Artif Life Robotics (2005) 9:222–233
DOI 10.1007/s10015-005-0349-5

ORIGINAL ARTICLE

Geoff Nitschke

Designing emergent cooperation: a pursuit–evasion game case study

Received: December 6, 2004 / Accepted: September 30, 2005

Abstract This research concerns a comparison of two neuroevolution approaches for the design of cooperative behavior in a group of simulated mobile robots. The first approach, termed single pool (SP), was characterized by robot neural controllers that were derived from a single genotype. The second approach, termed multiple pools (MP), was characterized by robot neural controllers that were derived from many different genotypes. The application domain implemented a pursuit–evasion game in which teams of robots of various sizes, termed predators, collectively worked to capture (immobilize) other robots, termed prey. The MP and SP approaches were tested, with and without lifetime learning, for the design of cooperative prey capture behavior within teams of predators. Results indicated that the MP approach was superior to the SP approach in terms of measures defined for prey-capture performance. Specifically, the MP approach facilitated behavioral specializations in the predator team facilitating emergent cooperative prey capture strategies that proved effective for the various team sizes tested.

Key words Cooperative behavior · Simulated multi-robot system · Pursuit–evasion · Emergence · Neuroevolution

1 Introduction

This article describes a set of experiments testing the efficacy of different neuroevolution approaches for the design of emergent, yet desired, forms of cooperative behavior in a simulated multirobot team. The approaches for the synthesis of cooperative behaviour were evaluated within teams of simulated Khepera robots.[1] The task domain is a collective pursuit–evasion game played by the simulated robots, where it was the collective task of pursuers (herein called predators) to immobilize at least one evader (herein called prey). A control experiment using a single predator and a single prey demonstrated that cooperation between at least two predators was needed to accomplish this task. Cooperative behavior was only evolved for a predator team, and each prey was able to move 20% faster than the predators. The behavior of each prey was not evolved, but instead used a previously evolved obstacle avoidance behavior. Functionally, each predator was the same in terms of movement and sensor capabilities. The predator team was rewarded a fitness which equated with the total time for which it was able to immobilize a prey.

The first neuroevolution approach tested was termed single pool (SP), in which each predator in the team was given an identical genotype. Hence, the corresponding neural controller (phenotype) for each of the predators in the team was the same. A variant of the SP approach was also tested (herein called the SP variant). The SP variant implemented a self-teaching neural network controller that allowed for lifetime learning. The second neuroevolution approach was termed multiple pools (MP), where each predator was given a different genotype, meaning that the neural controller (phenotype) at the beginning of each predator's lifetime was different. A variant of the MP approach was also tested (herein called the MP variant). The MP variant implemented a self-teaching neural controller for each predator in the team, allowing for lifetime learning in each.

For each neuro-evolution design approach, various group configurations (differing numbers of predators and prey) were compared. The two approaches were evaluated in terms of the time-period for which a prey was immobilized (fitness) during the course of a given simulation, and the geometrical stability of evolved prey capture strategies. Results indicated that the MP design approach, which facilitated emergent cooperative behavior by deriving complementary and specialized behavioral roles in the predator team, yielded superior performance in terms of the two measures defined to quantify the team's performance. This

G. Nitschke (✉)
Computational Intelligence Group, Department of Computer Science, Faculty of Sciences, De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands
Tel. +31-20-44-47506; Fax +31-20-44-47653
e-mail: nitschke@cs.vu.nl

superior performance proved to be consistent for all group configurations tested.

2 Related literature

The synthesis of collective behavior, particularly cooperation, using artificial evolution mechanisms as a design methodology is a research field in which there has been little work done in either simulated[2] or real-world[3] problem domains. Traditionally, the topic of emergent cooperative behavior has been studied within the field of multiagent systems using a top-down classical approach. Such approaches have achieved limited success given that it is extremely difficult to specify the mechanisms for cooperative behavior in all but the simplest problem domains. For example, emergent cooperation is commonly studied within abstract problem domains, such as the iterated prisoners dilemma (IPD)[4] or other multiagent scenarios[5] that operate within a game theory domain. Axelrod[6] explored the conditions under which fundamentally selfish agents were more likely to spontaneously cooperate. To perform this study, Axelrod[6] used the IPD game, which offers a long-term incentive for cooperation but a short-term incentive for defection. The IPD game has also been extended to account for the emergence of cooperation between more than two players,[7] as well as different forms of cooperation, such as nonmutual cooperation, where any altruistic behavior exhibited increases the likelihood of cooperative behavior being reciprocated at a later time.[8] Aside from classical artificial intelligence applications of cooperation to distributed problem domains, game theoretic models, or those modelling abstract problem domains such as sugar-scape[9] and other similar models,[2] research on the emergent cooperation using situated and embodied agents[10] has received relatively little attention due to the inherent complexity of such agents operating in the real world.

Concepts such as self-organization, emergence, and evolution are now thought by many researchers to pose a reasonable alternative to traditionalist artificial intelligence design approaches to multiagent cooperation. For example, artificial evolution has been used successfully for the derivation of cooperative pursuit strategies in the pursuit–evasion domain,[11,12] as well as to attain an ecological equilibrium between groups of predators and prey.[13] A common problem that confounds such research is that cooperative behaviors that emerge from interacting constituents of the system are difficult to analyze, as it is nontrivial to determine what mechanisms are responsible for what behaviors.

The utilization of neuroevolution[14] for design of desired yet emergent cooperative behaviors remains a relatively unexplored area of research in the pursuit and evasion domain[15] and related predator–prey systems[13] using multiple predators and prey. Various approaches have been used to study the pursuit–evasion domain, where the task is for multiple predators to capture a prey by surrounding it,[16,17] although few researchers have investigated emergent cooperation in these systems, with notable exceptions such as Denzinger and Fuchs,[18] Haynes and Sen,[11] and Yong and Miikkulainen.[12] Emergent cooperation in the pursuit–evasion domain has obvious real-world applications such as the formulation of military, reconnaissance, or search and rescue strategies in environments for which there are relatively few possibilities of specifying cooperative behaviors a priori.

3 Experimental setup and neuroevolution design approaches

Experiments were conducted in simulation using an extended version of the Evorobot Khepera simulator.[19]

3.1 Agents, controllers, and environment

The embodiment of each predator and prey was assumed to be a Khepera mobile robot.[1] As presented in Fig. 1A, the robots used as prey were equipped with eight infrared proximity sensors, as well as a light on the top. This light could be detected by the predator's light sensors, and was used so each predator could distinguish fellow predators from a prey. As presented in Fig. 2A, robots used as predators were equipped with eight infrared proximity and eight light sensors positioned on their periphery. Both the predators and the prey utilized a purely reactive behavior. As illustrated in Figs. 2B and 1B, respectively, behavior was controlled by a feed-forward neural controller, directly con-

