2006 International Joint Conference on Neural Networks
Sheraton Vancouver Wall Centre Hotel, Vancouver, BC, Canada
July 16-21, 2006

Multiagent team formation performed by operant learning: an animat approach

D. A. Gutnisky, R. Zelmann, and B. S. Zanutto, *Member, IEEE*

Abstract— **An animat approach to dynamic team formation in a group of distributed robots is studied. The goal is that robots learn to align with the others in order to form a row or a column without having communication among them, just local sensing and a reinforcement signal. The action of the robot is controlled by a biologically plausible neural network model of operant learning. The remarkable performance achieved by the proposed model allows the building of new Artificial Intelligence agents based on neurobiology, psychology and ethology research.**

Index Terms—**Operant behavior, Multiagent System, Neural Networks, Reinforcement Learning**

I. INTRODUCTION

In the last few years, multiagent systems engaged in collective behavior have been of growing interest among Artificial Intelligence researchers[1][2][3][4]. This interest comes from several reasons[5]. First, tasks may be too complex to be accomplished by a single robot, or a better performance can be obtained by performing the same tasks with multiple robots. Second, constructing a group of simple robots can be easier, more flexible, more fault-tolerant and it might have economical benefits compared to using a single powerful robot. Third, even though there are many definitions of cooperation in the robotics literature[6][7][8] there is a general agreement that given a specific task, a multiagent system performing cooperative behavior involves an increase in the total utility of the system by an underlying cooperation mechanism[5]. Last, animals' adaptation and social interaction skills can be borrowed from natural sciences by building autonomous robots inspired by biology, ethology, psychology and neurobiology research.

In this work, we have studied the problem of achieving global behavior in a group of simulated agents with only local sensing and a reinforcement signal. The global goal is to establish and maintain a specific geometrical shape (a row or a column). Additionally, each robot receives a reinforcement signal depending only on its individual action, and not on specific global situations. Thus, the proposed algorithm allows the accomplishment of the task, learning by trial-and-error and without having a centralized control or explicit communication. Moreover, as in Mataric proposal[9], there are no leaders and agents are not specifically designed to cooperate, instead cooperation is implicit by the design of the task and the reinforcement signals.

Even though, the fact that there have been a variety of approaches to create global behavior in multiagent systems[1][10][11][12], little research has been made to incorporate biologically inspired models of animal learning to solve coordination tasks in mobile robots. We studied how a model that learns to associate stimuli and rewards can learn to achieve global-level formation coordination without communicating or having global knowledge of the other robots' position.

Algorithms proposed to solve coordination problems are not usually restricted by biological constraints. Touretzky and Saksida[13] said that mobile robots trained by methods such as Q-Learning[14][15] have not come close to matching the sophistication, versatility and adaptation of the animals. They suggest that closer attention be paid to animal training literature and a serious attempt to model the effects described there may yield benefits of immediate value to robot learning researchers, and also provide a new computationally oriented perspective on animal learning.

On the other hand, simple artificial animals ('animats'), which operate as autonomous, adaptive robots in the real-world, can serve both as models of biology behavior and as a radical alternative to conventional methods of designing intelligent systems[16]. McFarland and Bösser [17] pointed out that it has been recently realized, that robots might better be designed along the zoological lines of primitive animals than along the traditional lines of autocratic control. Dean[18] said that the anim at approach is the most recent attempt to simulate the adaptive behavior characteristic of animals as well as the acquisition of this competence. Moreover, he suggested that both in analysis and design, the animat approach borrows heavily from ethology, psychology, neurobiology and evolutionary biology.

Animat studies can also provide models of emergent behavior in biological systems[19], also Dean[18] pointed out that: “for AI and robotics researchers, understanding the mechanisms behind adaptive behavior is secondary to creating them, but natural scientists can hope for tools and concepts to aid understanding of biological systems”. It is clear that if we knew how animals control their behavior this might provide us new ideas about how to make robots do it.

We studied the ability of the model of operant behavior presented in [20][21] to learn how to make formations with only local information. The hypothesis of this adaptive neural network model of aversive and appetitive behavior comes from theories of behavior, experimental results in animals and neurobiological evidence. Although the investigation in adaptive neuronal networks is not a new area, most of it has been used to explain behavior experiments in animals and they have not been used to control autonomous robots.

