Symbiosis: Using Predator-Prey Games as a Test Bed for Studying Competitive Co-evolution

Fani A. Tzima, Andreas L.Symeonidis and Pericles A. Mitkas
Electrical and Computer Engineering Dept.,
Aristotle University of Thessaloniki, GR541 24, Thessaloniki, Greece
Tel.: +30 2310 99 6359, Fax: +30 2310 99 6398
Email: fani@olympus.ee.auth.gr, asymeon@ee.auth.gr, mitkas@eng.auth.gr

Abstract— *The animat approach constitutes an intriguing attempt to study and comprehend the behavior of adaptive, learning entities in complex environments. Further inspired by the notions of co-evolution and evolutionary "arms races", we have developed Symbiosis, a virtual ecosystem that hosts two self-organizing, combating species – preys and predators. All animats live and evolve in this shared environment, they are self-maintaining and engage in a series of vital activities – nutrition, growth, communication – with the ultimate goals of survival and reproduction. The main objective of Symbiosis is to study the behavior of ecosystem members, especially in terms of the emergent learning mechanisms and the effect of co-evolution on the evolved behavioral strategies. In this direction, several indicators are used to assess individual behavior, with the overall effectiveness metric depending strongly on the animats net energy gain and reproduction rate. Several experiments have been conducted with the developed simulator under various environmental conditions. Overall experimental results support our original hypothesis that co-evolution is a driving factor in the animat learning procedure.*

I. Introduction

During the last decades, multi-agent systems (MAS) have received increasing attention in the artificial intelligence community [1]. Research efforts in the area focus, amongst others, on the investigation of autonomous and rational behavior of agents (software programs or robots), as well as their interaction and coordination, in fields such as robotics [2], information retrieval and management [3], and simulation [4], [5], [6].

In these domains intelligence constitutes a vital part of agent behavior, since it implies a certain degree of autonomy. In other words, it is of paramount importance for agents to be provided with the ability to make independent decisions. Nevertheless, in most dynamic domains, it is impossible for the designer to predefine all potential situations an agent may encounter and, therefore, the agent needs to be equipped with the ability to learn from and adapt to its environment.

Exactly in this direction, animats – artificial animals – and the animat approach constitute an intriguing attempt to study and comprehend the behavior of adaptive, learning entities in complex and probably hostile environments [7]. Using concepts borrowed from evolutionary Biology, the animat approach models animals as autonomous agents interacting with their environment and capable of learning multiple disjunctive concepts through experience [8]. In this context, environmental aspects are strongly coupled with an effective learning mechanism, while societal structures and their impact on animat evolution also constitute an important model factor.

On the other hand, co-evolution (i.e. the parallel evolution of multiple combating species with coupled fitness) introduces several interesting features that may potentially enhance the adaptation power of artificial evolution [9] or other types of bio-inspired learning systems. In particular, competing populations may increase their levels of complexity and effectiveness through their mutual interaction, which escalates to an evolutionary "arms race".

In direct analogy with the theory of co-evolution in ecology, evolutionary computation defines co-evolution as iterated adaptation involving "arms-races", either between learning species [10] or between a learner and its learning environment [11]. Early examples of co-evolutionary learning include the pioneering work by Hillis on sorting networks [12], by Tesauro on a self-playing Backgammon learner [13], by Angeline and Pollack on co-evolving Tic-tac-toe players [14]. In the adaptive behavior community, there are also numerous efforts focusing on co-ecolution in predator/prey games [15].

Advancing one step further on the way co-evolution and evolutionary arms races affect predator/prey systems, we have developed *Symbiosis*, a virtual ecosystem that hosts two self-organizing, combating species: preys, which are herbivorous, and predators, which are carnivorous and consume preys. All animats live and evolve in this shared environment, they are self-maintaining and engage in a series of vital activities, with the ultimate goals of survival and reproduction.

Following the animat approach, our strategy has a "problem side" (harder environments, increased efficiency) and a "solution side" (learning mechanism, adaptation) [16]. Most of the research on animat-like systems tends to emphasize the solution side: a certain experimental environment is selected and focus remains on testing and refining a particular architecture in that environment. The problem, in this case, is that results provide insufficient insight into the properties of the problem solved and the solution's efficiency. Thus, focusing more on the problem side, our main objective in this paper is the study of animat learning and evolved effectiveness under various scenarios of environmental conditions and co-evolution. In this direction, several indicators have been employed to assess individual behavior. The overall effectiveness metric depends strongly on the animats net energy gain and, therefore, its reproduction rate. Under the prism of co-evolution, we argue that the two species' effectiveness metrics are implicitly coupled and interdependent, since an animat's effectiveness reflects how well this individual performs against the opponent population.