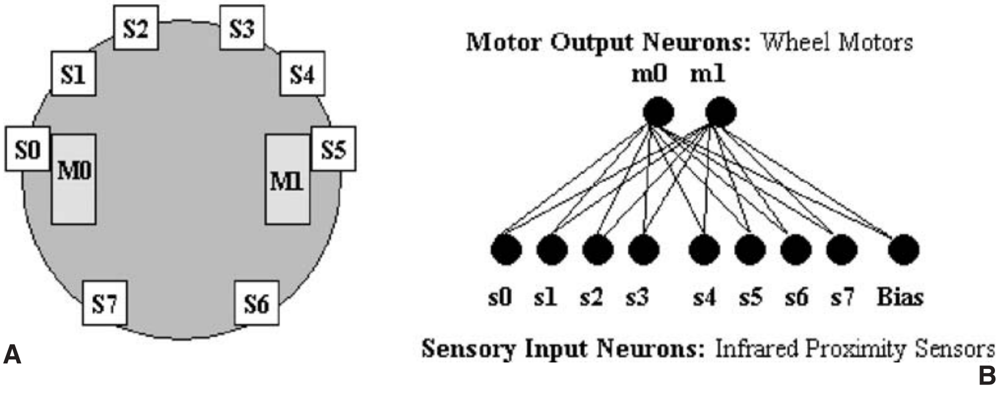

Fig. 1. A Sensor configuration for the simulated *prey* Khepera robots. **B** A two-layer feed-forward neural network was used as the prey controller (weights not evolved in the pursuit–evasion simulations)

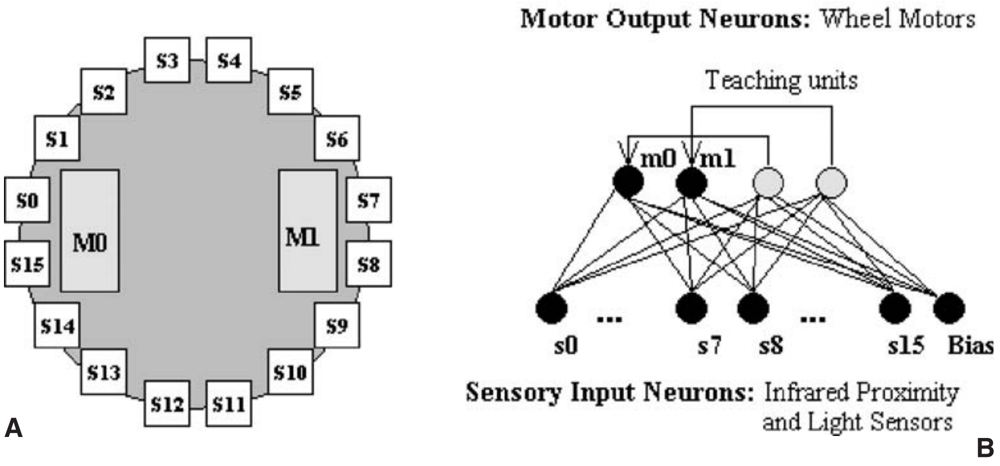

Fig. 2. A Sensor configuration for the simulated *predator* Khepera robots. **B** A two-layer feed-forward "self-teaching" neural network was used as the predator controller. Outputs of the teaching units were used by the motor units as teaching input in order to change the weights of connections leading to the motor units

necting sensory input units to motor output units. For both predators and prey, the activation of each output unit was used to update the speed value of the corresponding wheel every 100 ms. In order to allow the prey to move 20% faster than the speed of the predators, the activation value to each output unit was multiplied by 1.2 before setting the wheel speed.

In order to produce lifetime learning in the predator controller, the connection weights of the teaching network (presented in Fig. 2B as teaching weights) did not change during a predator's lifetime. However, standard connection weights did change based on the teaching input provided by the teaching weights. The two motor units used this teaching input in order to learn using back propagation.[20]

Since it was the standard weights that controlled a predator's behavior in the environment, the predator's behavior also changed. Hence, adaptation of these standard weights within a predator's lifetime denoted lifetime learning. Lifetime weight changes were then passed on to the next generation of controllers by having genetic recombination and crossover operators applied after the final life cycle of the controller. Predator controllers that did not utilize lifetime learning did not have teaching weights, so the standard weights changed only via genetic operators applied at the end of each controller's lifetime, during the course of an evolutionary process.

The simulation environment corresponded to a 1000 cm × 1000 cm arena with no obstacles. When the predator and prey robots were placed in the environment, sensory input was received via the input units, and activation values were passed to the two motor units and the teaching units. The activation value of the two motor units was used to move the robots, thus changing the sensory input for the next simulation cycle. This cycle was then repeated.

3.2 Simulating the predators, prey, and the environment

There are several facts that must be taken into account when designing a simulation model of a robot and its environment. Miglino et al.[21] elucidated that the responses of the infrared sensors of different Kheperas vary because of slight differences in the environment, such as ambient light settings, color, and the shape of the objects. Given that sensors respond in a significantly different way from other sensors when exposed to the same external stimulus, two different Khepera robots may perform very differently from each other in identical conditions because of the differences in the sensory characteristics. Thus, to simulate the predator and prey robots and the environment as accurately as possible, a sampling procedure described by Nolfi and Floreano[22] was reproduced in order to compute the activation state of the infrared and ambient light sensors. A real Khepera was made to turn 360° at different distances with respect to a wall and to an obstacle, whilst the activation levels of the sensors were recorded. For the predators, the activation levels of each of its eight infrared sensors and eight light sensors were recorded for 180 orientations and 20 different random distances. The same was done for the prey, but only for eight infrared sensors. The resulting samples were then used by the simulator to set the activation levels of the simulated prey and predator agents depending on their current positions in the simulated environment.

Conservative position noise[21] was added to make the simulated agents perceive objects as if they were farther or closer (with respect to a randomly selected axis) than they really were, thus producing affects similar to those in a physical environment that result from differences between the illumination of objects, shadows, or slight physical differences between objects of the same type. The same procedure was used to sample the state of the ambient light sensors for different orientations and distances with respect to a light bulb. These recorded values, together with a geometrical simulation of shadows, were then used in a simulation to set the activation state of the ambient light sensors. To simulate the Khepera's motors, the approach described by Jakobi et al.[23] was used. Using the speed sensors of Khepera and letting it move in the real environment, one was able to set the activation level of the neural controller input units, and to compute the displacement of the robot in the simulated environment.

3.3 Evolution of the prey controller

Prior to the prey being placed in the pursuit–evasion simulation environment with a predator team, the prey controller was evolved (in simulation) for static and dynamic obstacle avoidance. The training environment of the prey was a 1000 cm × 1000 cm rectangular arena surrounded by walls with obstacles placed in random positions.

In the first set of training scenarios, in which static obstacle avoidance behavior was being evolved, between 2 and 10 (randomly determined) uniform round objects, each of 5.5 cm diameter, were placed within the environment. This obstacle type was used because they are perceived in the same way independently from the point of view.[21] The obstacle diameter of 5.5 cm was used as this is the diameter of a Khepera robot. A minimum of 2 and a maximum of 10 objects were used, as this was the minimum and maximum predator team size tested in the pursuit–evasion simulations.