The model of operant learning has not been designed specifically to solve this task. It was originally proposed as a theory of operant learning to explain formally most of the relevant psychological experiments of appetitive and aversive stimulus in animals. In this way, the purpose of this paper is to approach some autonomous robot problems from models inspired in psychology, biology and neuroscience. Moreover, we have shown that this operant learning model is able to learn how to cooperate in the Prisoner Dilemma game [22], and it outperforms Q-Learning in an obstacle avoidance task [23]. These results suggested us that the same model could be able to learn how to cooperate in simple multiagent formation task.

I. Operant Conditioning

Psychologists have identified classical and operant conditioning as two primary forms of learning that enable animals to acquire relevant characteristics of their environment in order to get reinforcements or to avoid punishments. Classical conditioning is an open-loop experiment procedure where the controlled stimuli delivered by the experimenter are not contingent on the animal’s behavior. The learning occurs by repeated association of a conditioned stimulus (CS) with an unconditioned stimulus (US) that elicits an unconditioned response (UCR). For example, in Pavlov’s experiment [24], the dog hears a bell (the CS) and after a short time, a piece of meat is presented (US) that elicits a salivation UCR. After repeating the experiment several times, the sole presentation of the CS elicits the salivation response (conditioned response or CR).

On the other hand, operant conditioning is a closed-loop experiment procedure, in the sense that stimuli received by the animal are contingent on its behavior. The animal learns to perform the actions that led to reward and to avoid the actions to the ones that led to punishment. For example, a rat can be trained to press a key when he sees a red light as CS, in order to receive a food reward (US).

II. Model of Operant Conditioning

Zanutto and Lew[20][21] presented a neural network model that, based on biological plausible hypotheses, explains the relevant features of operant conditioning for appetitive and aversive stimuli[25][26] (for further details about the model, its psychological, neurobiological and anatomical bases, see [20][21][25]).

Behavioral experiments suggest that learning is driven by changes in the expectation about the future salient events, mainly reward and punishment. In operant and classical conditioning, the conditioned stimulus (*CS*) anticipates the unconditioned stimulus (*US*). Rescorla and Wagner[27] proposed that animals learn comparing what they expect from a given situation and what actually happens. As Staddon[28] has pointed out, animals act as the *CS* allows them to elaborate an expectation or prediction of the unconditioned stimulus. Furthermore, there are neural substrates of prediction and reward, such as the involvement of dopamine neurons of the ventral tegmental area (VTA) and sustantia nigra, identified with the processing of prediction and reward [29].

The model is shown in Fig. 1. The inputs to the model are: all the conditioned stimuli (*CS*s), the unconditioned stimulus (*US*) and the outputs are all the possible responses of the animal (*Rs*). The network has three basic functional blocks. The stimuli and response traces, the prediction neuron, and the response neurons.

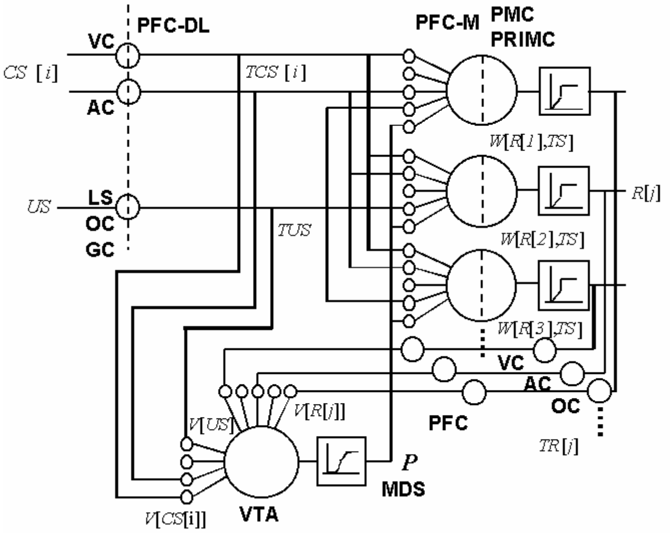

Fig. 1. The Neural network model of operant learning. There is an artificial neuron computing the prediction (*P*), and one for each response *R[j]*. VC: visual cortex; AC: auditory cortex; OC: olfatory cortex; GC: gustatory cortex; LS: limbic system; PFC: prefrontal cortex; PMC: premotor cortex; PRIMC: primary motor cortex; MDS: mesocortical dopaminergic system; VTA: ventral tegmental area. The traces (*TCS*[*i*]) representing the short-term memory of the conditioned stimuli are computed in the PFC-M[30], the unconditioned stimulus (*TUS*), and the responses (*TR*[*j*]) are inputs of the neurons computing *P* and *R[j]*. The synaptic weights *V*[CS[*i*]], *V*[R[*j*]] and *V*[*US*] represent the associations between the inputs and *P*. The synaptic weights *W*[*R[j],TS*], are associations between *P*, *TCS*[*i*] and *TUS*, and the PFC-L[31].

