Vicarious Value Learning and Inference in Human-Human and Human-Robot Interaction

1st Robert Lowe
Department of Applied IT
University of Gothenburg
Sweden
robert.lowe@gu.se

2nd Alexander Almér
Department of Applied IT
University of Gothenburg
Sweden
alexander.almer@gu.se

3rd Pierre Gander
Department of Applied IT
University of Gothenburg
Sweden
pierre.gander@gu.se

4th Christian Balkenius
Department of Philosophy
Lund University
Sweden
christian.balkenius@lucs.lu.se

***Abstract*—Among the biggest challenges for researchers of human-robot interaction is imbuing robots with lifelong learning capacities that allow efficient interactions between humans and robots. In order to address this challenge we are developing computational mechanisms for a humanoid robotic agent utilizing both system 1 and system 2-like cognitive processing capabilities. At the core of this processing is a *Social Affective Appraisal* model that allows for vicarious value learning and inference. Using a multi-dimensional reinforcement learning approach the robotic agent learns affective *value-based* functions (system 1). This learning can ground representations of affective relations (predicates) relevant to interacting agents. In this article we discuss the existing theoretical basis for developing our neural network model as a system 1-like process. We also discuss initial ideas for developing system 2-like top-down/generative affective (*semantic relation-based*) processing. The aim of the symbolic-connectionist architectural development is to promote autonomous capabilities in humanoid robots for interacting efficiently/intelligently (recombinant application of learned associations) with humans in changing and challenging environments.**

***Index Terms*—Social Affective Appraisal, Reinforcement Learning, Humanoid Robots, Artificial General Intelligence.**

I. Introduction

Much research focus in Artificial Intelligence in the past ten years has been on Deep Neural Network (aka Deep Learning) applications. These Deep Learning architectures have typically been domain focused, e.g. on pattern (such as image) classification or time series regression/classification problems. By contrast the area of Artificial General Intelligence (strong AI) that concerns autonomous agents capable of adapting to and learning from new interactive scenarios has been less affected by the Deep Learning revolution. One exception to this has been in the area of Deep learning applied reinforcement learning (e.g. through companies like Google-owned DeepMind). Notwithstanding, aspects of interactive intelligence including affective appraisal of social signals and semantic-relational knowledge development remain understudied.

For Artificial agents to be 'generally intelligent' they must be able to interact with, and learn from, their social and non-social environment in efficient ways. For such agents to be useful in the broad domain of human-robot interaction, they must be able to recognize affective states in human (and possibly non-human) agents and learn from them so as to circumvent laborious trial and error individualized learning. Reinforcement learning [1] has provided a tool for allowing artificial agents to learn from experience according to ('good' and 'bad') feedback signals. Such learning with minimal instruction from a human designer permits agents to learn how to achieve goals in their own way according to their own physical constraints. For over 100 years the importance of reinforcing signals has been recognized by animal learning researchers. It has also been at the heart of several influential theories of affective and emotion appraisal. [2] and [3], for example, contended that, following [4], [5])[1], dimensions of reward representation, including anticipatory reward acquisition expectation and anticipatory reward omission, could precipitate differential response control. Similarly, Rolls [6], [7] suggested that emotions, as states that facilitate goal (i.e. reward) driven behaviour, can be elicited by primary or secondary rewards and punishers or omission/premature termination thereof. Reward omission/termination is contended to elicit frustrative, including anger, based affective states, for example. [8], [9], [10], and more recently [11], [12], [13], adapted the Rescorla-Wagner [14] reinforcement learning algorithm to show how representation of omission and acquisition functions of reward could predict certain animal learning data beyond that of the basic reinforcement learning algorithm and in some cases circumvent trial and error learning through utilizing components of existing knowledge. Value-based learning has been applied in the social context. [15], for example, in their review, have highlighted two competing hypotheses for how individuals may learn from others' affective or value based states: i) through vicarious learning – individuals may recognize another as a social object but process the value-based activity of that agent as if it were their own, ii) through social value representation – individuals separately represent their own valuations of stimuli/objects in the environment and those of others. See also [16] for a further review of the field.

In this article we present ongoing work into developing neural network control architectures for humanoid robots at the core of which lies a reinforcement learning approach for implementing affective appraisal/processing of social and

[1]Amsel identified a 'frustration effect' of increased behavioural activity following omission of an anticipated reward.