The prey performed an obstacle avoidance behavior by moving forward as fast as possible, and moving in as straight a line as possible and keeping as far away from objects as possible. In order to evaluate prey performance, Eq. 1[21] was used.

$$F = \sum_{i=1}^{1000} V_i\left(1 - DV_i^2\right)\left(1 - I_i\right) \tag{1}$$

where Vi was the average rotation speed of the two wheels, DVi was the algebraic difference between the signed speed values of the wheels, and Ii was the activation values of the proximity sensor with the highest activity at time i. The fitness function was how far and have fast the prey could move without colliding with an obstacle.

To evolve the obstacle avoidance behavior, a variation of a standard genetic algorithm[24] was used. The initial population consisted of 100 randomly generated genotype strings that encoded the connection weights of 100 corresponding neural controllers. Initially, one genotype was randomly selected from the population and decoded into the prey controller. Each of the 100 genotypes in the population of genotypes was systematically decoded into a neural controller and tested.

At the end of a generation, after all genotypes in the population had been tested and assigned a fitness, reproduction took place. Reproduction involved the selection of the fittest 20% of genotypes from the population of genotypes. Ten sets of parent genotypes were randomly selected from within this 20% portion of genotypes, and each set of parents produced five child genotypes. Single-point crossover was used to recombine two parent genotypes, and mutation of a random value between –1.0 and +1.0 was applied to each gene with a 0.05 degree of probability.

As illustrated in Fig. 1B, each prey neural controller consisted of eight sensory neurons encoding the state of the infrared sensors, and a bias unit directly connected with two motor neurons that controlled the speed of the two wheels. The activation state of the bias unit was always 1.0. The genetic encoding scheme was a direct one-to-one mapping. In this case, each connection weight corresponded to a floating point number in the interval [–10, +10]. The genotype of each prey controller thus consisted of 17 floating point values representing the 16 weights and the bias of the neural network. Network architecture and learning rate were fixed and identical for all controllers.

Ten replications of the experiment to evolve a static obstacle avoidance behavior were made, where each experiment ran for 500 generations. The prey "lived" for 50 epochs, where each epoch consisted of 1000 cycles of simulation time. Each epoch constituted a test scenario, which tested different randomly generated obstacle positions and orientations and starting positions for the prey. At the end of the prey's lifetime, a fitness value was then assigned back to the genotype. The assigned fitness of an individual prey genotype was the sum of all its fitness for all epochs of its life.

In the second set of training scenarios, in which dynamic obstacle avoidance behavior was being evolved, the prey with the fittest static obstacle avoidance controller was placed in the same environment again, but with between two and ten (randomly determined) predators. Here, the predator controller was a heuristic pursuit behavior in which the predator moved in a straight line, at maximum speed, and in a random direction (using its infrared sensors to follow an obstacle avoidance behavior) until the prey was detected with it light sensors. The predator would then move in a straight line toward the prey. If a predator collided with the prey, the prey was awarded zero fitness, and a new epoch was started. If the prey reached the end of an epoch within its lifetime without colliding with a predator, then it was awarded a fitness equal to its average speed during the epoch (1000 cycles of simulation time). The fitness of a given controller in a given generation was equal to the sum of its fitness for all epochs of its lifetime. After ten experimental replications, the controller with the highest fitness was then selected as the controller to be used in the pursuit–evasion simulations.

3.4 Neuroevolution for adaptive predator controllers

Two neuroevolution design approaches were comparatively tested and evaluated for the task of having a predator team immobilize either one or two prey. The design approaches tested were termed single pool (SP) and multiple pools (MP). Both approaches were tested with and without lifetime learning in the neural controllers. Approaches using lifetime learning were known as the variant approaches. As illustrated in Figs. 3 and 4, respectively, the SP approach employed a single population of genotypes, whilst the MP approach employed multiple populations of genotypes. Each genotype was encoded with the parameters necessary to perform a direct mapping from a set of neural network weights (genotype) to a neural controller (phenotype) for a predator. Each genotype population was 100 randomly generated genotype strings that encoded the connection weights of 100 corresponding neural controllers.

To derive prey capture strategies, a standard genetic algorithm[24] was applied to genotype strings in order to evolve a corresponding set of weights that, when mapped to a neural controller, would prove effective in the formation of a prey capture strategy. For the SP approach, *n* genotypes were selected from the population of genotypes and decoded into *n* controllers, where *n* depended upon the predator team size, meaning that the predator team was homogenous. The fitness assigned to each predator was simply the fitness calculated for the single genotype that specified the predator team. The main advantage of this approach was its simplicity in terms of behavioral encoding and calculation of team fitness. In the SP variant approach (with lifetime learning), the difference was that individual phenotypes were able to adapt during their lifetime as a result of a self-teaching neural controller (presented in Fig. 2B). This learning process was affected by both genetic and environmental factors. Thus, predator phenotypes were able to adapt to environmental influences throughout the predators lifetime, which affected the fitness calculated for the team genotype, which in turn influenced the selection process in successive generations. The advantage of the SP variant was that it allowed for behavioral specialization in predators without being affected by the problem of needing to estimate a fitness contribution of different predators to the team as a whole.

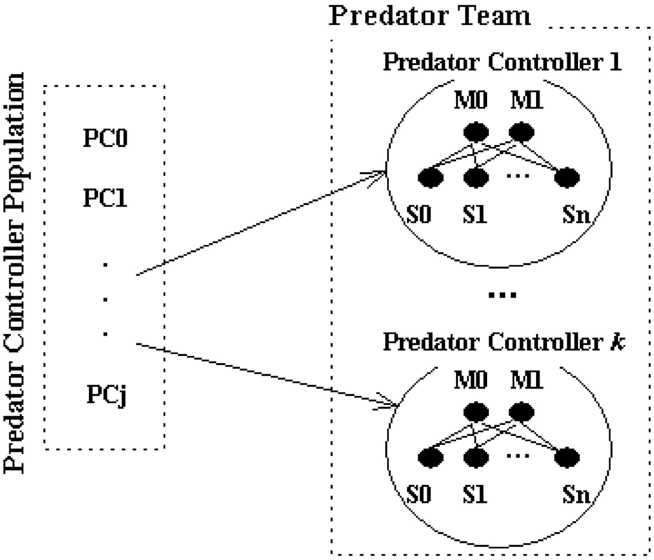

Fig. 3. Single pool (SP) approach. Each predator phenotype (a two-layer feed-forward neural controller) corresponds to a genotype selected from a single population of genotypes and copied *n* times

For the MP approach, *one* genotype was selected from each of *n* populations of genotypes and decoded into *n* controllers, where *n* depended upon the predator team size, meaning that the predator team was heterogeneous. Given that selection operated within each of the genotype populations, and each predator controller corresponded to a genotype from a different population, predator behavior tended to be dissimilar. In the MP variant approach (with lifetime learning), behavior was adaptive within a predator's lifetime as a result of a self-teaching neural controller (Fig. 2B). The advantage of the MP approach was that it encouraged behavioral specialization in the group of predators, in that the evolutionary setup provided for more genetic diversity. The disadvantage was that the assignment of individual predator fitness was an approximation. Specifically, an equal fitness score was assigned to each of the genotypes as a means of deriving the contribution of each predator to the performance of the team as a whole. So the fact that only an estimation of the fitness of a predator existed constituted a problem that may have prevented the selection of the best individual genotypes across successive generations. This evaluation is discussed in the following section.