In the model, the prediction neuron has all the stimuli and the responses traces as inputs. The synaptic weights are modified by the Rescorla-Wagner rule, except the US that remains fixed. The response neurons have all the CS and US traces and the prediction as inputs. If it exceeds a certain threshold, the learning in the responses will be hebbian[32] in the appetitive, and antihebbian in the aversive case. The reverse if the prediction is under the threshold. When one of the response neurons exceeds a certain level, the associated response is executed. The model equations are provided in the appendix.

III. Implementation

We developed an application to simulate an environment where a group of homogenous robots move freely, and receive rewards depending on specific local conditions. The environment is a grid of 5 by 5 and no agent can occupy the same cell. Each of the five robots has five different signals as inputs: North, South, East, West and “Double Union”. The first four indicate the direction that has more number of robots. The signal “Double Union” is received when an agent has in the same line two other agents, one on each side of it, or when it is besides a wall and another agent is adjacent to it (see Fig. 2). Robots have five possible actions, to move in either direction or to reaffirm the “Double Union”, that is, to remain in the same position. In each time step robots decide to take an action depending on the allowed moves, (i.e. they cannot move to an occupied cell, and the action “Double Union” can be executed only if its signal is presented). Agents always have to move, except when they are blocked, or when they have the possibility of taking the “Double Union” action.

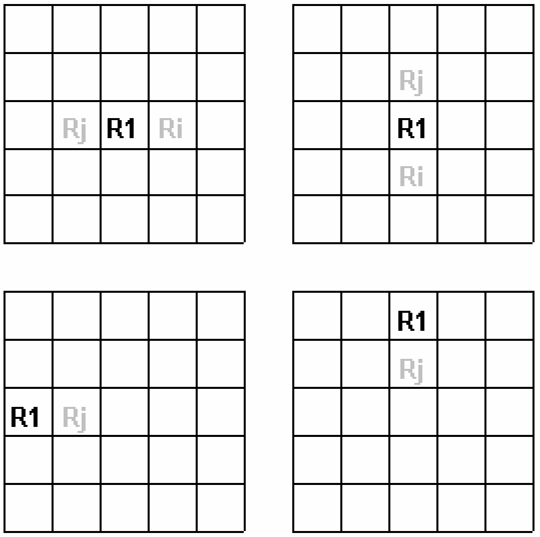

Fig. 2. The signal “Double Union” is presented to the robot 1 when it has two other robots adjacent to it, or when it is adjacent to a wall and to another robot. If in this situation robot 1 takes the action to reaffirm the “Double Union”, it will receive a reward, and stay in the same position.

Agents move in turns (the order is selected at random). If after a move the agent lies on a cell that is adjacent to an occupied cell, the first agent receives a reward (an US in the model). However, there is no reward if it remains in the same position. The other possibility to obtain a reward is when an agent has a signal of “Double Union”, and performs the correspondent action. The only way that all the robots make a stable formation is that all of them take the action of “Double Union” when they have adjacent robots in both of their sides.

IV. Experimental Results

Each experiment starts by putting the five agents distributed in the grid at random. An step consists of one action of each of the robots. The experiment finishes when the five robots form a row or a column and it is maintained during 10 steps, and a new experiment begins by again placing them randomly in the grid. To get an average of the performance, we made 50 repetitions and averaged the time to make a formation in function of the steps performed.

We compared the performance of the operant model with other two types of agents: random and programmed. Random agents move randomly, but when the “Double Union” signal is present, they execute the action that reaffirms the “Double Union” (otherwise it would be very improbable that they could maintain a formation for 10 consecutive steps). Programmed agents always move in the direction where there are most agents.