978-1-7281-3891-6/19/$31.00 ©2019 European Union

non-social signals. The article breaks down as follows: In Section II we present the affective appraisal reinforcement learning algorithm; in Section III we discuss its use according to human-human interaction experiments; in Section IV we consider the potential for human-robot interaction variants of the human-human experiments; finally, in Section V we discuss a cognitive architecture, in development, that permits System 1 and System 2-like adaptive control whereby the former provides *value-based* control and the latter provides top-down generative modulation of affective and object categories using *semantic relation-based* control. We suggest that the two systems may interact to promote lifelong learning through recombinant usage of learned associations, as well as predicate and object knowledge.

II. An Affective Computational Model of Emotion and Action Selection

A. Associative Two-Process Theory

The Affective Appraisal model we have developed is based on the Associative Two-Process (ATP) theory of [17]. In reference to standard stimulus-response trial-based learning using pigeons, rats and humans (see [18] for overview) ATP theory identifies Stimulus-Response (S-R) and Stimulus-Outcome Expectancy-Response (S-E-R) routes where E represents an expectation of an outcome. ATP theory is not unique in positing a role for outcome expectancies in learning and behaviour (see [19], [20]) but through extensive research through the lens of this theoretical perspective the role of outcome expectancies in decision making has been highlighted.

ATP theory indicates that the outcome expectancy route is formed according to two associatively learned components. Firstly there are S-E associations – Pavlovian associations – and secondly there are E-R associations whereby outcome expectations can substitute for, compete with, or facilitate, the external stimulus in guiding instrumental responding. The functional division of S-R learning into two processes has been validated by use of transfer-of-control paradigms wherein the original, learned S-E and E-R contingencies are experimentally manipulated leading to testable hypotheses concerning the pattern of initial responding to these new contingencies (see [21] for original experiment and [12] for model of the data).

By way of example, figure 1 schematizes a transfer-of-control scenario. As is typical for the paradigm, there are three phases. Each of these phases consists of a number of independent trials for learning: presentations of a stimulus, response options, and then a non-negative outcome if the correct response is chosen.

The phases break down as follows: Firstly, there is an instrumental learning phase where the two components (S-E and E-R) of the goal-directed route can be learned as well as the S-R route. Secondly, a Pavlovian (contingency change) learning phase is presented where new S-E associations are made. Finally, a second instrumental phase is utilized, which uses previously experienced stimuli and responses but introduces novel S-R pairings. This serves as a test of transfer of the knowledge of the components (S-E and E-R) learned in the first two phases that provide the relevant building blocks for the S-E-R process to select the correct response in phase 2.

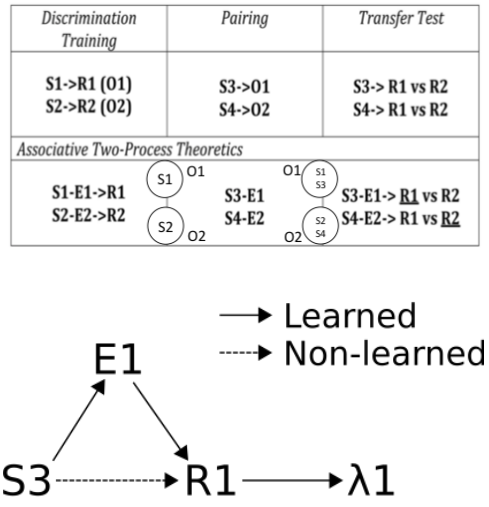

Fig. 1. Transfer-of-control paradigm. The conditioning consists of three phases: Phase 1 (Discrimination Training) – an initial instrumental phase where different stimulus-response (S-R) pairings (S1-R1, S2-R2) yield different outcomes (O1, O2); Phase 2 (Pairing) – a Pavlovian learning phase where new stimuli are presented and associated with previously experienced outcomes; Phase 3 (Transfer Test) – an instrumental transfer phase where the stimuli from phase 2 are re-presented as are the response options from Phase 1. ATP theory predicts that responding in the transfer test (phase 3) will be based on already existing S-E and E-R associations learned from the first two phases where the theorized preferred selections (underlined Rs) are shown in the top diagram and the S3→R1 choice process is schematized on the bottom (adapted from [18]).