For either the SP or MP approach, the 100 genotypes in a population of genotypes was systematically decoded into a neural controller and tested for each predator in the team.

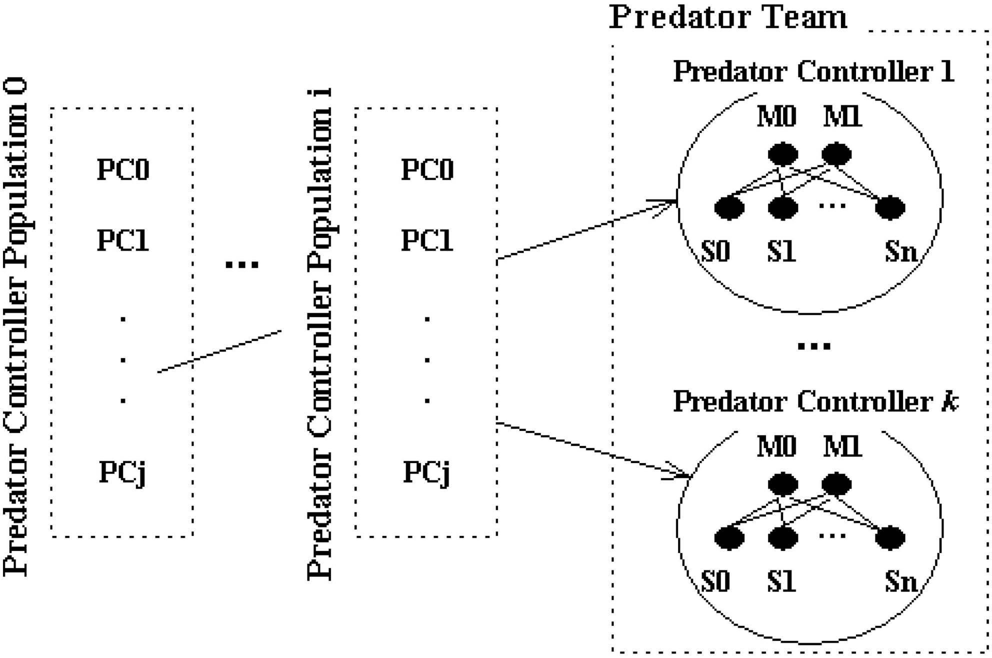

Fig. 4. Multiple pools (MP) approach. Predator phenotypes (a two-layer feed-forward neural controller) correspond to *n* different genotypes selected from *n* separate populations of genotypes

If the predator controllers were derived from the same population of genotypes (SP approach), then a genotype was selected randomly from the population and then flagged as having been selected for a specific predator in the team, so that it would not be selected and decoded as the controller for that predator again.

As illustrated in Fig. 2B, each predator neural controller consisted of 17 sensory neurons encoding the state of infrared and light sensors, and a bias unit directly connected with two motor neurons that controlled the speed of the two wheels. The activation state of the bias unit was always 1.0. The genetic encoding scheme was a direct one-to-one mapping. In this case, each connection weight corresponded to a floating-point number in the interval [–10, +10]. The genotype of each individual thus consisted of 33 floating-point values representing the 32 weights and the bias of the neural network. Network architecture and learning rate are fixed and identical for all predator controllers.

Each predator controller was tested for a "lifetime" of 50 epochs, where each epoch (testing random starting positions and orientations) lasted for 1000 cycles of simulation time. At the end of the controller's lifetime, a fitness was then assigned back to the genotype. In each epoch the fitness of an individual predator controller was equal to the number of simulation cycles for which the speed of the prey equaled zero. The assigned fitness of the corresponding genotype was then the sum of all its fitness value for all epochs of its lifetime.

At the end of a predator's (controllers) lifetime, reproduction took place. Reproduction involved the selection of the fittest 20% of genotypes from the population of genotypes. Ten sets of parent genotypes were randomly selected from within this elite portion of genotypes, and each set of parents produced 5 child genotypes. As illustrated in Fig. 5, single-point crossover was used to recombine two parent genotypes, and mutation of a random value between –1.0 and +1.0 was applied to each gene with a 0.05 degree of probability. This process was repeated for the 500 generations that a simulation ran.

3.5 Fitness evaluation of the predator team

For the SP approach and its variant, a single genotype specified the entire predator team. That is, predators were clones of each other, so the evaluation of team performance in this case was not problematic. The performance of a predator team executed under the SP approaches was simply measured as the fitness value assigned to the genotype that specified the team. In contrast to these approaches, a predator team using the MP approach and its variant was specified by *n* genotypes selected from *n* different populations. Hence, each genotype must be assigned an individual fitness score, and team performance evaluation needed to be computed by estimating the fitness contribution of each genotype to the team as a whole. A method of evaluation widely known as *fitness sharing*[25] was implemented for the MP approaches, where an equal fitness score was assigned to each individual genotype, thereby assuming that each individual contributed to team performance equally. The advantage of this method was that fitness for individual genotypes was easily calculated, and there was no disparity between team fitness and the fitness of individual team members.

3.6 Cooperative behavior control experiment

To demonstrate that cooperative behavior was required in order for the prey to be immobilized, a control experiment testing eight different pursuit–evasion scenarios was executed. For all scenarios, a single prey, utilizing the fittest controller from the dynamic obstacle avoidance experiment, was placed in the simulated environment with a single predator. As with the standard predator controller evolution experiments (Sect. 3.4), the predator utilized was a self-teaching two-layered feed forward controller in which the weights were adapted either by an evolutionary process or also via lifetime learning.

The first scenario tested one prey and one predator using the SP approach for the evolution of prey capture behavior. The second scenario tested one prey and one predator using

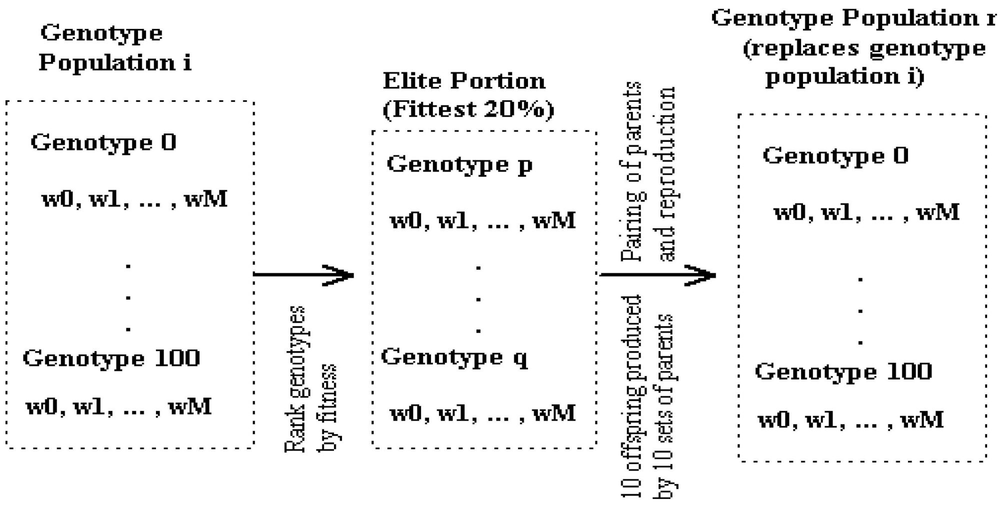

Fig. 5. The reproduction scheme used for the production of offspring predator genotypes. An elite portion (the fittest 20%) of genotypes is selected from the population at the end of the predator teams' lifetime; recombination and mutation operators were applied. The elite portion then produces offspring genotypes that replace the former population