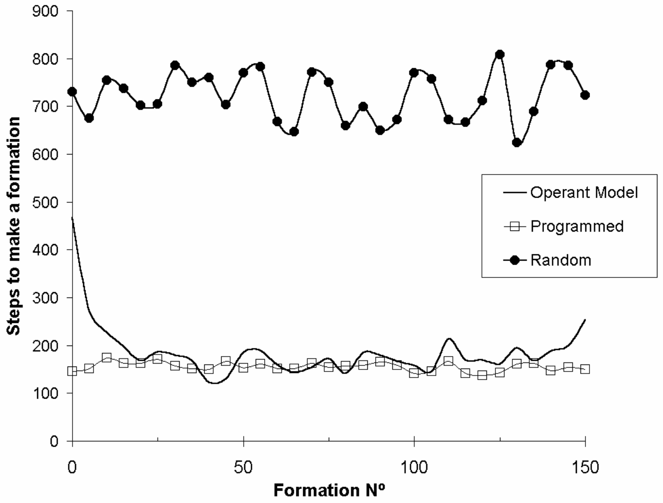

Fig 3. Average steps performed per agent in function of the number of formations completed.

Fig. 3 shows the amount of steps required to make a formation for each type of agent. It can be seen that at the beginning the operant model performance is near the random one, but after making a few formations, it achieves almost the same performance as the programmed agent.

This improvement occurs because operant model agents gradually learn to get more rewards (see Fig. 4) and as a consequence, they are able to achieve the collective property of making formations. Fig. 4 shows the amount of rewards that all agents received per step. Random and Programmed agents received rewards under the same conditions as the operant model.

The gradual rise in rewards obtained by operant agents is correlated with Fig. 5 where the percentage of correct responses in function of the number of steps is shown. Here, a correct response is considered when the agent moves in the direction of the activated CS. Operant agents are able to achieve almost a 90% of correct response, with a performance similar to the maximum one (considering the restrictions of the environment). This shows that the emergent property of collective behavior achieved is robust from mistakes performed by individual agents.

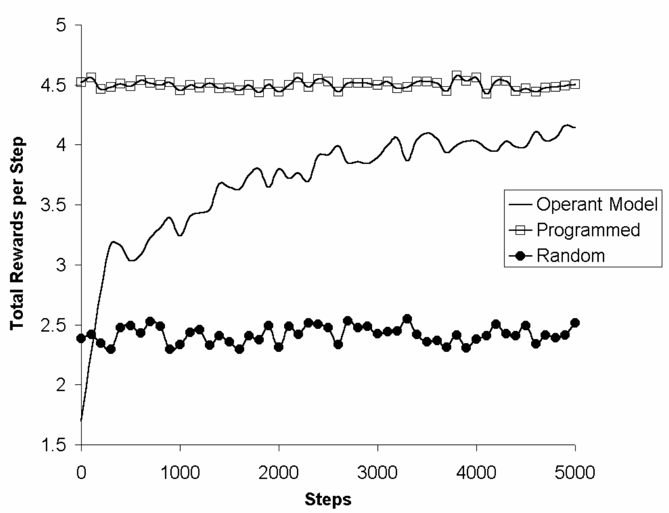

Fig 4. Average of the total rewards obtained by the five agents in function of the step number.

.

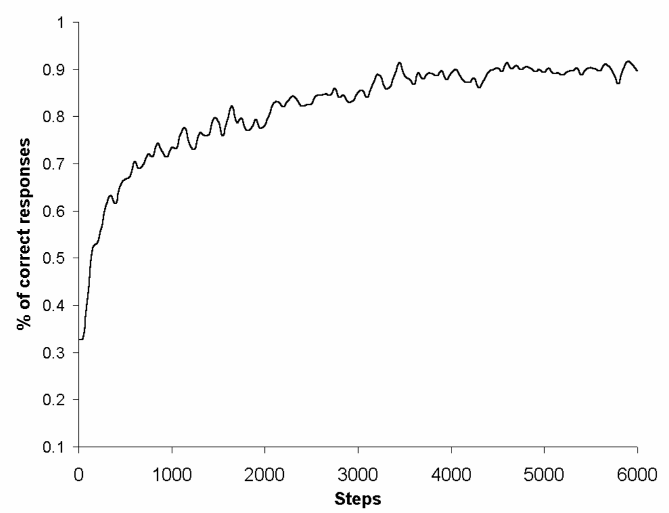

Fig 5. Percentage of correct response made by operant learning agents in function of the step number.

Fig. 6 shows the probability of making a formation in function of the number of moves done. It can be seen that operant and programmed agents are almost indistinguishable.