In the specific transfer-of-control example given in figure 1 (top), over the first two phases outcomes (O1 and O2) are common to S1 and S3, and S2 and S4, respectively (given that in phase 1 the correct responses are made to obtain those outcomes). As a result, when phase 3 (transfer test) occurs, since the animal/human has learned S1 and S3 according to the same outcome (O1) that is, it has formed S1-E1 and S3-E1 associations S3 automatically cues the response associated with E1 (learned in phase 1), in this case E1 substituting for the external stimulus. No new learning is required for this in spite of the fact that the subject has not been exposed to the particular S-R pairing (S3-R1) previously.

ATP postulates, therefore, that by way of a (dual-route) structured learning process, a type of transitive inference is possible to find correct responses in the test phase without the requirement of learning. S3-R1 associations have not been learned at the beginning of the test phase, but previous experience allows for a transitive performance of the form A→C (S3-R1) derived from A→B (S3-E1), B→C (E1-R1).

The transfer-of-control problem does not entail a designed transitive inference problem but animals and humans appear to resolve this problem by utilizing internal 'hidden' stimuli (expectancy states) through which inference can be made.

B. Affective Associative Two-Process Modelling

The Affective Associative Two-Process model that we have developed [11], [12], [13] merges Associative Mediational Theory [2], [3] and Associative Two-Process theory [17] and is depicted as a feedforward neural network in figure 2. It does so by modelling the differential expectancies of ATP in terms of differential reinforcement outcomes. In such cases, differential outcomes can take the form of differential reinforcement magnitudes [21] or differential omission rates/probabilities (as studied by [2], [3]). The S-R route (horizontal arrow at the bottom of the figure) provides the relation to be inferred in the absence of explicit learning of this association. The connections between successive layers provides the means for inference (when associatively learned). This process implements the S-E and E-R route and occurs as follows: i. the omission and magnitude value dimensions of the external stimuli (S1, S2, etc.) are learned and processed, ii. these values are input into affective value states ('optimistic' reward acquisition inputs and 'pessimistic' reward omission inputs) and are non-linearly transformed (via differentially parameterized logistic functions) so at to allow for 'categorized' outputs to iii. form associations with responses. This categorization disambiguates the control that affective states can have over responding. In this model, the E (expectancy) component can thus be seen as having two stages: i. value dimension computation, ii. affective value computation. The model, as it builds on, and can collapse to, the [10] model (and in turn that of [14]).

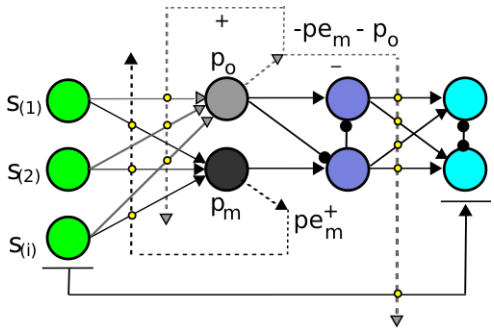

Fig. 2. Neural network model of the theorized Affective-Associative Two-Process (mathematical details found in [13]). Green nodes provide stimulus inputs to value nodes (omission node = grey, magnitude node = black). These nodes provide input to affective nodes (blue) whose activations correspond to 'pessimistic' (upper) and 'optimistic' (lower) values. These nodes in turn can be associatively linked to a response layer (cyan nodes). Connections from stimuli (green nodes) to responses (cyan nodes) constitute the second process of the associative two-process. Associative links are updated by prediction error signals. Key: pe_m = magnitude prediction error, p_m = magnitude prediction, p_o = omission prediction.

III. Vicarious Value Learning in Human-Human Interaction

In spite of the numerous forms of transfer-of-control experimental set up and the consistent result found in animals and humans performing as individual subjects, to the best of the authors' knowledge, such training procedures have yet to be applied to human (or animal) participants in a social context. In [16] we suggested that viewing the Pavlovian (second) stage as a form of passive observation in a social context would allow to test hypotheses as to whether participants can transfer knowledge learned from their own instrumental experience (phase 1) and also knowledge learned within the passive social context (phase 2) to the standard test phase (phase 3). Drawing on a review of how value systems are represented for self and other in joint activities [15], we hypothesized that subjects would be able to vicariously learn the value observed for the other where the context is non-competitive. This is to say that participants experience (the presentation of) other's stimuli and outcomes as if they were their own and update their value function accordingly.

