive with respect to the transparency gained from these emotional expressions. A later study [20] noticed a slight increase in transparency but also proposed a larger scale study.

III. Joy, Distress, Hope and Fear in the TDRL Theory of Emotion

The TDRL Theory of Emotion proposes that all emotions are manifestations of temporal difference errors [5], [12]. Emotion is defined as *valenced appraisals in reaction to (mental) events providing feedback to modify action tendencies, grounded in primary reinforcers* [12]. This idea of relating emotions to the processing of goals and progress is in line with recent cognitive theories of emotion [21], [22].

We now briefly explain what the TD error in RL is using SARSA as an example. SARSA is the simplest model-free RL method existing, and is an on-policy variant of Q-learning [23]. It learns action values $Q(s, a)$ based on repeated observations of states, actions, and rewards. It is called on-policy because adaption of the Q-values is based on the actual actions chosen, not on the best possible actions such as in Q-learning. It is called model-free because SARSA does not learn or use information about state transitions, only about Q-values. The Temporal Difference (the learning signal for the action values) is defined as follows:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \overbrace{(r_t + \gamma * Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t))}^{\text{Temporal Difference}} \quad (1)$$

In the TDRL Theory of Emotion, Joy is proposed to be the manifestation of a positive TD, while distress is a negative TD. As such, Joy and Distress are defined as follows [12]:

$$if(TD > 0) \Rightarrow Joy = TD \quad (2)$$

$$if(TD < 0) \Rightarrow Distress = TD \quad (3)$$

Hope is the anticipation (forward simulated) of a positive TD, while fear is the anticipation of a negative TD [12]. We use these definitions of joy, distress, hope, and fear as a basis for modeling robot emotion in this paper.

IV. Method

In this study, 61 adult participants (mean age 30.64, SD 13.147) were recruited to teach a robot three different colors (red, green, and blue). The robot uses SARSA to learn a policy. In the start state (see Figure 1), the robot asks the participants which color to change its eyes to, and the participant could respond with one of the three colors. The robot would then choose one of three actions [eyes-red, eyes-green, eyes-blue], and subsequently asks the participant if the color was correct, which the participant could confirm or deny. The actual reward was hardwired to make sure errors in interpretation of the reward were not possible. No mismatch between user response and hardwired reward occurred. The robot used ϵ-greedy with ϵ-decay for its action selection process during learning. Before the experiment, the participants were informed that participation was voluntary and that they would have 10 minutes to teach the robot the three colors, which would be more than enough time so that they would not feel rushed. After convergence, the task ended.

A. Experimental manipulation

A between-subject experiment was conducted, where each participant was randomly assigned to one of three different robot modes. The first mode (mode 1) had no simulated emotions, the second mode (mode 2) had simulated emotions joy, distress, hope, and fear, and the third mode (mode 3) had

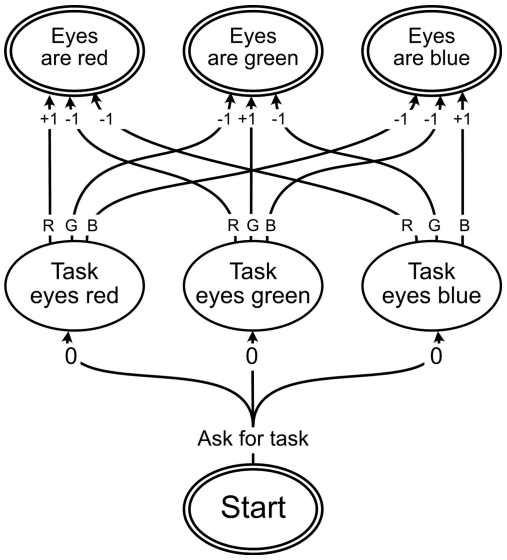


Fig. 1: A visualisation of the MDP of the experiment, with seven states, four actions, and their associated rewards

these simulated emotions and matching emotional attributions (expressed vocally by the robot as the cause of its emotion).