In order to mainly focus on the key aspects of adversarial co-existence and learning, inter- and intra-species interactions have been seriously taken into account. The realization of *Symbiosis* as a multi-agent system (MAS), and specifically as two separate agent communities, ensures heterogeneity, while simultaneously allowing for the formation and study of collective behavioral patterns. It is worth noting that, despite the absence of an explicit communication protocol, all animats implicitly communicate with each other through their shared environment, as they live and cope with the environmental changes imposed by other animats decisions. These decisions are determined by an agent learning mechanism modeled using classifier systems and genetic algorithms [17].

The rest of the paper is structured as follows: Section II describes the implemented system, specifying the characteristics of the structural elements of *Symbiosis*. Section III provides two series of experiments conducted by the use of *Symbiosis*, in order to investigate: (i) the effect of the genetic rule creation mechanism on agent performance, and (ii) the robustness of the learning mechanism, provided a "taxonomy" of environments. Finally, Section IV concludes this work and proposes some thoughts on future research directions.

II. System Overview

Environment model

The environment space of the agents in *Symbiosis* is a two-dimensional (xxy) grid, where each cell represents a site of the environment and can be referenced through its coordinates. The environment contains food, traps, obstacles and agents (preys or predators), while the remainder of the space is covered by vacant cells. It should also be pointed out that no movement is allowed beyond the grid borders, i.e. the grid is surrounded by obstacles on all sides. Fig. 1 illustrates an example of a 10x10 environment, along with the color conventions and numerical values associated with each cell type.

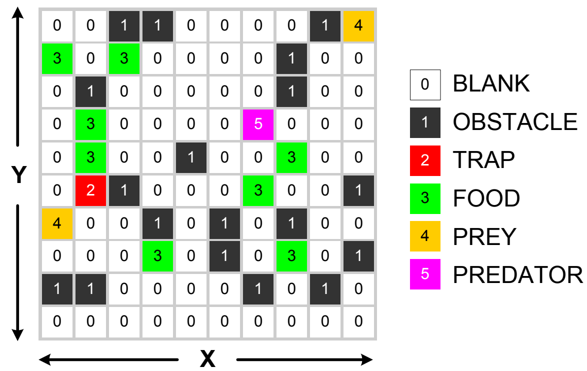

Fig. 1. Sample environment in Symbiosis

Environment initialization may be done automatically or by the user. In the latter case, the percentage of each type of objects in the grid must be specified through the user interface. *Symbiosis* then generates an ecosystem with the specified analogy of objects scattered on the grid, following a normal distribution, where agents are randomly dispatched in vacant cells.

Agent model

The agent model in *Symbiosis* incorporates essential features of a living organism in a diverse and uncertain environment. More specifically, in order to achieve their goals, agents have to adapt to their environment's initially unknown and changing conditions, maintain a positive energy flow and produce offspring throughout their lifetime. The adaptation procedure of *Symbiosis* agents involves a learning mechanism based on personal experience, as well as credit received through interaction with the environment.

An agent's action cycle in each time step involves:

- (i) looking around and forming the current vision vector,
- (ii) deciding on the next target position,
- (iii) moving to the target position on the grid,
- (iv) consuming food/prey (if any in the current position),
- (v) updating the rule-base (the agent's internal representation of the environment) based on the outcome of the last move
- (vi) reproducing if certain conditions are met.

In the following sections we briefly discuss each of these steps, with the exception of the decision making procedure (step (ii)). We will further elaborate on this in section II-C.

Agent Sight—Instead of using the von Neumann [5] or the Moore neighborhood [8], *Symbiosis* agents are able to sense (see) an nxn area of the environment. For this extended

Moore neighborhood, a vision vector of $n\text{x}(n-1)$ values is constructed by taking, in a row-major fashion, the contents of the cells around the agents position. Fig. 2 illustrates a 5x5 vision field and the corresponding vision vector. The agents position is not included in the vision vector.