In [22] we report the findings of two experiments that adapted the transfer-of-control set up (figure 1, top) to a social context. In both experiments subjects were required to sit in front of a monitor and over a number of independent trials click on one of two option tabs (responses) following the presentation of a number of different images (stimuli). Figure 3 provides a trial progression diagram for experiments 1 and 2 in phase 1 (instrumental learning phase). Subjects are required through trial and error learning to match the presented stimulus to the correct response option (1 or 2) in order to achieve a reward outcome (high or low points), the value of which attaches differential outcome expectations to the different stimuli (two in phase 1, four in phase 2 and 3) according to Associative Two-Process theory. The experimental condition entails consistent reward value attached to the correct stimulus-response pairs while the control condition (non-differential outcomes condition) randomizes the rewarding outcomes. Controls were in place for various order effects. Phase 3 was much the same as phase 1 except there were four different stimuli (presented in phase 2).

While phase 2 in Experiment 1 entails no instrumental component to the trials – no response options are presented – and only stimuli and outcomes are presented (permitting Pavlovian learning), in Experiment 2 outcomes are no longer presented but instead continuous video footage of an actress culminating in a 'happy' or 'frustrated' facial expression (as validated using Noldus' FaceReader [23]) – see figure 4. In Experiment 1 subjects were told to learn from the videoed performance of a fellow participant (whose face was absent and outcomes shown) on a second monitor. An area of contention for us concerned whether subjects truly perceived the social presence of another (though this was somewhat evaluated by questionnaire feedback). In Experiment 2, by adding the video of the actress and providing an alternative measure of a social presence signal (EEG readings using OpenBCI's Mark IV headset with 8 electrodes worn by all participants), we hoped to clarify better that the transfer-of-control phenomenon found in Experiment 1 could be more reasonably argued to be based on social signals. [22] provides details of the findings including a hypothesize social transfer-of-control effect not reported here.

Our Social Affective Appraisal model (figure 5) predicts

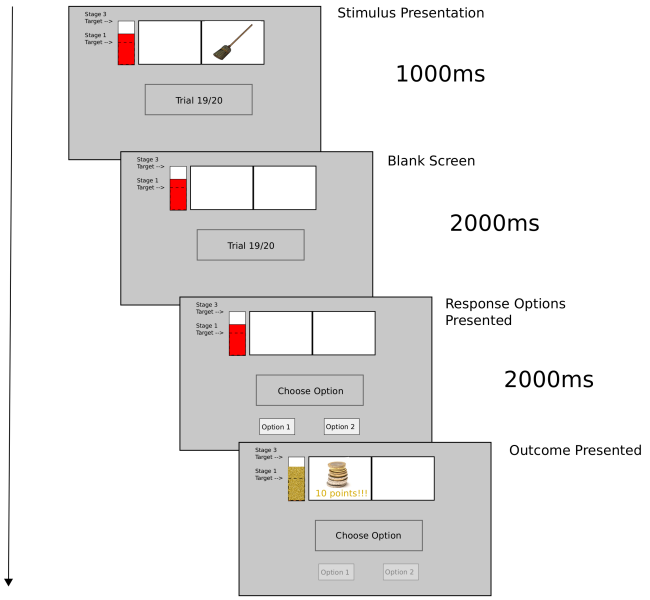

Fig. 3. Phase 1 trial progression diagram for Experiments 1 and 2. Subjects are required to pair the initially presented stimulus (first screenshot right hand panel) to one of two response options (third screenshot) to obtain the rewarding outcome (fourth screenshot). Phase 3 followed the same sequence but with four different stimuli to be associated with the two response options.