The emotions of the robot were based on the TD value in Equation 1. Joy and distress are based on the actually received TD and calculated as follows:

$$if(TD > 0.2) \Rightarrow Joy \tag{4}$$

$$if(TD < -0.2) \Rightarrow Distress \tag{5}$$

Hope and fear are based on the forward simulated TDs in the MDP, using a greedy simulation policy for the robot. TDs are calculated for the greedy actions in the *Task* states (see Figure 1). At the start state, the maximum absolute value of the three simulated TDs for the three possible stochastic outcomes (human picks red, green or blue) is selected and that TD is used for the emotion to express. If that TD is positive, the emotion is hope, if that TD is negative, the emotion is fear. Only those transitions that are actually observed are used. The calculations for hope and fear are:

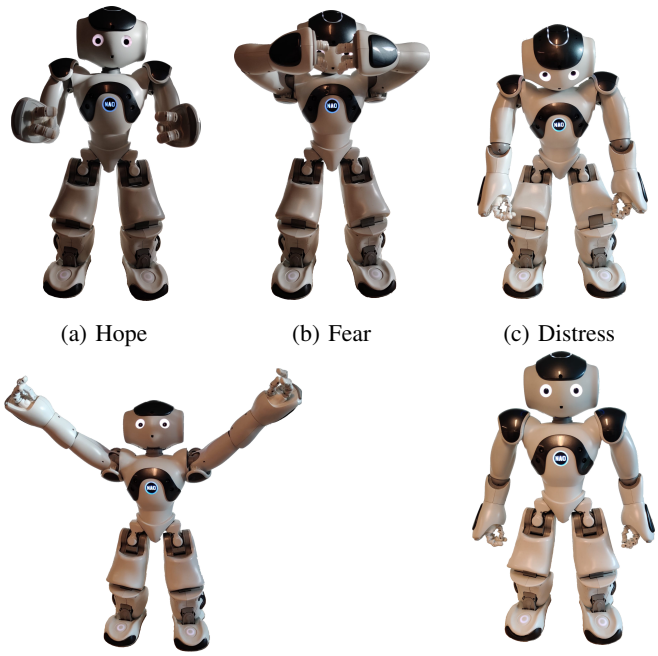
$$\begin{aligned} &TD = signed(\max(|TD_{a,s \in Taskstates}(argmax(Q(s,a)))|)) \\ &if(TD > 0.2) \Rightarrow Hope \\ &if(TD < -0.2) \Rightarrow Fear \end{aligned} \tag{6}$$

Joy and distress are expressed right after receiving the TD update for an action (i.e., after the transition to a new state). Hope and Fear are expressed *in* a state, before the robot action to ask the user for a command. Multiple emotions can be expressed in a row, for example, upon ariving in an outcome (color) state the robot can express distress if it calculates a negative TD due to a wrong choice, and then fear in the start state for the possible wrong choice it may make in the future.

As an emotion is either expressed or not (no intensity), we introduced a small threshold for the emotion elicitation, so that the robot only expresses an emotion when a significant change in the TD occurs. This was done to stop the robot expressing emotions when converging on the learning task.

The robot mode with matching emotional attributions explains to the user what color it is feeling hope or fear for (precisely: the TD associated with a transition to an outcome state). This explanation is in addition to expressing the emotion. As such, in this condition the robot explains the cause of the prospect-based emotions, the attribution of hope and fear. The robot only explains this to the user in the start start. If the robot were to simulate hope, it will add the statement "Ik hoop dat het {color} wordt!" or "I am hoping for {color}!" after the hope expression.

The emotions of hope, fear, joy, and distress were expressed using both the body language of the robot and verbal expressions. The poses used for these emotions were based on previous research by Thoma et al. [24] and Wu et al. [25], and are shown in Figure 2. The verbal expressions used for the Dutch robot mode and the English mode can be seen in Tables I and II, respectively. During the interaction, the algorithm randomly choses one of the three expressions to avoid repetition of statements.



(a) Hope (b) Fear (c) Distress

(d) Joy (e) Neutral

Fig. 2: The NAO in the poses for the 4 emotions and it's neutral pose.

B. Measurements

Learning progress (and final outcome) is measured based on the number of color outcome states towards which the Q-table has a maximum value transition above zero (robot has learned

TABLE I: The statements in Dutch made by the robot for different emotions and for the neutral mode without any emotions

Emotion	Expression 1	Expression 2	Expression 3
Hope	Ik heb er zin in	Kom maar op!	Dat gaat wel weer goed komen.
Fear	Oei dit vind ik spannend.	O nee dit gaat vast niet.	O nee, dit gaat fout.
Joy	Hoera	Jippie	Wat fijn
Distress	Drommels	Helaas	Wat jammer
Neutral	Oké	Bedankt	Prima

the correct action). The Q-table will have a value greater than zero when the correct color has been performed at least once.

