

Emergence of Diversity in a Group of Identical Bio-robots

Regular Paper

Alessandra Vitanza^{1*}, Luca Patanè¹ and Paolo Arena¹

¹ University of Catania, Catania, Italy

*Corresponding author(s) E-mail: avitanza@dieei.unict.it

Received 19 November 2014; Accepted 20 March 2015

DOI: 10.5772/60545

© 2015 Author(s). Licensee InTech. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Learning capabilities, often guided by competition/cooperation, play a fundamental role ubiquitously in living beings. Moreover, several behaviours, such as feeding or courtship, involve environment exploration and exploitation that include local competition and lead to a global benefit for the colony. This can be considered as a form of global cooperation, even if the single agent is not aware of the overall effect. This paper aims at demonstrating that identical bio-robots, endowed with simple neural controllers, can evolve diversified behaviours and roles when competing for the same resources in the same arena. These behaviours produce also a benefit in terms of time and energy spent in the whole group. The robots are asked for a classical foraging task structured through the cyclic activation of the resources. The result is that each single robot, while competing to reach the maximum number of available targets, tends to prefer a specific sequence of sub-tasks. This indirectly leads to the global result of task partitioning, where the cumulative energy spent, in terms of the overall travelled distance and time needed to complete the task, tends to be minimized. A campaign of simulation experiments is conducted using different number of robots and scenarios: the common emergent result obtained is the role Specialization of each robot. The description of the neural controller and specialization mechanisms are reported in details and discussed.

Keywords Bio-robotics, Multi-robot Systems, Collaborative Algorithms, Dynamic Simulation, Labour Division

1. Introduction

Coordination and collaboration among robots are the results of self-organized behaviours: social insects provide brilliant solutions for foraging, migration, mating and others [1, 2]. Transferring these characteristics to future biorobotic systems will assure both flexibility in space- and time-varying environments and high robustness to faults in the single agents [2, 3]. On the other hand, even within the same ecological niche, individuals of the same species compete for resources. This is mostly clear in simple insects like flies, which do not show apparent cooperation capabilities, but indeed compete for food and mating [4].

Indeed the boundary between cooperation and competition is rather subtle: in a sense it can be argued that simple brains mainly compete for resources. Such a competition is of course mediated by the environment: limited resources can be exhausted by other individuals even if a cycle of regeneration can be considered. The single agent behaviour and the environment co-evolve, within the agent life cycle, to reach a global equilibrium for the colony. The environment acts so as to shape the local competitive behaviour of

the single agents, to give rise to a global cooperative strategy, leading to an equilibrium state.

An open question in animal social behaviour is related to the existence of a kind of social brain guiding the individual behaviours through the environment: could global order emerge from the local behaviour of the agents which simply compete for survival? We tried to answer to this question demonstrating through a series of simulation campaigns how a simple form of cooperation (i.e. task partitioning), can arise in a small number of competitive roving robots, endowed with the same neural controller. The robots succeed in adapting to different sequences of environment induced stimuli. Here there is no need, in principle, for a kind of super organism. The environment imposes rules and the global benefit for the group of robots can arise from the local competition mechanisms among equally endowed individuals, even in absence of a direct communication among them. Furthermore, the definition of a series of tasks is frequently met in Nature: there are different activities that have to be performed in given time windows during the daily cycle, and a task division among the individuals is requested, even if all of them are equally able to perform all the tasks [5].

The proposed approach starts from the results of previous works [6, 7, 8], where the Specialization strategy was introduced and a first collaborative algorithm was formalized to fulfil an overall mission. Many research activities face with specialization and the relationships between reward mechanism, behavioural diversity and optimization of performance [9]. The aim of our work is to investigate and quantify the influence of the environmental mediation on an operant conditioning at the level of the single agent. The open problems, we want to focus on, regard the emergence of task partitioning induced by the environment and how the obtained solutions are robust to the robot starting positions. By unravelling the interaction among robots and the environment we identified the variables that lead to different solutions, proposing performance indexes to evaluate the final robot behaviour. The scenario considered in this work is a basic foraging task performed in a simulation environment with different arrangements of food sources and starting position of robots. When a target is retrieved, another one becomes active following a predefined sequence, as an environment-dependent rule. The role of the environment is important to indirectly influence the single agent behaviours depending on the other elements of the group. In the proposed work the task partitioning is addressed, allowing a decrease of the energy spent by the whole group while accomplishing the task. The results show that agents evolve modelling their own basically competitive capabilities to perform globally collaborative strategies, starting from a homogeneous initial situation and exploiting the environmental mediation.