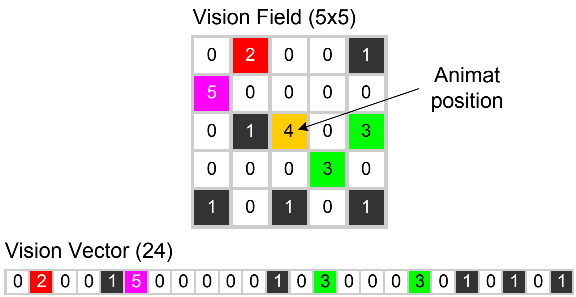

Fig. 2. Agent vision field (5x5)

In order to introduce the notion of uncertainty into *Symbiosis*, another kind of object has been added to the environment. This object may look like food (or prey in the case of predators), while it actually is a trap, with a given probability p (user-specified). Moreover, a prey may not recognize a predator and perceive the corresponding cell as empty, with the same probability p. By tuning the probability parameter p, the user can define more or less reliable environments, where the agent learns to trust its senses in various degrees.

Agent Decision Making—Agents can move towards any cell within their vision field, as long as this cell is not occupied by another agent or any obstacle. The agent decides on a target cell based on a set of classifiers implementing its movement rules. The movement decision involves checking the current vision vector against the set of classifiers, finding the optimal match and moving towards the proposed cell (Fig. 3).

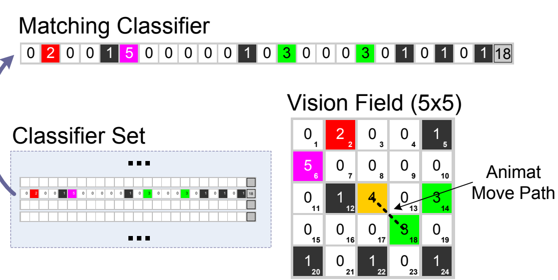

Fig. 3. Agent decision making based on classifiers (movement rules)

Agent Movement—The route that the agent will follow towards its destination is cost dependent. The shortest path is selected among all alternatives, based on the A* (A-star) algorithm [18]. This, of course, presupposes maintaining a graph representation of the environment, where obstacles, agents and recognized unsafe cells (traps or predators for preys) are unreachable nodes.

Agent Reproduction—An agent may reproduce, if its energy balance is equal or greater than z energy units (eu). Only one offspring is born each time and it inherits its parents perception of the environment (in direct analogy with evolved traits being passed on from parent to offspring in nature). The newly born agent appears in a neighboring position in the grid, while its birth results in the reduction of the parent's energy by one half.

Knowledge Model – Evaluation Mechanism

One of the core features of *Symbiosis* is the ability of its members to self-organize and augment their intelligence. Agents in *Symbiosis* have an ongoing interaction with the dynamic environment, wherein they are forced to make decisions continuously. Moreover, they must interleave action selection and learning from outcomes, in an environment that provides them with both immediate and long-term feedback on their actions. Therefore, their learning mechanism must be properly formulated so as to allow for real-time decision evaluation and corresponding revision of the underlying knowledge model.

The learning mechanism we have implemented comprises three parts: (i) a set of classifiers, implementing the agent knowledge model, (ii) a classifier evaluation mechanism, allowing for real-time decision evaluation, and (iii) a rule creation mechanism, comprising a random and a genetic process of rule creation.

Classifier System—The classifier system implements the agent's knowledge model of the environment. Each classifier consists of two parts, the detectors and the effectors, and can be represented as in Eq. (1):

$$\text{If} < \text{detector} > \text{then} < \text{effector} > \qquad (1)$$

This type of classifier can be directly mapped to the structure of an agent message:

$$< \text{ClassifierRule} >::=< \text{condition} >:< \text{movement} > \qquad (2)$$

where the <condition> clause is a stream of 0s, 1s and wildcards (#), while the <movement> clause is a stream of 0s and 1s. Every time an agent moves to a new position within its vision field, a new vision vector is generated. This vector is transformed into a bit stream (Fig. 4), which facilitates comparison to the detectors of the classifier set.