All possible values for learning progress are:

0) no color learned
1) one color learned
2) two colors learned
3) three colors learned

For measuring the learning process, the following was collected:

- The progress of robot learning over iterations
- The ratio of exploration commands given by the user
- The ratio of exploitation commands given by the user
- The ratio of inefficient exploitations commands by the user (i.e., when the user selects a learned color when there are unknown colors left)
- The ratio of the user selecting another command the next trial, dependent on whether the previous command was correctly executed by the robot or not (i.e., the probability of "switching" to another color command depending on wether the robot correctly executed the last command)

After the experiment, the participants were asked to answer two questionnaires: the Godspeed Questionnaire [26] and the User Experience Questionnaire (UEQ) [27]. The Godspeed questionare was used to measure the perceived Anthropomorphism, Animacy, Likeability, and Intelligence. The UEQ was used to find the perceived Novelty, Stimulation, Efficiency, and Attractiveness.

C. Reliability checks

It was decided to not use the scores for perspicuity and dependability from the UEQ as the consistency of these scores

Emotion	Expression 1	Expression 2	Expression 3
Hope	I am looking forward to this	"Let's go!	Okay. Let's go
Fear	This is a bit scary for me	O no, it will go wrong again	Oh no, it will go wrong
Joy	Hooray	Nice	Lovely
Distress	O bother	Let's pretend that did not happen	How unfortunate
Neutral	Okay	Thank you	Check

TABLE II: The statements in English made by the robot for different emotions and for the neutral mode without any emotions

was very low. The Cronbach's Alpha for perspicuity was .291 and for dependability it was .494. For the other UEQ scores, the Cronbach's Alpha for attractiveness was .876, for efficiency it was .661, for it stimulation it was .733, and for novelty it was .628. For Godspeed Cronbach's Alpha was calculated as well. For anthropomorphisrm it was .791, for animacy it was .637, for Likeability it was .847, and for Perceived intelligence it was .623.

There was no difference in age, gender or number of participants between the conditions. Condition one had 20 participants, mean age of 30.20 years (SD = 13.950) and consisted of 11 males, 8 females, and one other. Condition two had 20 participants, a mean age of 28.95 years (SD = 9.288), with 13 males, and 7 females. Condition three had 21 participants, a mean age of 32.67 years (SD = 15.631) and 13 males and 8 females.

Finally, we asked subjects to rate their experience with robots and computer science, and checked with an ANOVA if conditions would predict a difference in experience, which was not the case [F(2,58=0.076, p=0.927].

V. Results

To investigate the effect of the different robot modes on the user experience a MANOVA test has been conducted on the Godspeed Questionnaire and the UEQ. Godspeed measured the user experience on perceived anthropomorphism, animacy, likeability, and intelligence. This MANOVA did not show any significant effects [F(8,108) = 1.325, $p = .239$]. The UEQ looked at the user experience on perceived novelty, stimulation, efficiency, and attractiveness. This MANOVA did not show any significant effects as well, [F(8,108) = 1.647, $p = .120$].

The p-values resulting from the univariate ANOVAs for each dependent variable can be seen in table III, showing that only novelty has a significant [F(2, 58) = 3.485, $p = .037$] effect, with a near significant effect for animacy [F(2, 58) = 3.103, $p = .052$].

To test for differences between the different robot modes, we performed Post-Hoc tests. For completeness, we report corrected (Bonferroni) and uncorrected (LSD) comparisons. Please note that table III shows the uncorrected p values. Corrected values can be calculated by dividing the uncorrected p values by 3 (the number of condition comparisons).

With Bonferroni correction for multiple comparisons [28], the only significant difference is found for perceived novelty between the no emotions and the emotions condition [Mean = 3.2411, SD = 0.57141; Mean = 3.6875, SD = 0.53186 respectively, with $p = 0.034$].