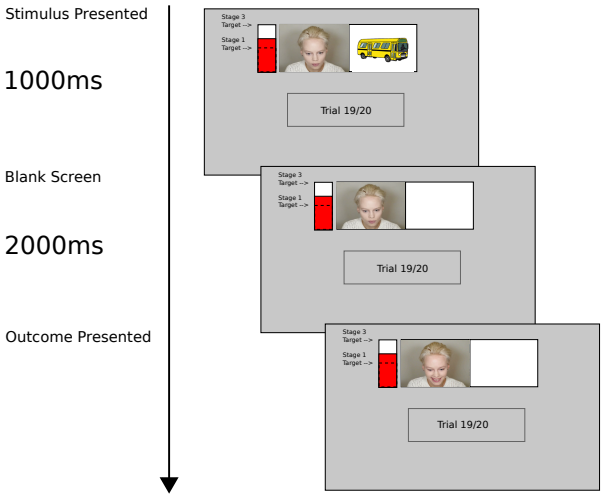

Fig. 4. Phase 2, Experiment 2. Experimental subjects (perceivers) are instructed to learn from the (stooge) subject in the video (left panel). This Pavlovian learning phase differed in Experiment 1 by presenting the actual outcomes, to be paired with a given stimulus, as opposed to emotional expressions of the stooge substituting for the direct outcome information.

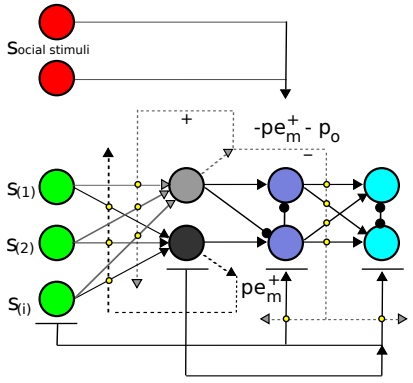

Fig. 5. Social Affective Appraisal Model. The Affective-ATP model presented in figure 2 is adapted here in accordance with [19]. Direct associative links between stimuli and affective states (blue nodes) are enabled as well as fixed connections between the value dimensions (omission, grey node; magnitude, black node) and the responses (cyan nodes). Social stimuli have direct links to the affective nodes of the perceiving agent thereby encoding 'contagion'.

the same findings as for individual transfer-of-control experiments where the value function is assumed to be vicariously learned so that differential outcomes associated with particular stimuli for the perceived other are learned as if they were part of the value function of the perceiving subject. In Experiment 1 this is hypothesized to manifest through direct association of experimental stimuli (S) to differential rewarding outcomes (E). In Experiment 2 this is hypothesized to manifest through indirect association whereby the experimental stimuli (S) are associated with the perceiving subject's outcome expectations (value function) via emotional contagion – the subject perceives the positive/optimistic (pessimistic) emotion, which triggers a representation of that subject's own positive/optimistic (pessimistic) outcome valuation. The latter hypothesis assumes the existence of a direct link between external stimuli representations (S) and the affective states [19] through which mirror neurons may connect. The representation of affective states of others (e.g. expressed through encoding facial action units – [23], [24]) is simplistic here as agents' tendency to experience emotional contagion may be highly context specific. We provide this path to explain, in principle, how subjects could vicariously learn the value functions of others and bring that knowledge to bear on their own response selection (see [16] for more discussion).

IV. Vicarious Value Learning in Human-Robot Interaction

In section III we described the rationale, motivation and performance of subjects on two experiments which hypothesized a particular type of inferential learning based on the Associative Two-Process theory. As a temporal difference learning algorithm which is able to valuate stimuli in spite of inter-stimulus intervals (delays) between external stimulus and reward presentation, the computational model thus far developed we view as suitable for a humanoid robot. The robot can serve thereby to substitute for a human in a number of follow up experiments to the ones described in section III:

1) Controlled human-robot interaction: The robot substi-

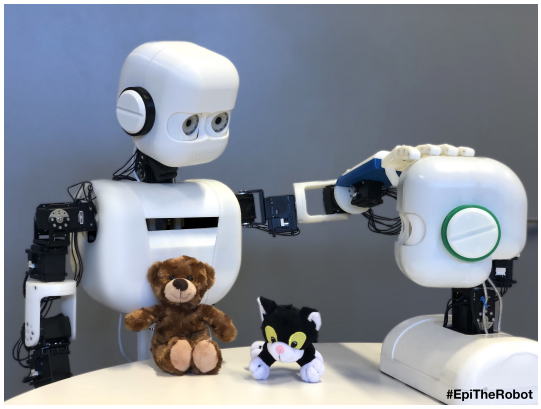

Fig. 6. Epi the robot. Epi is a humanoid robot with actuators (e.g. gripping hands) suitable for carrying out several basic interactive and goal-directed tasks (e.g. stacking). The Epi robot head (right) also has stereoscopic vision and a capacity for pupil dilation and LED expressive capabilities - differential iris colouration and smile/frown to express affective valence and arousal states.

tutes directly for the human stooge in Experiment 2.