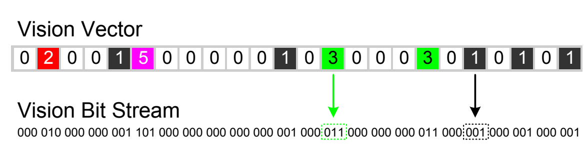

Fig. 4. Transforming the vision vector into a bit stream

In case the detector part of a classifier stored in the agent's knowledge base matches the vision bit stream, the corresponding effector is activated and the agent moves towards the proposed cell. However, it is quite probable to have more than one classifiers matching the incoming vision bit stream. This problem is addressed by the classifier evaluation mechanism.

Classifier Evaluation mechanism—A slightly modified version of the bucket brigade algorithm [17] is employed for evaluating classifiers. In case there are multiple matches, a bidding process between the candidate classifiers is initiated. Bidding is based on the strength of each classifier rule, a quantity that represents the success rate of a classifier with respect to its prior decisions (i.e. when a classifier rule sends an agent to a food cell, then its strength is increased and vice versa). The bid that a classifier C participating in an auction is going to place at time t, is given by:

$$bid_C(t) = C_{bid} \cdot strength_C(t) \tag{3}$$

where C_{bid} is a constant value that determines the strength percentage each classifier is willing to place in the bid. The classifier that places the higher bid is announced winner of the auction and it is activated. The strength for C is now (at time t+1):

$$\begin{aligned} strength_C(t\text{+}1) = strength_C(t) + energy_C(t\text{+}1) \\ -energy_C(t) - bid_C(t) - tax_C(t) \end{aligned} \tag{4}$$

where $energy_C(t\text{+}1)$ is the current energy of the agent (after having followed the route to the indicated cell and having consumed the food there, or having fallen into the trap), $energy_C(t)$ is the energy of the agent on time t, and $tax_C(t)$ a value for participating in the auction (all participating classifiers pay the tax). Finally, C rewards the classifier C' which was the previous activated rule (at time t–1) and led to the activation of C. The new strength value of C' is now:

$$\begin{aligned} strength_{C'}(t\text{+}1) = strength_{C'}(t) \\ +bid_C(t) - tax_C(t) \end{aligned} \tag{5}$$

In this way, the classifier set is constantly evaluated and the best performing rules gain strength.

Rule creation mechanism—If no classifier detector matches the current vision vector, then a new classifier is constructed, with the vision bit stream as a detector, and a random effector. Moreover, in order to increase the variety of the knowledge base, a genetic algorithm mechanism is employed, which generates new classifiers by applying selection, crossover and mutation operators on the best-performing classifiers already residing in the rule set. The products of the genetic rule creation mechanism replace the weakest (in terms of strength) classifiers, therefore providing a notion of "rejuvenation" of the agent's mentality. The invocation interval of this process is user-defined.

Assessment Indicators

One of the primary objectives of the developed system is to study agent learning and their evolved effectiveness under various scenarios of environmental conditions and with respect to co-evolution. Thus, a series of environmental and agent assessment indicators are defined, in order to better comprehend how changes in the ecosystem affect its members, and how changes in the behavior of the agents affect their environment.

Environmental indicators—Environmental indicators include resource availability, environmental variety and environmental reliability and are defined according to the following equations:

- *Resource availability* depends (i) on the total amount of resources (food or preys) available and (ii) on their distribution pattern. Thus, it is defined as the ratio of the total resouces r available in the environment to the total harvesting distance d_r:

  $$a = \frac{\sum r}{\sum d_r} \tag{6}$$

  The denominator of Eq. (6) is calculated as the sum of steps needed in order to collect all resources, starting from a random position and moving always to the nearest resource cell.
- *Environmental variety* v of an ecosystem is defined as the ratio of the distinct vision vectors vv that may exist to the occupied cells of the grid oc (cells that contain either resources, traps or obstacles):

  $$v = \frac{\sum vv}{\sum oc} \tag{7}$$

- *Environmental reliability* r is defined as:

  $$r = 1 - p \tag{8}$$

  where p is the vision error probability, i.e. the probability that an agent will misinterpret the vision vector.