Without correction for multiple comparisons, the LSD Post Hoc test showed more significant differences (See Table). In particular, we found a significant difference between the means of Animacy for the robot without emotional expressions [Mean = 2.8917, SD = .57297] and the robot mode with emotional expressions [Mean = 3.2667, SD = .49971] with $p = .027$. We also found a significant difference between the mean of Animacy for the robot without emotional expressions and the

mean of Animacy with emotional expressions and attribution [Mean = 3.2222, SD = .48971] with $p = .027$.

A near signiciant difference was found between the means of Attractiveness for the robot without emotional exressions [Mean = 3.5298, SD = 0.70933] and the robot mode with emotional expressions [Mean = 3.8209, SD = .61618], $p = 0.053$.

In figure 3 the mean scores for the User Experience questionnaire can be seen. The results of the UEQ have been scaled to the scale of the Godspeed questionnaire, so both questionnaires are on a scale from 1 to 5.

None of the other results on learning progress/outcome and learning process were significantly different between the robot conditions. We observed no differences for the number of iterations needed to converge, the color selection of the human, the exploration/exploitation behavior of the human, the learning progress over time, or the exploration/exploitation varying over time (see plots and bar charts).

	Post-Hoc			ANOVA
Score \ Between modes	1-2	1-3	2-3	
Attractiveness	.053	.132	.639	.127
Efficiency	.862	.370	.284	.512
Stimulation	.172	.313	.705	.366
Novelty	.011	.296	.115	.037
Perceived intelligence	.821	.890	.928	.974
Likeability	.108	.205	.712	.238
Animacy	.027	.047	.786	.052
Anthropomorphism	.586	.092	.250	.223
LS3 Location	.401	.577	.769	.692
Loops in LS0	.313	.749	418	.583
Loops in LS1	.775	.667	.473	.768
Loops in LS2	.395	.895	.467	.655
Inefficient Exploitation Ratio	.929	.429	.483	.683
Exploitation Ratio	.719	.745	.492	.788
Exploration Ratio	.719	.745	.492	.788
Different command after failure	.428	.170	.563	.385
Different command after succes	.641	.497	.835	.785
Same command after failure	.909	.444	.379	.628
Same command after succes	.569	.872	.462	.741

TABLE III: Post Hoc comparisons between robot modes and univariate ANOVAs. For the Post-hoc comparisons the table shows the uncorrected p values. Corrected p values can be calculated by multiplying the uncorrected p values by 3 (the number of condition comparisons).Yellow values indicate uncorrected significant differences (LSD), but not corrected significance (Bonferroni). Green indicates corrected significant differences (Bonferroni). For the ANOVA green indicates significant at the 0.05 level, light green indicates near significance at the 0.1 level. All values have been calculated with a univariance general linear model. For all ANOVA's the degrees of freedom were: F(2, 58).

VI. Conclusion and Discussion

Our results demonstrate minimal differences between a robot without emotional expressions, one with emotional expressions grounded in the learing process, and one with emo-

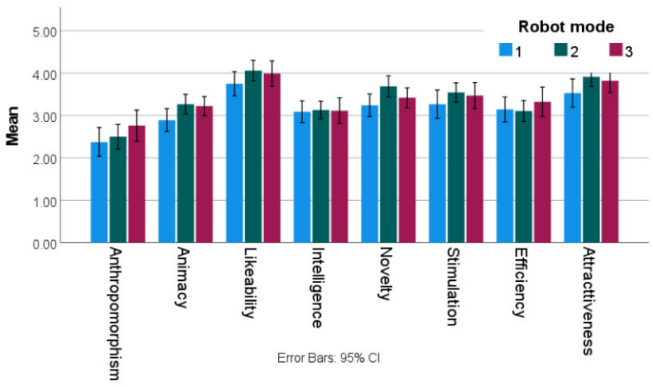


Fig. 3: Error bars and means of the Godspeeds scores for Anthropomorphism, Animacy, Likeability, and Intelligence and the scaled means for the UEQ scores for Novelty, Stimulation, Efficiency, and Attractiveness for the different robot modes with error bars