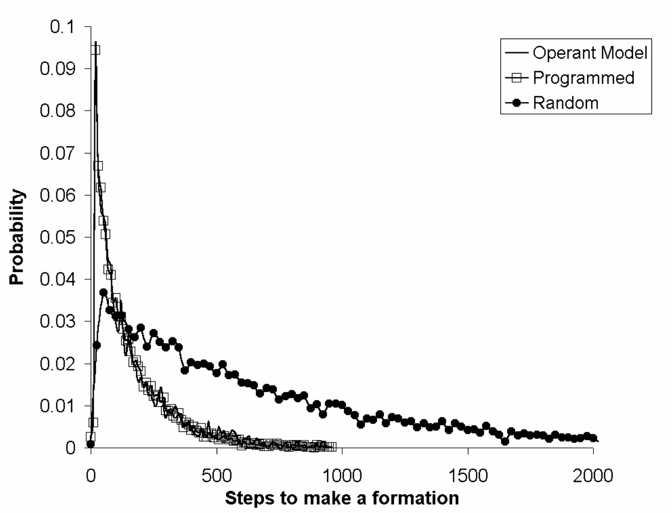

Fig 6. Estimated probability density function of the steps needed to achieve a formation.

Fig. 7 was made by counting the amount of times that agents occupied each cell. We made graphs for the three types of agents, and in two blocks of trials. In the first block we counted the time each cell was occupied between the start of the experiment and 10% of the time that was required to make the formation. The second block was between 90% and 100% of the formation time. It can be seen that programmed and operant learning agents tend to move around the center, and in the horizontal and vertical line that cross it. In contrast, random agents tend to move in a more disperse way. In the case of programmed and operant learning agents, there is almost no difference in the occupancy graphs between 0-10% and 90-100% of time to formation. This means that agents can rapidly join and stay close to each other, but most of the time is employed to rearrange in such way that the formation is completed.

Fig. 8 shows the number of steps needed to make a formation for each type of agent with their correspondent error bars. The figure clearly shows that there is almost no difference in the performance achieved by the programmed and operant learning model.

V. Discussion

Most of the work in multiagent systems has assumed that cooperation is explicitly designed into the system[5]. On the other side, another approach is to study how cooperation can arise from selfish agents. Kube & Zhang[33] pointed out that: “Designing autonomous robots that accomplish useful tasks is a challenging and still elusive goal of scientific research. The main hypothesis of the approach lies on the hope that such a population of machines will achieve a higher level of competence due to an emergent property of the system making it more than just the sum of the parts”.

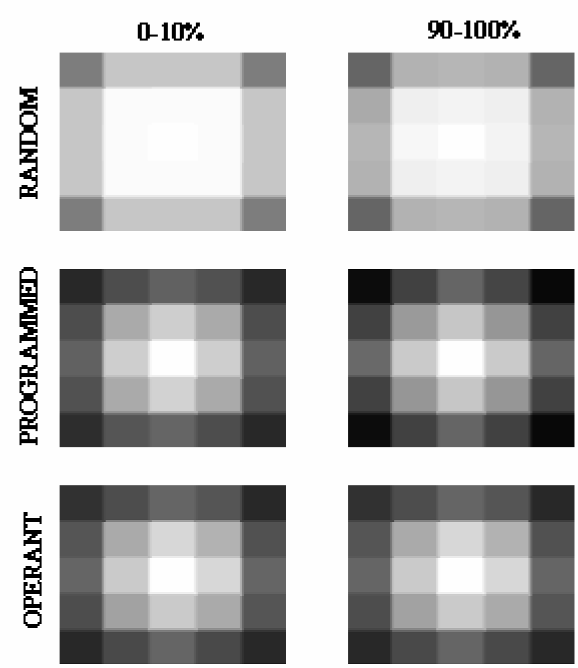

Fig 7. Density graph of occupied cells of the five agents. The first column is the average occupancy density graph for each cell between 0 to 10% of the time needed to make a formation (from top to down, random, programmed and operant learning agents). The second column is the average occupancy density graph between 90 and 100% of the time needed to make a formation. White colored cells represent cells that are occupied longer than black ones.