2. The Spiking-based Neural Controller

A brief overview of the Control Algorithm and the Neural Network (NN) controller endowed in each robot are here reported. Further details on Specialization learning and NN model are available in [7, 8, 10].

2.1 The Algorithm

In the experimental setup, the environment contains differently coloured targets on the floor, which are cyclically activated in a mutually exclusive way. In particular, at the beginning of the cycle only the first target of the sequence is visible; when that target is retrieved, the second one becomes active, following the predefined sequence. This mechanism permits to obtain an environment-dependent rule that guides the targets exploration. At the end of the sequence, a Reward signal (Rw) is activated and then the cycle begins again. The single robots, while competing to forage the available targets, indirectly perform a task partitioning, exploiting the environmental mediation.

Each robot starts with the same ability to identify and reach all the targets in the arena. If a target is present in the environment the robots move toward it with a fixed speed, otherwise they rotate looking for targets, performing a fixed clockwise rotation on the spot (about 45). These rules imply that a target can be reached by a more distant but well oriented robot than by another which is nearer but badly oriented. No direct communication is introduced among the agents: they compete for reaching the same target once this is activated. When the cycle is concluded and the Rw is activated for all robots, a learning phase is then performed. Following the training algorithm described below, each robot increases its attractiveness for the targets reached, contemporary decreasing its interest for the non-reached ones. The final result, emerging from this scenario, is a spontaneous labour division among the robots, which become refractory to those targets they are not able to reach. This learning mechanism can be considered as an agent distributed operant conditioning: it acts in different ways on the neural architectures in each robot, leading to the development of different skills. The presence of a global reward signal contributes to generate diversity and specialization. The targets are here circular spots on the ground, able to be reached by all the robots: due to the collision avoidance strategies embedded into the robot control system, competitive behaviours are encouraged. The environmental setup and the other robots play a fundamental role, biasing the final behaviour of each single robot. The block diagram of the Control System Architecture is shown in Fig. 1, where a spiking-based neural controller selects the robot behaviour depending on the information acquired by the sensory system. The retrieved targets are stored in a two-level memory. When the sequence of targets is completed the reward signal, generated by the environment, triggers the learning process. This consists, in our case, in a Threshold adaptation applied to the vision neurons, depending on the data

stored in the memory that is then emptied to be ready for the next foraging session. In particular, the *Short-Term Memory (STM)* permits to memorize the targets retrieved during the activation sequence and acts every R_w activation; instead the *Long-Term Memory (LTM)* retains visited targets for longer time providing low-frequency adjustments in the threshold adaptation.

2.2 The Neural Model

The computational structure acting as a nonlinear controller embedded in each robot, was derived from modelling the learning mechanisms in the fruit fly *Drosophila melanogaster*. Within the insect brain, the Mushroom Bodies (MBs) and the Central Complex (CX) are the most studied neural assemblies for their enhanced characteristics in olfactory and visual learning; for example, rewarding and punishing olfactory associations were peculiarly addressed into the MBs of the insect brain [11, 12]. Efficient computational models were recently designed and implemented, which resulted useful for addressing more complex behaviours like attention, expectation and decision making [13]. On the other side, visual learning and visual targeting were addressed in the CX. A complete, updated insect brain computational architecture was recently presented in [13]. Here the neural controller is a reduced model of the whole architecture, retaining the essential features needed for the assigned task.

The developed control architecture is a multi-layer neural network similar to one fan of the Fan-Shaped Body, an area within the CX devoted to visual feature discrimination and memory [13]. The structure uses a class I Izhikevich neuron model [14] and is made up of two modules: one for visual

target recognition, and another for obstacle avoidance. In fact, the fruit fly CX receives information also from the mechanoreceptors distributed on the body. Moreover, output neurons in our architecture directly act on the motors through the mediation of inter-neurons. This also resembles the CX which is connected with pre-motor areas for controlling locomotion in both tactic and phobic behaviours elicited by the visual sense.