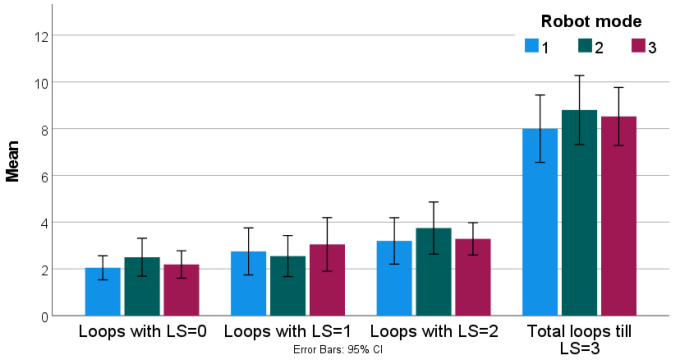


Fig. 4: Bar plot with error bars of the mean of the number of iterations the robot took to learn 1, 2, and 3 correct colors, with the total sum of iterations for the complete experiment.

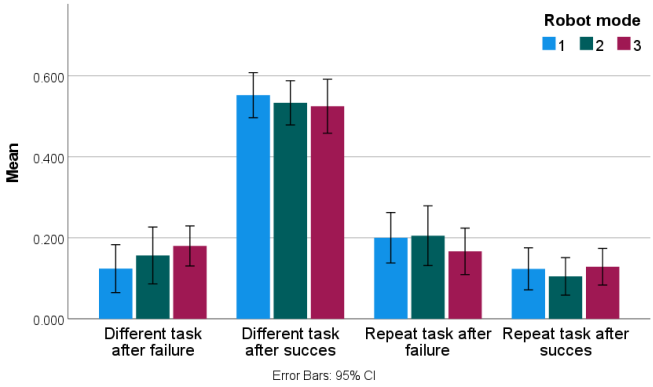


Fig. 5: Error bars and means of the ratios for the command selected by the user, dependent on the robot's success on the previous command

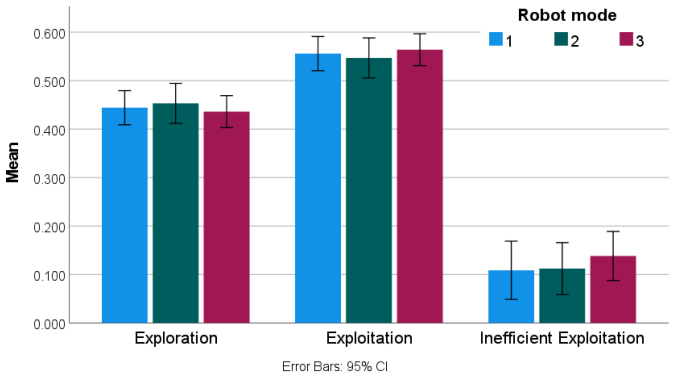


Fig. 6: Error bars and the means of the ratio of inefficient exploitation commands, exploration commands, and exploitation commands

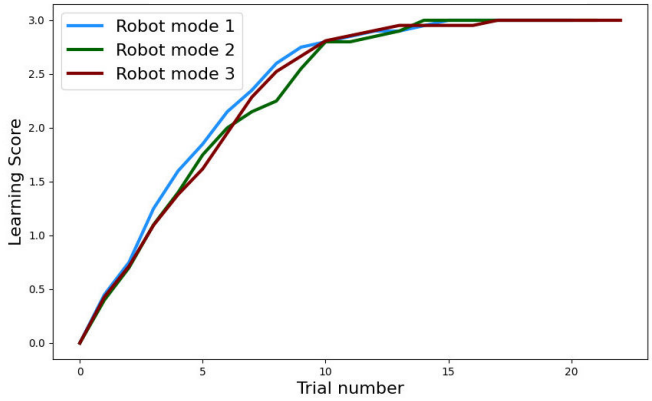


Fig. 7: Learning curves for three robot conditions, calculated as means of the number of correct colors learned over the iterations. Each curve is the mean over the particpants in that condition.