2) Controlled robot-human interaction: The robot substitutes for the human subject in Experiment 2.
3) HRI in the wild: Human and robot interact in a more naturalistic version of the transfer-of-control task.

We have considered the humanoid robot Epi (developed at LUCS – Lund University Cognitive Science – group by Christian Balkenius and Birger Johansson) as suitable for the proposed experimental scenarios listed above. As can be seen in figure 6, Epi is usable as head (suitable for scenarios 1 and 2 listed above) or with full upper torso and arms (suitable for scenario 3). Epi is the first robot, to our knowledge, that has pupil dilation capability and can be used for attentional filtering as well as for communication purposes. Epi can signal arousal through pupil dilation in interaction, and has different colouration of its irises as well as a grid of LEDs for signaling 'smiles' so that Epi can signal differential affective states.

In the above-mentioned scenarios, for 1) we envisage a comparatively straightforward substitution of the human stooge for the Epi robot that can use its valence and arousal expressive capabilities to convey responses to differential outcomes. This experimental scenario would provide an interesting validation of the potential to use a robot in such interactive scenarios and would only require the robot to learn/respond in a non-social context (e.g. using the Affective-ATP model depicted in figure 2). In the second experimental scenario, Epi's ability to substitute for the human subject would require an ability to interpret the affective expressions of the stooge. Most rudimentarily this could consist of face identification and perception of the appropriate facial action units encoded by existing software (e.g. using [23]). Ultimately, we will seek to develop our own basic affective recognition system (see accompanying workshop paper for initial work). In scenario 3) we envisage using both the Epi head and full Epi (head with upper torso and arms). The scenario would entail Epi (and the human) carrying out tasks where success on the task requires perceiving stimulus cues and affective expressions or outcomes of the human partner. The actions of the human/Epi partner would not be perceptible to each agent thereby necessitating vicarious learning of each other's value functions as for scenarios 1 and 2 (and Experiments 1 and 2 in section III). Such vicarious and inferential learning helps the agents circumvent a correspondence problem concerning the physical means by which goal-achieving actions are carried out. All the agents need attend to are contextual cues in the environment ('stimuli') and cues concerning how well the task is going (goal outcomes and affective expressions of the other agent). Human-robot interaction exploiting this Social Affective Appraisal mechanism (figure 5) thus permits reduced emphasis on continual learning and instead permits a sort of recombinant learning [25] wherein agents use their existing knowledge and vicarious learning capabilities to make (implicit) inferences, best guesses, in order to bootstrap learning of new stimulus-response relations in the environment. Such lifelong learning skills [25] may be considered critical for autonomous and artificial general intelligence in interactive robots and the study of how this can be successfully transferred to a robot is a major motivation for our ongoing and future work.

V. Discussion: From Feedforward Affective Appraisal to Generative Relational Processing

In the previous sections we have described the application of an affective computational (neural network) model, empirically validated, that through i) dual processes of learning valuates stimuli according to ii) different reinforcement dimensions, and iii) appraises the stimuli according to affective qualities associated with differential responses so as to maximize reward. The dual process and differential valuations of stimuli allow for inferential capabilities wherein agents are able to associate new stimuli with affective states that, in turn, have previously been associated with adaptive responses. This 'inference' enables agents to best guess appropriate responses based on reuse/recombining of existing knowledge without having to directly learn new stimuli-response associations.