the SP approach with lifetime learning. The third scenario tested one prey and one predator using the MP approach without lifetime learning. The fourth scenario tested one prey and one predator using the MP approach with lifetime learning. The next four scenarios tested the same experimental setup as the former four scenarios, except that two predators were tested instead of one. The configuration of the evolutionary process was the same as that used for the predator controller evolution (Sect. 3.4). After 10 replications of each experiment, where each replication was executed for 500 generations, results indicated that for all experimental setups, a single predator was unable to immobilize the prey for any period of time. However, two predators were able to immobilize the prey for short periods of time in some instances. The results of using the latter four experimental setups with two predators are presented in Sect. 4.

4 Comparison of approaches

Ten replications of experiments testing the SP and MP neuroevolution approaches were made, where each experimental replication ran for 500 generations. The predator team "lived" for 50 epochs. Each epoch consisted of 1000 cycles of simulation time. Each epoch constituted a test scenario where all predators and prey were tested for different, randomly generated orientations and starting positions in the environment. This specific configuration for the number of experimental replications, generations, epochs, and cycles of simulation time was utilized given the success of such experimental setups in previous evolutionary robotics experiments.[26–28] Ten different group configurations of predators and prey were tested for both the SP and MP approaches and their variants. These group configurations were named and defined as follows: group type 1, 2 predators and 1 prey, group type 2, 3 predators and 1 prey; group type 3, 4 predators and 1 prey; group type 4, 5 predators and 1 prey; group type 5, 6 predators and 1 prey; group type 6, 2 predators and 2 prey; group type 7, 3 predators and 2 prey; group type 8, 4 predators and 2 prey; group type 9, 5 predators and 2 prey; group type 10, 6 predators and 2 prey.

5 Evaluation of cooperative behavior

In order to quantify the effectiveness of emergent prey-capture strategies, two different measures were used to evaluate performance. The first was predator team fitness, where the fitness awarded to the team was equal to the period of time for which a prey was immobilized during the predator teams' lifetime. The second measure was a statistical index termed the group stability index. Adapted from Baldassarre et al.,[29] this index measured how stable a particular geometric formation of a group of predators was with respect to the prey for a given time period. Specifically, the index measured how much the relative position of each predator changed with respect to the other predators and prey for a given time in the simulation. For example, if the predators formed a circle about the position of a prey, the index indicated for how long the predators maintained this circle formation. If the predators were able to hold a certain formation for an extended period of time, the index was high, indicating high group stability.

6 Results

Ten sets of experiments were executed to test each group type, where each experiment set tested the SP and MP approaches and their variants.

6.1 Cooperative behavior performance

Figure 6 illustrates the average fitness attained for all predator group types using the SP and MP design approaches, and their respective variants. The fitness value presented for each group type equates with the average time for which each prey in each of the experiments was immobilized by the predator team. The average was taken over the 10 replications executed for each experiment. Figure 7 presents the average group stability index (GSI) attained for all predator group types using the SP and MP design approaches, and their respective variants. A GSI value close to 1 represented high group stability in terms of the predators forming and moving in a particular group structure whilst maintaining close proximity to a prey, whereas a GSI value close to zero indicated poor group stability and therefore relatively little or no group structure.

6.2 Evolved behavior

For all ten group types tested using the SP design approach, only two cooperative prey-capture strategies consistently emerged. These strategies, termed *entrapment* and *encirclement*, are illustrated in Fig. 8A and 8B, respectively, and briefly described below.

In the encirclement strategy, at least three, and at most four, predators moved to circle the prey, each moving in the same direction in close proximity to the prey. The predators would gradually move closer the prey, eventually forcing it to become immobile. The strategy was only successful for immobilizing a prey for relatively short periods of time, given that the predators were not able to coordinate their movements for extended periods.

Similarly, the entrapment strategy utilized at least three, and at most four, predators, where all moved simultaneously towards a prey from different directions in order to immobilize it within a triangular or square formation. As with the encirclement strategy, all predators remained in close proximity to the prey, except that they would "knock" against it in order to prevent its escape. The entrapment strategy was also hindered by a lack of coordination

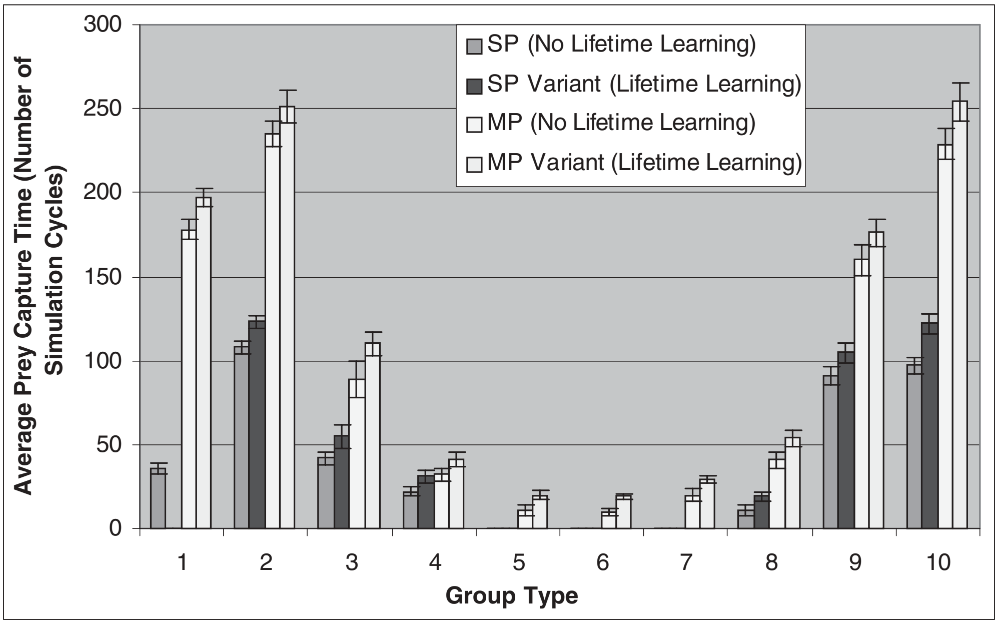

Fig. 6. Measured fitness for experiments testing 10 different predator group types (see text for explanation). Average fitness was taken over the 10 replications made for each experiment. *SP*, single pool design approach; *MP*, multiple pools design approach