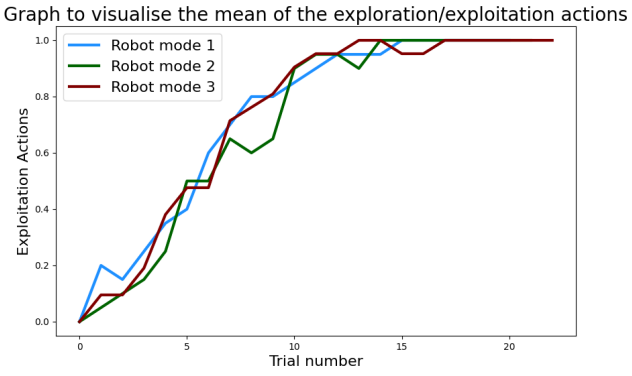


Fig. 8: Mean of the exploration/exploitation tradeoff throughout the iterations. A value of 1 indicates exploitation and a value of 0 indicates exploration. Each curve is the mean over the particpants in that condition.

tional expressions and the emotional attribution that belongs to the emotion.

This means that for simple tasks, emotional expressions grounded in RL do not help nor hurt. As the expressions of the robots were very clear and perceived as intended by participants, the emotional model was extensively evaluated in previous work, pilot experiments were done extensively to test the setup, and all of the participants understood the task, and the differences between the conditions are very clearly observable, we feel it is highly unlikely this lack of a clear result should be explained by a lack of manipulation or a methodological flaw in the setup. Also, as there is no negative effect on the user experience either, the expressions were apparently seen as natural, and did not hinder the human in the task.

The learning task was carefully constructed: it is simple enough to understand, involves an actual small RL problem (with a small MDP, not just a bandid - i.e., it was not an action value learing problem), it runs very smoothly and interactively on the NAO robot, and it was a clear teaching challenge, that was easy to monitor by the human teacher as well.

However, in retrospect we believe the task is not suitable, and therefore feel this insight is an important contribution to the field. We believe the reason of the minimal effect is threefold.

First, the task is simple and easy to oversee for a human. As the emotions are simulated based on the ”mental state” of the RL system of the robot, the additional information this emotion gives to the human teacher to understand the state of the agent is perhaps too little. To test if such emotion simulation is beneficial, the task needs to be more complex.

Second, even though the emotion expressed by the robot in this task provides information about the state of the robot, in this particular task, there is no reason for the human to adapt the teaching strategy, as the goal is to teach the robot the colors, wehther or not it is happy, sad, hopeful or fearful. To explain: what should the human do if the robot fears a particular color command from the human ? Not teach it that color? To test if emotion expression grounded in RL is beneficial, the emotion needs to be informative about how the user may *influence* the robot’s process to the benefit of the robots policy.

Third, as the human was in complete control, deciding which color the robot had to try, there was never a moment when the user had to intervene. For example, if fear is a genuine signal of danger ahead, the human teacher may stop the robot from going on and steer it towards another area of the task. Or, if the robot autonomously explores, and when the human sees hope, based on a falsely predicted future benefit, the human can stop the robot. For this, the robot needs to have some autonomy in the task, and a better impact measure of robot conditions may be the number of (constructive) interventions the human teacher performs.

It is important that future work on interactive robot learning with human teachers, where emotions are used as social signals to the teacher, takes into account these three aspects:

- The task complexity needs to be such that the human teacher cannot oversee the complete process easily.
- Emotions need to convey information about how the human can change the actions of the robot.
- The robot needs to have some meaningful autonomy in solving the task so that the human can intervene triggered by the emotion of the robot.

On a positive note, emotion expression, when properly grounded in the learning process apparently did not hinder the human teacher either.

VII. Ethical Impact statement

Prior ethical approval was obtained from Delft Univerity of Technology. Subjects provided written agreement for their participation. The time investment for the subjects was reasonable (approx. 15-30 minutes including debriefing). The environmental impact of the energy used for the robot and analysis is neglectable, no large scale cloud resources have been used. The work could have a minor impact on society, as it addresses the topic of robots with emotions. We would like to stress that robots do not feel anything in the sense that biological agents do. These emotions are simulations.

References

[1] R. R. Hoffman, S. T. Mueller, G. Klein, and J. Litman, "Metrics for explainable ai: Challenges and prospects," *arXiv preprint arXiv:1812.04608*, 2018.

[2] J. Y. Chen and M. J. Barnes, "Human–agent teaming for multirobot control: A review of human factors issues," *IEEE Transactions on Human-Machine Systems*, vol. 44, no. 1, pp. 13–29, 2014.

[3] 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*. IEEE, 2005, pp. 708–713.