Whilst the use of control conditions in Experiment 1 and 2 (section III) provided a means for falsification of our Social Affective Appraisal model variant of the Affective-ATP model, it is also conceivable that human subjects are not using a purely associative learning approach in order to achieve transfer-of-control. In [22] we guarded against simple strategies such as 'process of elimination' by having four stimuli in the third phase of the transfer-of-control conditions. Nevertheless, subjects may still be applying rules that top-down mitigate associative learning. Moreover, we are interested in how System 1 (associative learning based) and System 2 (top-down relational knowledge based) -like processes interact during learning and decision making – [26]. Experimental approaches exist (e,g. [27]) that attempt to tease out different learning approaches according to such systems by comparing decision making behaviour (response selection) of subjects to typical profiles of associative learning based on implicit transitive inference and to that of more rule-based explicit understanding

(where the subject has understood the rules and where correct response selection tends to occur much earlier in trials). In [27] it was found that a subset of subjects would use System 1-like learning and another subset would use System 2-like knowledge validated by comparing individual decision making profiles to subjective reports of task rule comprehension. We will seek to adapt the experimental tasks described in this article to allow for situations that sometimes favour System 1-like approaches (e.g. where response speed is of the essence) and other times favour System 2-like approaches. We will seek to evaluate under what conditions the two type of systems are used and how they might interface as a means to validate and develop a cognitive control architecture, rooted in affective appraisal/generation, that can be used on the Epi robot. Space precludes detailed description of our hypothesized architecture but in [26] we describe a hypothetical interface between a semantic-relational ('symbolic connectionist', adapting LISA [28], and DORA [29]) offline architecture and our own Social Affective Appraisal model further elaborated to incorporate deep parallel distributed processing of object features. Through bottom-up parallel processing social and non-social invariant object representation may be learned in the style of standard (e.g. convolutional neural networks) deep learning architectures. Through top-down local processing of object-predicate relations affective and object (including feature) representations may be generated facilitating learning of objects and predicates in (rule-relevant) propositions (e.g. in predicate calculus terms *frustrates(Stimulus, John)*. Lifelong learning for artificial generally intelligent agents may thereby exploit online learning of object and predicate features which are in turn refined by relational knowledge. Relational knowledge may also be continuously developed through analogical reasoning as in DORA (see [26], [29]) recombining learned objects and predicates to form new relations. Our long-term goal is to imbue in our interactive robot – relation learning, inferential capacities, attention-based rapid object and predicate learning – that are grounded in online, autonomous and homeostatic behaviour [30] and reward but also punishment systems (see also [31], [32]).

REFERENCES

[1] R.S. Sutton, and A.G. Barto, Reinforcement Learning: An introduction, 1st ed. MIT Press, 1998.

[2] J.B. Overmier, and J.A. Lawry, "Pavlovian conditioning and the mediation of behavior," The Psychology of Learning and Motivation, vol. 13, pp. 1–55, 1979.

[3] J. Kruse, and J.B. Overmier, "Anticipation of reward omission as a cue for choice behavior," Learning and Motivation, vol. 13(4), pp. 505–525, 1982.

[4] A. Amsel, "The role of frustrative nonreward in noncontinuous reward situations," Psychol. Bull., vol. 55, pp. 102–119, 1958.

[5] A. Amsel, Frustration theory: an analysis of dispositional learning and memory, Cambridge University Press, Cambridge, 1992.

[6] E.T. Rolls, The brain and emotion. Oxford Univ. Press, Oxford, 1999.

[7] E.T. Rolls, The Brain, Emotion, and Depression. Oxford University Press, 2018.

[8] C. Balkenius, and J. Morén, "Emotional learning: a computational model of the amygdala," Cybern. Syst. Int. J., vol. 32, pp. 611–636, 2001.

[9] J. Morén, Emotion and Learning - A Computational Model of the Amygdala. Lund University Cognitive Studies, 93, 2002.

[10] C. Balkenius, J. Morén, and S. Winberg, "Interactions between motivation, emotion and attention: from biology to robotics," In Proceedings of the Ninth International Conference on Epigenetic Robotics, vol. 149, pp. 25–32). Lund University Cognitive Studies, 2009.

[11] R. Lowe, Y. Sandamirskaya, and E. Billing, "A neural dynamic model of associative two-process theory: The differential outcomes effect and infant development," In 4th International Conference on Development and Learning and on Epigenetic Robotics, pp. 440–447, 2014.

[12] R. Lowe, and E. Billing, "Affective-Associative Two-Process theory: A neural network investigation of adaptive behaviour in differential outcomes training," Adaptive Behavior, vol. 25(1), pp. 5–23, 2017.