Agent performance indicators—Agent performance indicators include energy, effectiveness, age, resource consumption rate, trap collision rate, unknown situation encounter rate and reproduction rate. They are defined according to the following equations:

- *Energy* en is they key feature of an agent: at birth, agents have an initial energy level, which fluctuates depending on their actions (eating increases energy, falling in traps and colliding with obstacles reduces it, moving burns it, etc.). When an agent runs out of energy, it dies.
- An integer ag_i defines the *age* of agent i, measured in ecosystem epochs. It should be denoted that the energy loss rate increases, when the organism "ages" i.e. it exceeds a pre-specified number of epochs.
- The *resource consumption rate* indicator rcr is defined as:

  $$rcr = \frac{\sum r_{encounter}}{s} \tag{9}$$

  where the numerator represents the number of resource cells the agent hits, while the denominator represents the moves that the agent has performed.
- *Trap collision rate* is defined as:

  $$tcr = \frac{\sum t_{collision}}{s} \tag{10}$$

where $\sum t_{collision}$ is the number of traps the agent has collided with.

- The *unknown situation rate* usr is defined as:

$$usr = \frac{\sum unknownSituation}{s} \tag{11}$$

where $\sum unknownSituation$ is the number of total unknown situations (no classifier matches the vision vector) that the agent has encountered.

- The agent *reproduction rate* is defined as:

$$rr = \frac{\sum reproduction}{s} \tag{12}$$

where $\sum reproduction$ is the number of offspring the agent has given.

- *Effectiveness* e is defined as:

$$e = 1 + \frac{eur - elr}{ear} \tag{13}$$

where eur is the energy uptake rate, elr the energy loss rate and ear the energy availability rate, which are given respectively by Eq. (14), (15) and (16) :

$$eur = rcr \cdot ResourceCost \tag{14}$$

$$\begin{aligned} elr = tcr \cdot TrapCost + MotionLossRate \\ + BirthLossRate + LossByPredatorsRate \end{aligned} \tag{15}$$

$$ear = a \cdot speedOfMotion \tag{16}$$

- Finally, *net effectiveness* (e_{NET}) focuses on preys' effectiveness, only in terms of resource exploitation, trap avoidance etc., excluding their performance in the task of evading from predators. Thus, it is defined as in Eq. (13), with the difference that the energy loss by predators is not taken into account when calculating the energy loss rate (see Eq. (15)).

III. Experiments And Results

Several experiments have been conducted with the developed simulator, in order to explore the nature and extent of the effects of co-evolution pressure. Our experimental approach involves the following process: the two animat populations (with their conflicting needs) and their learning mechanisms (that aim to satisfy these needs), are tested against (i) differentiation of the learning capabilities of one of the two species, (ii) variation of the invocation frequency of the genetic rule creation mechanism and (iii) variation of the environmental conditions (resource availability, environmental variety, environmental reliability, food refresh rate). In all cases, our primary goal is to test the robustness of the learning mechanism and investigate whether (and to what extent) co-evolution of the two species affects the learning process. In all cases, agent performance metrics are used as indicators of the success (or failure) of the agent learning mechanism.

Results presented hereafter focus on the prey population and are structured in two series of experiments (see Table I for parameters). Experiment *Series A* concerns the effect of the genetic rule creation mechanism on agent performance, while *Series B* investigates the robustness of the learning mechanism, provided a "taxonomy" of environments. In both series of experiments, results are comparatively presented between two cases: (i) both species employing the learning mechanism prestated, and (ii) learning preys versus predators employing a simple "prey-hunting" strategy.

TABLE I
Experiment Parameters (both Series A and B)

Parameter	Value
Grid dimensions	50x50
Initial prey population	80
Initial predator population	8
Initial prey energy	500
Initial predator energy	800
Ageing steps	2000
Food refresh epochs (unless otherwise stated)	10
Food cost	10
Prey cost	20
Trap cost	-10
Classifier set size	3000
GA Crossover rate	0.7
GA Mutation rate	0.01

Varying the genetic rule creation mechanism

Four experiments were conducted to study the effect of the genetic rule creation mechanism on agent performance. Given a certain environment with resource availability $a = 0.3368$, environmental variety $v = 2.3334$ and environmental reliability $r = 0.95$, different genetic algorithm invocation steps (values 100, 200, 500 and 50 – experiments $A1$ through $A4$ respectively) were applied. In all cases, co-evolution of the two species resulted in an increase of the agent performance indicators. Indicatively, for GA rate = 100 (which proved to be the optimal value), preys' mean effectiveness improved by 14.8% and mean net effectiveness by 11.9%, when both species co-evolved using the learning mechanism. Figure 5 illustrates the preys' (denoised) mean net effectiveness throughout experiment Exp_{A1a} (both species learning) with respect to Exp_{A1b} ("naive" predators).