[4] 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. [Online]. Available: http://ijr.sagepub.com/content/32/11/1238.abstract

[5] J. Broekens and M. Chetouani, "Towards transparent robot learning through tdrl-based emotional expressions," *IEEE Transactions on Affective Computing*, vol. 12, no. 2, pp. 352–362, 2019.

[6] 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. [Online]. Available: http://emr.sagepub.com/content/5/2/119.short

[7] A. C. A. Ortony, Gerald L. Clore, "The cognitive structure of emotions," *Cambridge University Press*, 1988.

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

[9] 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. 158–173.

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

[11] C. M. de Melo, P. J. Carnevale, S. J. Read, and J. Gratch, "Reading people's minds from emotion expressions in interdependent decision making," *Journal of Personality and Social Psychology*, vol. 106, no. 1, p. 73, 2014.

[12] J. Broekens, "A temporal difference reinforcement learning theory of emotion: A unified view on emotion, cognition and adaptive behavior," *arXiv preprint arXiv:1807.08941*, 2018.

[13] 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

[14] J. Broekens and L. Dai, "A tdrl model for the emotion of regret," in *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 2019, Conference Proceedings, pp. 150–156.

[15] L. Dai and J. Broekens, "Simulating fear as anticipation of temporal differences: an experimental investigation," in *2021 9th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*. IEEE, 2021, pp. 1–8.

[16] T. M. Moerland, J. Broekens, and C. M. Jonker, "Emotion in reinforcement learning agents and robots: a survey," *Machine Learning*, vol. 107, pp. 443–480, 2018.

[17] T. Moerland, J. Broekens, and C. Jonker, *Fear and Hope Emerge from Anticipation in Model-Based Reinforcement Learning*. AAAI Press, 2016, pp. 848–854.

[18] M. B. Moussa and N. Magnenat-Thalmann, "Toward socially responsible agents: integrating attachment and learning in emotional decision-making," *Computer Animation and Virtual Worlds*, vol. 24, no. 3-4, pp. 327–334, 2013.

[19] A. Rossi, M. M. Scheunemann, G. L'Arco, and S. Rossi, "Evaluation of a humanoid robot's emotional gestures for transparent interaction," in *Social Robotics: 13th International Conference, ICSR 2021, Singapore, Singapore, November 10–13, 2021, Proceedings 13*. Springer, 2021, pp. 397–407.

[20] G. Angelopoulos, A. Rossi, G. L'Arco, and S. Rossi, "Transparent interactive reinforcement learning using emotional behaviours," in *Social Robotics: 14th International Conference, ICSR 2022, Florence, Italy, December 13–16, 2022, Proceedings, Part I*. Springer, 2023, pp. 300–311.

[21] R. Reisenzein, "Emotional experience in the computational belief-desire theory of emotion," *Emotion Review*, vol. 1, no. 3, pp. 214–222, 2009.

[22] A. Moors, Y. Boddez, and J. De Houwer, "The power of goal-directed processes in the causation of emotional and other actions," *Emotion Review*, vol. 9, no. 4, pp. 310–318, 2017.

[23] G. A. Rummery and M. Niranjan, *On-line Q-learning using connectionist systems*. University of Cambridge, Department of Engineering Cambridge, UK, 1994, vol. 37.

[24] P. Thoma, D. S. Bauser, and B. Suchan, "Besst (bochum emotional stimulus set)—a pilot validation study of a stimulus set containing emotional bodies and faces from frontal and averted views," *Psychiatry Research*, vol. 209, no. 1, pp. 98–109, 2013.

[25] J. Wu, Y. Zhang, S. Sun, Q. Li, and X. Zhao, "Generalized zero-shot emotion recognition from body gestures," *Applied Intelligence*, pp. 1–19, 2022.

[26] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *International journal of social robotics*, vol. 1, pp. 71–81, 2009.

[27] M. Schrepp, A. Hinderks, and J. Thomaschewski, "Applying the user experience questionnaire (ueq) in different evaluation scenarios," in *Design, User Experience, and Usability. Theories, Methods, and Tools for Designing the User Experience: Third International Conference, DUXU 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceedings, Part I 3*. Springer, 2014, pp. 383–392.

[28] W. Haynes, "Bonferroni correction," *Encyclopedia of systems biology*, pp. 154–154, 2013.