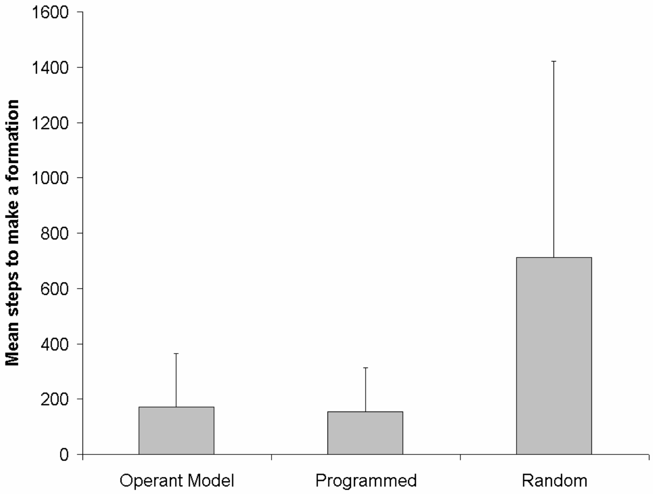

Fig 8. Mean steps needed to make a formation for each of the three types of simulated agents. Error bars represent s.d.

In nature such emergent property is found in many cases [34][35][36]. Despite multiagent cooperation, and specifically team formations having a growing interest from the Artificial Intelligence research community[1][10][11][12], little study has been made in building robots with models based on the neural mechanism that allows animals to learn. The architecture of the neural network of the operant model we presented relies on a neuron that predicts the reinforcement, where its main objective is to control the learning of the response neurons in order to take the actions that are more likely to be linked to a reward in a specific situation.

We showed that with operant capacities a group of robots is able to make formations. Despite its simplicity, the model is able to explain many characteristical behaviors observed in animals and, as the present work showed, it is suitable to control autonomous robots in order to have the emergent property of making formations. With simple hypotheses borrowed from psychological and neurobiological experiments, the proposed operant model is a clear example of how research in control of autonomous robots can be benefited from research in psychology and neuroscience.

VI. Conclusion

We studied how to make formations with a group of distributed robots without global information. The actions of the robots were controlled by a biological plausible neural network model of operant learning. This model, built with the aim of explaining real behavioral experiments in animals, can solve complex tasks in different environments from those where animal experiments are made. Thus, the results obtained suggest a further study of the feasibility of using operant learning as a method for controlling animats.

Appendix

In this purposed model all stimuli generates a short-term memory given by:

$$TS_n = TS_{n-1} \cdot (1-\alpha) + \alpha \cdot S_{n-1}$$

$$TR[i]_n = TR[i]_{n-1} \cdot (1-\beta_R) + \varepsilon \cdot (1 - TR[i]_{n-1}) \cdot R[i]_n$$

Where S is: *CS*[*1*], *CS*[*2*]…, *or US…;* n is the time unit. Here Greek letters represents constants. When S_n is greater than 0, α=ε, and when S_n=0, α=β.

The prediction (P) is calculated as:

$$P_n = \xi \Big/ \left(1 + e^{-\upsilon(X_n - \sigma)}\right)$$

$$X_n = V[US]_n \cdot TUS_n + \sum_{i=1}^{NCS} V[i]_n \cdot TCS[i]_n + \sum_{j=1}^{NR} V[j+NCS]_n \cdot TR[j]_n$$

Where V represents the synaptic weights of the neuron that computes the prediction, these are the associations between the inputs and P.

i: conditioned stimulus index and j: responses index

NCS: is the number of conditioned stimuli,

NR: is the number of the possible responses.

The inputs to response neurons are: P, *TCS*s and TUS. The output of the response neurons ($R[j]$) is calculated as:

$$R[j]_n = f_2(Y[j]_n)$$

$$Y[j]_n = W[j][P]_n \cdot P_n + W[j][US]_n \cdot TUS_n + \sum_{i=1}^{NCS} W[j][i]_n \cdot TCS_n[i] + noise(n)$$

$$f_2 = 0 \; if \; Y[j]_n < 0; \; f_2 = 1 \; if \; Y[j]_n > \mu; \; else \; f_2 = Y[j]_n$$

Where W represents the synaptic weights of the neuron that computes $R[j]$, and *noise(n)*: white noise (amplitude=1/32).

i: conditioned stimulus index

μ : is the threshold.

The animal executes the response *R[j]* when *Y[j]* is greater than μ..