Varying the environmental parameters

Following the animat approach and focusing mainly on the problem side, we have defined a "taxonomy" of environments (Table II), according to their resource availability (a), environmental variety (v), environmental reliability (r) and food refresh rate. Then, keeping the GA rate to 100 (optimal value – experiment *Series A*), we have conducted numerous experiments in each of these environments, trying to investigate the effect of co-evolution on preys' effectiveness. For all environment configurations, we observed a remarkable improvement of preys' performance indicators when both species were learning. It is worth noting, that the maximum improvement (16.42%) of mean effectiveness was achieved in Exp_{B7},

TABLE II
EXPERIMENT SERIES B: ENVIRONMENT TAXONOMY

Experiment Series B: Environment Taxonomy										
	Exp_{B1}	Exp_{B2}	Exp_{B3}	Exp_{B4}	Exp_{B5}	Exp_{B6}	Exp_{B7}	Exp_{B8}	Exp_{B9}	Exp_{B10}
Resource Availability	$\mathrm{H}-0.44$	$\mathrm{H}-0.44$	$\mathrm{H}-0.44$	$\mathrm{H}-0.44$	$\mathrm{L}-0.32$	$\mathrm{L}-0.32$	$\mathrm{L}-0.31$	$\mathrm{L}-0.31$	$\mathrm{H}-0.44$	$\mathrm{L}-0.31$
Environmental Variety	$\mathrm{H}-2.22$	$\mathrm{H}-2.22$	$\mathrm{L}-1.20$	$\mathrm{L}-1.20$	$\mathrm{H}-3.66$	$\mathrm{H}-3.66$	$\mathrm{L}-1.82$	$\mathrm{L}-1.82$	$\mathrm{H}-2.22$	$\mathrm{L}-1.82$
Environmental Reliability	$\mathrm{H}-0.99$	$\mathrm{L}-0.60$	$\mathrm{H}-0.99$	$\mathrm{L}-0.60$	$\mathrm{H}-0.99$	$\mathrm{L}-0.60$	$\mathrm{H}-0.99$	$\mathrm{L}-0.60$	$\mathrm{H}-0.99$	$\mathrm{H}-0.99$
Food Refresh Rate	10	10	10	10	10	10	10	10	20	5

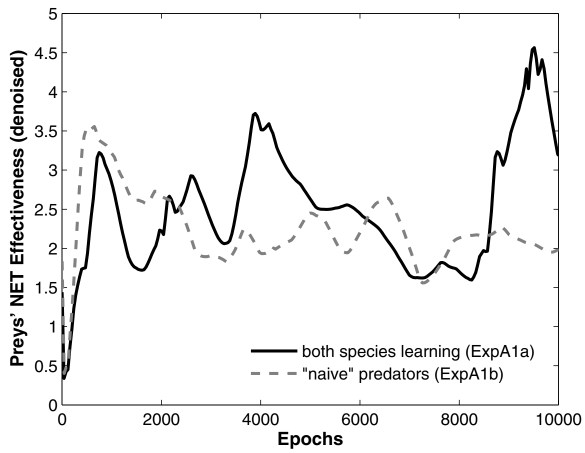

Fig. 5. Preys' denoised net effectiveness - Experiments A1

a "difficult" environment in terms of resource availability. Moreover, a 20.7% increase in mean net effectiveness was achieved in Exp_{B1}, despite its high environmental variety. Finally, in cases of low environmental reliability, such as Exp_{B4} and Exp_{B6}, we observed increased resource exploitation ($+9.2\%$ mean food encounter rate) and greater population density ($+23.1\%$ mean population) respectively.

IV. CONCLUSIONS

In this paper, we have presented a systematic approach to studying and comprehending the behavior of adaptive, learning entities in complex environments. Following the animat approach and incorporating the notions of co-evolution and evolutionary "arms races", we have developed *Symbiosis*, a virtual ecosystem that hosts two self-organizing, combating species – preys and predators. Software agent technology has been adopted in order to model and implement the autonomous nature of the participant entities, while a learning mechanism modeled using classifier systems and genetic algorithms has been incorporated into them. Current work includes an experimental evaluation of the effect of co-evolution on animat behavioral strategies and performance. Experiments indicate that co-evolution is indeed a driving factor in the animat learning procedure, regardless of the learning mechanism parameters or the complexity of the environment. Further work that will enhance the scope of the system, includes a more in depth study of predator behavior, as well as the efficiency of our learning mechanism in scenarios of explicit cooperation.