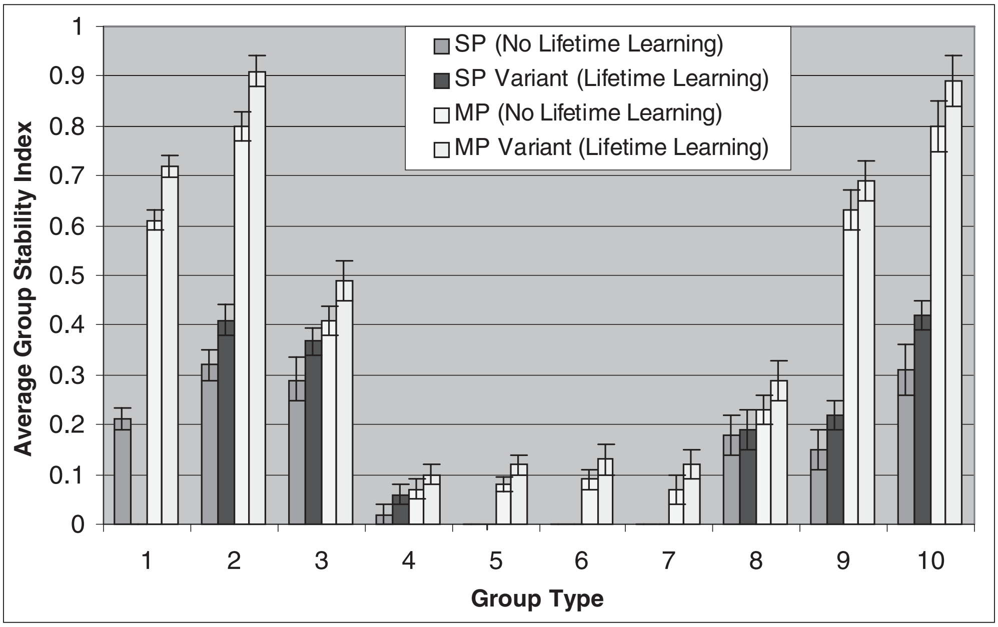

Fig. 7. Group stability index (GSI) measured for experiments testing 10 different predator group types (see text for explanation). Average GSI was taken over the ten replications made for each experiment

between the predators, and the number of prey-capture instances was less than in the encirclement strategy, although in an instance when a prey was cooperatively trapped, prey-capture time was longer compared to the encirclement strategy. For all ten group types tested using the SP variant design approach, only one emergent cooperative prey-capture strategy was observed. This was a derivative of the entrapment strategy and was termed *role switcher*.

In this strategy, a form of behavioral specialization emerged during the lifetime of the predators. The role switcher strategy used only three predators, where, as illustrated in Fig. 9, one predator moved to each side of the prey, while another predator, termed a *blocker*, moved around the flanking predators to approach the prey from the front, in order to immobilize the prey within a triangular formation. When the prey moved to escape, the flanking predators moved also, turning one way to force the prey in a specific direction. The blocker then moved around in order to confront the prey again. This system of entrapment, movement, and then entrapment continued for several times before the prey was able to evade the predators. Behavioral specialization was manifest such that a blocker predator always moved alongside one of the flanking predators, and the roles of flanker and blocker switched between two predators whenever a prey tried to evade the predators. This dynamic adoption and switching of roles served to

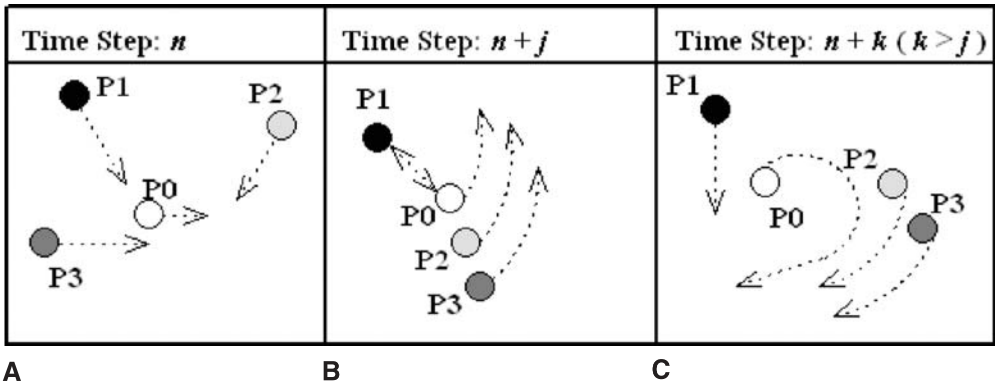

Fig. 8. An example of the role-switcher strategy that emerged under the SP variant approach and the MP approaches. P0, prey; P1, predator 1 (role knocker); P2, predator 2 (role rlanker); P3, predator 3 (role flanker)

coordinate the movements of the predators and thus effectively slow or immobilize a prey's movement.

Observing the fitness (prey capture time) results presented in Fig. 6, it is obvious that experiments using a predator to prey ratio of 2:1 (group types 2 and 10) yielded a greater performance, in terms of prey capture time, than the SP approach (under which the entrapment and encirclement strategies emerged).

For all ten group types tested using the MP and MP variant design approaches, only one emergent cooperative prey-capture strategy was observed. As with the SP variant experiments, the role switcher strategy was the only strategy that consistently emerged, although a specific difference was noted. Different predators adopted different roles from the beginning of their lifetimes. Specifically, particular predators always assumed the role of a *flanker*, while other predators always assumed the role of a blocker or that of an idle predator. This adoption of roles, maintained throughout the lifetime of the predators, served to increase the effectiveness of the role switcher strategy, and thus the performance of the MP approaches. Observing Fig. 6, it can be noted that the MP and MP variant approaches yielded a greater performance than the SP approaches. Also, the MP role switcher strategy was consistently effective at immobilizing a prey for all group types tested.

7 Analysis and discussion

In this section the cooperative prey capture strategies that emerged under the SP and MP approaches tested are discussed. The discussion relates the relative success of emergent prey-capture strategies to the group types tested and the performance measures defined.

SP. Two cooperative prey-capture strategies, each using at least three predators, consistently emerged for all except three predator group types. As is evident in Figs. 6 and 7, compared to the SP variant and MP approaches, these strategies performed poorly in terms of average team fitness. That is, for all except group types 1 and 8, the fitness of teams operating under the SP approach was lower than that of teams operating under the SP variant approach. The low fitness and group stability of the encirclement and entrapment strategies was a result of physical interference that occurred as three or more predators collectively approached a prey. This result was found to be due to confused infrared sensor readings of predators in close proximity to each other, and the fact that individual predators did not possess any memory, explicit form of communication, or coordination to facilitate a successful cooperative prey capture strategy. As illustrated in Figs. 6 and 7, this result was especially prevalent for group types 5, 6, and 7, where the minimum or maximum predator to prey ratio was utilized.