The synaptic weights (*V*) of the prediction neuron (bounded between –1 and 1) are calculated as follows:

$$V[S]_n = (2/(1 + e^{-\kappa \cdot VX[S]_n})) - 1;$$

$$VX[S]_n = VX[S]_{n-1} + \eta \cdot TS_n \cdot (US_n - X_n)$$

Where η can take two values, η_i controls the rise and η_d controls the decay of *VX[S]*$_n$. Values are bounded between –10 and 10. The synaptic weight *V*[*US*] of the prediction neuron is fixed to 0.1.

The synaptic weights of response neurons *j* are calculated by Hebbian or anti-Hebbian learning [32] depending on *P* as follows:

$$W[j][q]_n = \psi \cdot W[j][q]_{n-1} + \phi \cdot TQ_n \cdot TR_n[j] \cdot \Omega$$

Where *TQ* is: *P, TUS or TCSi*] and the respective index q is: *P, US or i*; the first term includes a momentum 16]. If *P* < λ then $\Omega = -\lambda_-$, otherwise $\Omega = \lambda_+$.

To simulate the animals exploratory behavior at the beginning of the experiment, the probability of generating random responses (*Pb*) decreases exponentially from a starting value (φ).

$$Pb_n = Pb_{n-1} \cdot \omega$$

In each trial a CS is presented to the model, then it makes a response, and depending on the CS and the response performed, it will receive the reward or not. Each trial consisted of 80 time steps of the algorithm. The CS was presented during 20 time steps, the US during 10 time steps, and the response persisted for 5 time steps.

The constants are: β=0.005, β_R=0.025, δ=.0001, ε=0.25, ϕ_a=0.01, ϕ_{pav}=0.07, γ=5, η_i=0.02, η_d=0.01, κ=0.2, λ_+=0.6, λ_-=0.06 μ=0.35, ρ=0.0003, σ=0.4, τ=0.45, υ=10, ω=0.9997, ξ=0.6, ψ=.9985, χ=20, φ=0.99, ζ=10, noise=0.03125, V_{US}=0.1. CS intensity=1, US intensity=6, Response intensity=1. There is also a context stimulus that it is presented all the time, and its intensity is of 0.15.

References

[1] J. Fredslund & M. J. Mataric, "A General, Local Algorithm for Robot Formations," *IEEE Transactions on Robotics and Automation,* special issue on Advances in Multi-Robot Systems, vol. 18, no, 5, pp. 837-846, Oct. 2002.

[2] C. Grinton, L. Sonenberg, and L. Sterling, "Exploring agent cooperation: studies with a simple pursuit game," in *Advanced Topics in Artificial Intelligence*, Berlin, 1997, pp. 96-105.

[3] M. Wooldridge, N. R. Jennings, "The Cooperative Problem-Solving Process," *Journal of Logic Computation*, vol. 9, no. 4, pp. 563-592, 1999.

[4] P. R. Cohen, H. R. Levesque, and I. Smith, "On Team Formation," in *Contemporary Action Theory*, J. Hintikka and R. Tuomela, Eds. Synthese, 1997.

[5] Y. Uny Cao, A. S. Fukunaga, and A. B. Kahng, "Cooperative mobile robotics: Antecedents and directions," *Autonomous Robots*, vol. 4, pp. 7-27, 1997.

[6] D. Barnes & J. Gray. Behavior synthesis for cooperant mobile robot control. In International Conference on Control, pp 1135-1140, 1991.

[7] M. Mataric. Interaction and Intelligent Behavior. PhD thesis, MIT, EECS, May 1994.

[8] S. Premvuti, and S. Yuta, "Consideration on the cooperation of multiple autonomous mobile robots," in *IEEE/RSJ IROS*, 1990, pp 59-63.

[9] M. J. Mataric, "Designing emergent behaviors: From local interactions to collective intelligence," in *From Animals to Animats 2, Second International Conference on Simulation of Adaptive Behavior (SAB92)*, 1992, pp. 432-441.

[10] K. Sugihara, and I. Suzuki, "Distributed algorithms for formation of geometric patterns with many mobile robots," *Journal of Robotic Systems,* vol. 13, no. 3, pp. 127-139, 1996.

[11] B. B. Werger, "Cooperation Without Deliberation: A Minimal Behavior-based Approach to Multi-Robot Teams," *Artificial Intelligence,* vol. 110, pp. 293-320, 1999.

[12] T. Blach, and R. C. Arkin, "Behavior-based Formation Control for Multi-robot Teams," IEEE Transactions on Robotics and Automation, vol. 14, no. 6, pp. 926-939, Dec. 1998.