REFERENCES

[1] E. Alonso, M. d'Inverno, D. Kudenko, M. Luck, and J. Noble, "Learning in multi-agent systems," *Knowledge Engineering Review*, vol. 16, no. 3, pp. 277–284, 2001.

[2] J. Walker, S. Garrett, and M. Wilson, "Evolving controllers for real robots: A survey of the literature," *Adaptive Behavior*, vol. 11, no. 3, pp. 179–203, 2003.

[3] M. Klusch, "Information agents: Theory and applications," *International Journal on Cooperative Information Systems*, vol. 10, no. 1-2, pp. 51–56, 2001.

[4] R. Conte, N. Gilbert, and J. S. Sichman, "MAS and social simulation: A suitable commitment," *Lecture Notes in Computer Science*, vol. 1534, pp. 1–9, 1998.

[5] J. M. Epstein and R. Axtell, *Growing artificial societies: social science from the bottom up*. Washington, DC, USA: The Brookings Institution, 1996.

[6] A. L. Symeonidis, S. Seroglou, V. Valtos, and P. A. Mitkas, "Biotope: an integrated framework for simulating distributed multiagent computational systems," *IEEE Transactions on Systems, Man and Cybernetics, Part A*, vol. 35, no. 3, pp. 420–432, 2005.

[7] S. W. Wilson, "Classifier systems and the animat problem," *Machine Learning*, vol. 2, no. 3, pp. 199–228, 1987.

[8] F. Krebs and H. Bossel, "Emergent value orientation in self-organization of an animat," *Ecological Modelling*, vol. 96, no. 1, pp. 143–164, 1997.

[9] S. Nolfi and D. Floreano, "How co-evolution can enhance the adaptive power of artificial evolution: Implications for evolutionary robotics," in *Proceedings of the First European Workshop on Evolutionary Robotics*. London, UK: Springer-Verlag, 1998, pp. 22–38.

[10] I. Tanev, M. Brzozowski, and K. Shimohara, "Evolution, generality and robustness of emerged surrounding behavior in continuous predators-prey pursuit problem," *Genetic Programming and Evolvable Machines*, vol. 6, no. 3, pp. 301–318, 2005.

[11] M. Ebner, "Coevolution and the red queen effect shape virtual plants," *Genetic Programming and Evolvable Machines*, vol. 7, no. 1, pp. 103–123, 2006.

[12] W. D. Hillis, "Co-evolving parasites improve simulated evolution as an optimization procedure," in *Artificial Life II*, C. L. et al., Ed. Addison Wesley, 1992, pp. 313–324.

[13] G. Tesauro, "Practical issues in temporal difference learning," *Machine Learning*, vol. 8, no. 3-4, pp. 257–277, 1992.

[14] P. J. Angeline and J. B. Pollack, "Competitive environments evolve better solutions for complex tasks," in *Proceedings of the 5th International Conference on Genetic Algorithms (GA-93)*, 1993, pp. 264–270. [Online]. Available: citeseer.ist.psu.edu/angeline93competitive.html

[15] D. Cliff and G. F. Miller, "Co-evolution of pursuit and evasion II: Simulation methods and results," in *From animals to animats 4*, P. Maes, M. Mataric, J. Meyer, J. Pollack, and S. Wilson, Eds. Cambridge, MA: MIT Press, 1996, pp. 506–515. [Online]. Available: citeseer.ist.psu.edu/cliff95coevolution.html

[16] S. W. Wilson, "The animat path to AI," in *Proceedings of the first international conference on simulation of adaptive behavior (From animals to animats)*, J.-A. Meyer and S. W. Wilson, Eds. Cambridge, MA, USA: MIT Press, 1991, pp. 15–21.

[17] J. H. Holland, *Adaptation in natural and artificial systems*. Cambridge, MA, USA: MIT Press, 1992.

[18] P. E. Hart, N. J. Nilsson, and B. Raphael, "A formal basis for the heuristic determination of minimum cost paths," *IEEE Transactions on Systems Science and Cybernetics SSC4*, vol. 2, pp. 100–107, 1968.