SP variant. In experiments using this approach, the role switcher strategy emerged for all except four of the group types tested. In the role switcher strategy, a form of dynamic behavioral specialization emerged in groups of at least three predators. This behavioral specialization was in the form of dynamic role adoption that emerged during the lifetime of a predator, and varied from predator to predator depending upon the group type being tested. This dynamic role adoption facilitated cooperation between the predators, affording the predator team a high average fitness and group stability compared to the SP approach. As presented in Figs. 6 and 7, this result was true for all group types tested except 1, 5, 6, and 7. The role switcher strategy emerged only when the predator to prey ratio was at least three predators to one prey. Hence, it did not emerge for group types 1 and 6, and it is theorized that for group type 5 the number of predators collectively attempting to immobilize a prey was too high, and for group type 7 the number of predators was too low to manage the successful immobilization of two prey. A dynamic assumption of roles during a predator's lifetime allowed for the formation of subgroups in a predator team, yielding a high team fitness and group stability for group type 10. As is evident from Figs. 6 and 7, the fitness and group stability attained for group type 10 was comparable to that attained for group type 2, where the ratio of predators to prey was also 3 to 1. The lower fitness and group stability evident in experiments testing group types 3, 4, 5, 8, and 9, compared to experiments testing group types 2 and 10, was a result of multiple predators attempting to assume the same behavioral role. That is, the strategy only persisted if there were three predators, two assuming the roles of flankers and one assuming the role of a knocker. Additional predators served only to confuse the strategy and cause it to fail.

It is theorized that behavioral specialization for the role switcher strategy emerged as an indirect result of interference that occurred when at least three predators collectively approached a prey. Such interference was observed in the SP experiments and often caused emergent strategies to fail prematurely, thereby making it more difficult for such strategies using at least three predators to be selected for, and propagated by, the evolutionary process. Whereas the role switcher strategy was able to achieve a high team fitness and group stability (evident in Figs. 6 and 7), as with the strategies emergent under the SP approach, it was at best only able to slow the prey, and never completely immobilizing it in experiments testing group types 5, 6, and 7 (Fig. 6). Figure 6 also presents a higher fitness for the SP variant (under which the role switcher strategy emerged) compared to the SP approach for all except group type 1. This result is readily explained, since this group type tested only 2 predators, and 2 prey, and the role switcher was only ever observed forming with 3 predators. Thus, the dynamic adoption of behavioral roles that defined the role switcher strategy only proved effective in an environment with an exact predator to prey ratio of 3:1. The interference that occurred between predators as they collectively approached a prey prevented the role switcher strategy from immobilizing a prey in other instances. However, the lifetime learning in the SP variant approach was able to exploit an environment with two prey via the dynamic formation of two specialized predator subgroups.

MP. In experiments testing the MP approach, only one cooperative prey-capture strategy emerged. This strategy was classified as a derivative of the role switcher strategy (observed in the SP variant experiments) and utilized at least two, and at most six, predators. The MP version of the role switcher strategy prey capture used a genetic-based specialization. In the SP variant experiments, the adoption of specialized behaviors was dependent upon the positions of the predators at a given time, whereas in the MP experiments, different predators initially behaved differently and assumed genetically predetermined roles prior to engaging in the role switcher strategy. That is, one predator always assumed the role of the blocker, whilst others always assumed a flanking role, or that of an idle predator. Idle predators served the purpose of reducing any potential interference between predators as they collectively approach a prey, and also increased the chance of success of the prey capture strategy by limiting the number of predators that constituted the strategy. The fact that the predators were genetically different is one explanation for the evolution of specialized behavioral roles that complemented each other in the emergence of the MP role switcher strategy.

In experiments testing group type 1, both predators in the simulation developed the complementary behaviors of flanker and knocker. In experiments testing group type 2, the three predators in the simulation developed the complementary behaviors of flanker, knocker, and idle predator. In experiments testing group type 3, one predator evolved the role of the knocker, another the role of an idle predator, and the other two both evolved the role of a flanker. Two predators having the same set behavioral role was detrimental to the role switcher strategy in that both contributed to interference within the team strategy (reflected in the results for group type 3 compared to other group types in Figs. 6 and 7). That is, if both flankers were within sensor range of the predator, and collectively attempted to approach the prey and simultaneously assume a flanking role, the strategy dispersed and the prey would escape. In experiments testing group type 4, one predator evolved the role of the knocker, another the role of an idle predator, and the other three evolved the role of flankers. As with experiments testing group type 3, the same problem of physical interference between team members with the same behavioral role occurred. This is reflected in Figs. 6 and 7 by an even lower team fitness and stability compared to that of group type 3. In experiments testing group type 5, one predator evolved the role of the knocker, another two the roles of an idle predator, and the other three evolved the roles of flanker. As can be highlighted in Figs. 6 and 7, the results were comparable to those of the experiment testing group type 4.

In the experiment testing group type 6, one predator evolved the role of the knocker and the other evolved the role of a flanker. However, as reflected in the results presented in Figs. 6 and 7, the presence of a second prey detracted from the effectiveness of the role switcher strategy given that if the second prey was in close proximity to the first, then there was an increased chance that separate predators would follow separate prey. In the experiment testing group type 7, one predator evolved the role of a knocker, the second evolved the role of a flanker, and the third evolved the role of an idle predator. As can be highlighted in Figs. 6 and 7, the results were similar to those of the experiment testing group type 6. In the experiments testing group types 8, 9, and 10, two specialized subgroups of predators always emerged. In the case of group type 8, two subgroups, each consisting of two predators taking the role of a knocker and a flanker, emerged. In the case of group type 9, the same two subgroups once again emerged, and the fifth predator developed the role of an idle agent. The results were the same for the case of group type 10, except that two predators developed the role of an idle agent. The effectiveness of teams comprising two specialized subgroups is illustrated in Figs. 6 and 7 (group types 8, 9, and 10), which present a high average fitness and group stability compared to other experiments testing scenarios with two prey agents.

It can thus be noted that the MP approaches were more effective than the SP approach at exploiting an environment containing two prey. Specifically, the MP approach facilitated the evolution of specialization at the individual level as well as at a subgroup level. That is, under the MP approach neural controllers evolved such that predators would assume a set behavioral role from the beginning until the end of their lifetimes. These behavioral roles were complementary, such that multiple predators in close proximity in collective pursuit of a prey were sufficient for the formation of an effective prey-capture strategy. The overall effectiveness of the MP prey-capture strategy, compared to those emergent under the SP approach, is evident from the higher fitness and group stability index presented in Figs. 6 and 7, respectively.

Additionally, observing the team fitness and stability results presented in Figs. 6 and 7, it can be noted that both the SP and MP variant approaches (those with lifetime learning) were consistently able to achieve a higher level of performance. These results are confirmed by previous results in evolutionary robotics experiments[26–28] that sought to investigate the impact of lifetime learning upon evolution in terms of deriving neural controllers to successfully accomplish a given task. In predator teams using the SP and MP approaches without lifetime learning, predator neural controllers evolved a general solution to the prey-capture task, and this solution was passed on through the genotype to successive generations. This general solution was not optimal because it could take into account the characteristics of particular prey-capture scenarios, for example, scenarios containing a single prey versus scenarios containing two prey. In the case of approaches using lifetime learning, it was theorized that the standard controller weights are incorporated the same general solution, and that learning was able to refine an inherited strategy by taking into consideration the specificity of a given predator–prey scenario. Hence, lifetime learning results imply that the genotypes (the inherited standard weights and the inherited teaching weights) of evolved predator controllers with lifetime learning incorporated not a predisposition to behave efficiently, but a *predisposition to learn* to behave efficiently.