[13] R. Lowe, A. Almér, E. Billing, Y. Sandamirskaya, and C. Balkenius, Biological Cybernetics. vol. 111 (365). https://doi.org/10.1007/s00422-017-0730-1, 2017.

[14] R.A. Rescorla, and A.R. Wagner, "A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement," In A.H. Black, and W.F. Prokasy (eds), Classical Conditioning II: Current Research and Theory, New York: Appleton-Century-Crofts, 1972.

[15] C.C. Ruff, and E. Fehr, "The neurobiology of rewards and values in social decision making," Nat. Rev. Neurosci., vol. 15, pp. 549–562. doi: 10.1038/nrn3776, 2014.

[16] R. Lowe, A. Almér, G. Lindblad, P. Gander, J. Michael, and C. Vesper, "Minimalist social-affective value for use in joint action: A neural-computational hypothesis," Frontiers in Computational Neuroscience, vol. 10. Available at: https://doi.org/10.3389/fncom.2016.00088, 2016.

[17] M.A. Trapold, and J.B. Overmier, "The second learning process in instrumental learning," In Classical Conditioning II: Current Research and Theory (pp. 427–452). New York: Appleton-Century-Crofts, 1972.

[18] P. Urcuioli, "Behavioral and associative effects of differential outcomes in discriminating learning," Learning and Behavior, vol. 33(1), pp. 1–21, 2005.

[19] R.N. Cardinal, J.A. Parkinson, J. Hall, and B.J. Everitt, "Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex," Neuroscience & Biobehavioral Reviews, vol. 26(3), pp. 321-352, 2002.

[20] S. de Wit, and A. Dickinson, "Associative theories of goal-directed behaviour: A case for animal-human translational models," Psychological Research PRPF, vol. 73(4), pp. 463–476, 2009.

[21] G.B. Peterson, and M.A. Trapold, "Effects of altering outcome expectancies on pigeons' delayed conditional discrimination performance," Learning and Motivation, vol. 11, pp. 267–288, 1980.

[22] J. Rittmo, R. Carlsson, P. Gander, C. Vesper, and R. Lowe, "'Vicarious Value Learning: Processing Other?s Affect within a Differential Outcomes Transfer of Control Task', in preparation.

[23] Noldus, FaceReader: Tool for automatic analysis of facial expression: Version 6.0. Wageningen, the Netherlands: Noldus Information Technology B.V, 2014.

[24] P. Ekman, W.V. Friesen, J.C. Hager, "Facial action coding system - the manual," Salt Lake City: Research Nexus, 2002.

[25] G. Anthes, "Lifelong learning in artificial neural networks,", Communications of the ACM, Vol. 62 (6), pp. 13–15, 2019.

[26] R. Lowe, A. Almér, and C. Balkenius, "Bridging Connectionism and Relational Cognition through Bi-directional Affective-Associative Processing,", Open Information Science, In Press.

[27] M.J. Frank, J.W. Rudy, W.B. Levy, and R.C. O'Reilly, "When logic fails: Implicit transitive inference in humans," Memory & Cognition, vol. 33(4), pp. 742–750, 2005.

[28] J.E. Hummel, and K.J. Holyoak, "Distributed representations of structure: A theory of analogical access and mapping," Psychological review, vol. 104(3), 427, 1997.

[29] L.A. Doumas, R.G. Morrison, and L.E. Richland, "Individual differences in relational learning and analogical reasoning: A computational model of longitudinal change," Frontiers in Psychology, 9, 2018.

[30] A. Montebelli, R. Lowe, and T. Ziemke, "The cognitive body: from dynamic modulation to anticipation," In Workshop on Anticipatory Behavior in Adaptive Learning Systems, Springer, Berlin, Heidelberg, pp. 132–151, 2008

[31] N. Navarro-Guerrero, R.J. Lowe, and S. Wermter, "Improving robot motor learning with negatively valenced reinforcement signals," Frontiers in neurorobotics, 11, 10, 2017.

[32] C. Li, R. Lowe, and T. Ziemke, "A novel approach to locomotion learning: Actor-Critic architecture using central pattern generators and dynamic motor primitives," Frontiers in neurorobotics, 8, 23, 2014.