[13] D. S. Touretzky, and M. L. Saksida, "Operant conditioning in skinnerbots," *Adaptive Behavior,* vol. 5, pp. 219-247, 1997.

[14] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement Learning: A survey," *Journal of Artificial Intelligence Research,* vol. 4, pp. 237-285, 1996.

[15] Watkins, C. J. C. H. (1989). Learning with Delayed Rewards. Ph.D. dissertation, Cambridge University, Psychology Department.

[16] R. I. Damper, R. L. B. French, and T. W. Scutt, "ARBIB: An autonomous robot based on inspiration from biology," *Robotics and Autonomous Systems*, vol. 31, no. 4, pp. 247-274, 2000.

[17] D. McFarland, and T. Bösser, *Intelligent Behavior in Animals and Robots*. Cambridge, MA: Bradford Books / MIT Press, 1993.

[18] J. Dean, "Animats and what they can tell us," *Trends in Cognitive Sciences*, vol. 2, no. 2, pp. 60-67, 1998.

[19] L. Steels, "Towards a theory of emergent functionality," in *From Animals to Animats: Proceedings of the First International Conference on Simulation of Adaptive Behavior*, 1991, pp 451–461.

[20] S. E. Lew, C. Wedemeyer, and B. S. Zanutto, "Role of unconditioned stimulus prediction in the operant learning: a Neural network model," in *Proc. of the IEEE* Intl. Joint Conf. on Neural Networks, Washington, DC, 2001, pp. 331-336.

[21] B. S. Zanutto, and S. Lew, "A neural network model of aversive behavior," in *Proceedings of the IASTED Neural Networks NN'2000*, Zürich, 2000, pp. 118-123.

[22] D. A. Gutnisky, and B. S. Zanutto, "Cooperation in the iterated prisoner's dilemma is learned by operant conditioning mechanisms," *Artificial Life,* vol. 10, no. 4, pp 433-461, 2004.

[23] D. A. Gutnisky, and B. S. Zanutto, "Learning obstacle avoidance with an operant behavior model," *Artificial Life,* vol. 10, no. 1, pp 65-81, 2004.

[24] I. P. Pavlov, *Conditioned reflexes*. Oxford: Oxford University Press, 1927.

[25] N. Schmajuk, and B. S. Zanutto, "Escape, avoidance and imitation: a neural network approach," *Adaptive Behavior,* vol. 6, pp. 63-129, 1997.

[26] N. Schmajuk, D. Urry, and B. S. Zanutto, "The frightening complexity of avoidance: a neural network approach," in *Models of Action*: *Mechanisms for Adaptive Behavior*, C. Wynne & J. Staddon, Eds. Hillsdale, NJ: Eribaum, 1998.

[27] R. A. Rescorla, and A. R. Wagner, "A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement," in *Classical conditioning II: Current research and theory*, A. H. Black & W. F. Prokasy, Eds. New York: Appleton-Century-Crofts, 1972.

[28] J. E. R. Staddon, *Adaptive behavior and learning*. Cambridge: Cambridge University Press, 1983.

[29] W. Schultz, P. Dayan, and R. Montague, "A neural substrate of prediction and reward," *Science,* vol. 275, pp. 1593-1598, 1997.

[30] P. S. Goldman-Rakic, "Circuitry of primate prefrontal cortex and regulation of behavior by representational memory," in *Handbook of Physiology: The Nervous System*, F. Plum, Ed. Bethesda, MD: American Physiology Society,1987, pp. 373–417.

[31] J. F. Bates, and P. S. Goldman-Rakic, "Prefrontal connections of medial motor areas in the rhesus monkey," *Journal of Comparative Neurobiology*, vol. 336, pp. 211–28, 1993.

[32] D. O. Hebb, *The organization of Behavior: A Neuropsychological Theory.* New York: Wiley, 1949.

[33] C. R. Kube, and H. Z. Zhang, "Collective robotics: from social insects to robots," *Adaptive Behavior*, vol. 2, no. 2, pp. 189-218, 1994.

[34] E. O. Wilson, *The insect societies.* Harvard University Press, 1971.

[35] R. Trivers, *Social Evolution.* Menlo Park, CA: Benjamin Cummings, 1985.

[36] L. A. Dugatkin, *Cooperation among Animals*. *An Evolutionary Approach*. Oxford Series in Ecology and Evolution. Oxford: Oxford Univ. Press., 1997.