8 Conclusions

This article has presented a set of experiments comparing two neuroevolution approaches, and their variants, for the synthesis of cooperative behavior within a team of simulated robots (termed predators), where the task of the team was to cooperatively immobilize one or two evader robots (termed prey). The performance of the team was tested for ten different group configurations of predators and prey. Team performance was measured in terms of *fitness* (calculated as the total time for which a prey was immobilized by a predator team over the course of the team's lifetime), and a *group stability index*. This index indicated how stable a particular geometric formation of predators about a prey was for a given period of time. The results presented indicated the MP approach to be superior in terms of these measures for all group types tested. The superiority of the MP approach was found to be a result of a genetic form of behavioral specialization that assigned behavioral roles at the beginning a predator's lifetime. The MP approach also facilitated the evolution of specialized subgroups of predators in scenarios using two prey. These specialized subgroups aided in reducing physical interference between predators as they collectively approached a prey. An analysis of emergent strategies revealed that behavioral specialization was a necessary aspect of the emergence of effective cooperative behavior in the described task domain. This was especially evident in the SP experiments, where a low fitness and group stability was observed. This was a result of physical interference occurring between predators as they collectively approached a prey. Also, approaches using lifetime learning were found to yield a superior performance compared to the same approaches without lifetime learning. Previous evolutionary robotics experiments[26–28] support this result where neural controllers with lifetime learning evolved a predisposition for learning effective task accomplishment.

A comparison with other emergent cooperative behavior approaches in the pursuit–evasion domain[11,12,18] is difficult given the real-world nature of the experiments described here. That is, although the robots were simulated, the environment was a continuous domain and the simulation incorporated noise in sensory data, namely confused infrared sensor readings resulting from two or more Kheperas being in close proximity to each other. These noisy sensor data were a key reason for interference between multiple predators as they collectively approached a prey. Also, a continuous environment does not allow for the selection of distinct sets of situation/action values that are possible in grid-world implementations[18] where a finite set of actions and resultant outcomes can be defined. However, the similarity in results with other evolutionary robotics experiments placed within real-world[3] and simulated[29] task environments that mandate cooperative behavior elicits support for the supposition that the adoption of specialized and complementary behaviors by different agents in a team facilitates emergent cooperative behavior.

Finally, experimental results highlighted that the MP neuroevolution methodology is an effective method for deriving specialized behaviors that facilitate emergent cooperative behavior in a simulated multirobot system with no explicit communication or coordination mechanisms.

References

1. Mondada F, Franzi E, Ienne, P (1993) Mobile robot miniaturization: a tool for investigation in control algorithms. Proceedings of the 3rd International Symposium on Experimental Robotics, Kyoto, Japan, October 28–30, 1993, pp 501–513
2. Iba H (1996) Emergent cooperation for multiple agents using genetic programming. Parallel problem solving from nature. Springer, Berlin
3. Quinn M (2000) Evolving cooperative homogeneous multi-robot teams. Proceedings of the International Conference on Intelligent Robots and Systems (IROS 2000), Takamatsu, Japan, October 31–November 5, 2000, pp 1798–1803
4. Axelrod R, Hamilton W (1981) The evolution of cooperation. Science 211(4489):1390–1396
5. Dugatkin L (1990) *N*-person games and the evolution of cooperation: a model based on predator inspection in fish. J Theor Biol 142:123–135
6. Axelrod R (1984) The evolution of cooperation. Basic Books, New York
7. Akiyama E, Kunihiko K (1995) Evolution of cooperation, differentiation, complexity, and diversity in an iterated three-person game. Artif Life 2:293–304

8. Oliphant M (1994) Evolving cooperation in the non-iterated prisoner's dilemma: the importance of spatial organization. Proceedings of the 4th Artificial Life Workshop, Cambridge, USA, October 28–30, pp 349–352
9. Epstein J, Axtell R (1996) Growing artificial Societies: social science from the bottom up. MIT Press, Cambridge
10. Pfeifer R, Scheier C (1999) Understanding intelligence. MIT Press, Cambridge
11. Haynes T, Sen S (1996) Evolving behavioral strategies in predators and prey. Adaptation and learning in multi-agent systems. Lecture notes in computer science. Springer, Berlin, pp 113–126
12. Yong C, Miikkulainen R (2001) Cooperative co-evolution of multi-agent systems. Technical Report AI01-287. Department of Computer Science, University of Texas, July 2001
13. Nishimura S, Ikegami T (1997) Emergence of collective strategies in a prey–Ppedator game model. Artif Life 3:243–260
14. Gomez F (2003) Robust non-linear control through neuroevolution. PhD Thesis, Department of Computer Science, University of Texas
15. Benda M, Jagannanthan V, Dodhiawalla R (1985) An optimal cooperation of knowledge sources. Technical BCS-G2010-28, Boeing AI Center, August
16. Korf R (1992) A simple solution to pursuit games. Working Papers of the 11th International Workshop on DAI, Geneva, Switzerland, February 25–29, 1992, pp 195–213
17. Levy R, Rosenschein J (1992) A game theoretic approach to distributed artificial intelligence and the pursuit problem. Decentralized AI III, Springer, Kaiserslautern, pp 129–146
18. Denzinger J, Fuchs M (1996) Experiments in learning prototypical situations for variants of the pursuit game. Proceedings of the 2nd International Conference on Multi-Agent Systems (ICMAS-96), Kyoto, Japan, December 9–13, 1996, pp 48–55
19. Nolfi S (2000) Evorobot 1.1 user manual. Technical report. Institute of Cognitive Sciences and Technologies, National Research Council of Italy, July
20. Rumelhart D, Hinton G, Williams R (1986) Learning internal representations by error propagation. Parallel Distributed Processing, vol 1. Foundations. MIT Press, Cambridge
21. Miglino O, Hautop H, Nolfi S (1995) Evolving mobile robots in simulated and real environments. Artif Life 2:417–434
22. Nolfi S, Floreano D (2000) Evolutionary robotics. MIT Press, Cambridge
23. Jakobi N, Husbands P, Harvey I (1995) Noise and the reality gap: the use of simulation in evolutionary robotics. Proceedings of the 3rd European Conference on Artificial Life (ECAL-95), Granada, Spain, June 4–6, 1995, pp 704–720
24. Eiben A, Smith J (2003) Introduction to evolutionary computing. Springer, Berlin
25. Bull L, Holland O (1997) Evolutionary computing in multi-agent environments: eusociality. Proceedings of the 2nd Annual Conference on Genetic Programming, San Francisco, USA, July 13–16, 1997, pp 347–352
26. Nolfi S, Elman J, Parisi D (1994) Learning and evolution in neural networks. Adapt Behav 3:5–28
27. Nolfi S, Floreano D (1999) Learning and evolution. Autonomous Robots 7:89–113
28. Nolfi S (2000) How learning and evolution interact: the case of a learning task which differs from the evolutionary task. Adapt Behav 7:231–236
29. Baldassarre G, Nolfi S, Parisi D (2002) Evolving mobile robots able to display collective behaviors. Proceedings of the International Workshop on Self-Organization and Evolution of Social Behavior, Monte Verità, Switzerland, Sepember 8–13, 2002, pp 11–22