The peculiarity of this structure is to allow learning both in the synaptic links among the neurons and in the threshold of each neuron. However, in the simulations presented here, synapses are already learned and fixed: a Hebbian learning method (STDP - Spike Timing Dependent Plasticity) was already applied, as shown in [10], to let all the robots show tactic behaviour for all the targets present in the environment. In particular, the module devoted to visual and target recognition was split in as many Sensing and Specialization Modules (S&SM) as the type of targets present in the environment (see Fig. 1).

The learning process obtained through Threshold adaptation Neuron (T_aN), was applied to induce hyperpolarization or depolarization into the visual neurons within each S&SM sub-group, to make them responsive only to a specific class of targets, as shown in [7].

The modified equation of neuron model, used for a generic T_aN , is here reported:

$$\begin{aligned} \dot{v} &= 0.04v^2 + 5v + 140 - u - g_a V_{th} + I_i \\ \dot{u} &= a(bv - u) \end{aligned} \quad (1)$$

with the spike-resetting

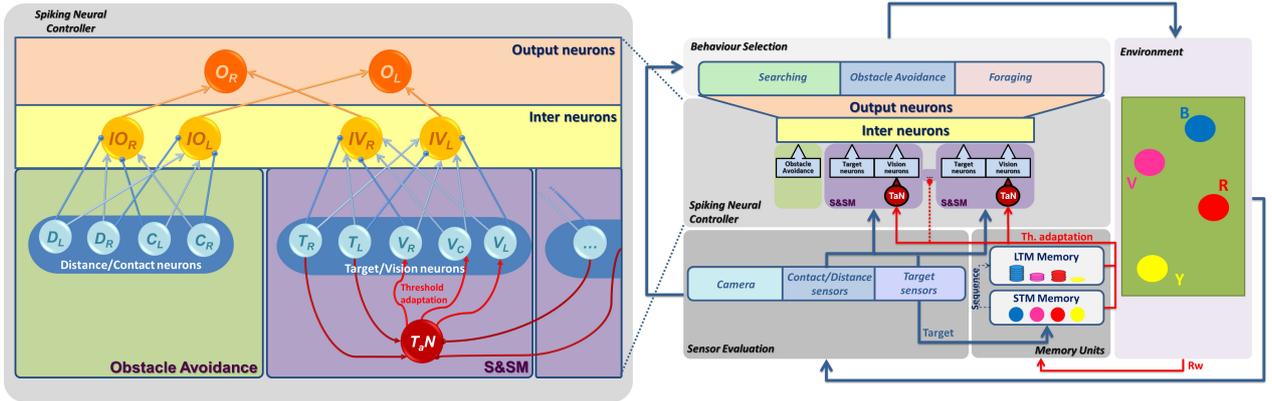


Figure 1. Block diagram of the control architecture. In *Sensor Evaluation* block, data from the sensors are collected. The *Spiking Neural Controller* block describes the neural network model used for the foraging. The number of *Sensing and Specialization Modules (S&SM)* is related to the different targets to recognize. T_aN is the Threshold adaptation neuron, devoted to induce Specialization learning. The threshold adaptation, applied to the neural structure, is here guided by two different memories: *Short-Term Memory (STM)* and *Long-Term Memory (LTM)*. The *Behaviour Selection* block indicates the most suitable action to be executed in the dynamic simulation Environment. Details of the spiking neural network are shown in the left-side: the three-layer structure permits to transduce inputs coming from target sensors and from vision sensors to guide the behavioural selection, the Obstacle Avoidance block induces reactive behaviours in order to avoid collisions. In the input layer we can distinguish four classes of input neurons Contact (C_x), Distance (D_x) related to the Obstacle Avoidance block and Target (T_x), Vision (V_x) related to S&SM, where $x=L, R, C$ indicates Left, Right and Central side. In the inter layer the sensory information are collected in the Obstacle and Vision classes (IO_x and IV_x) and finally the output layer contains two neurons that are used to control the speed of the Left and Right motors of the system (O_x).

$$\text{if } v \geq 0.03, \text{ then } \begin{cases} v \leftarrow c \\ u \leftarrow u + d \end{cases} \quad (2)$$

where: $a=0.02$, $b=-0.1$, $c=-55$, $d=6$; v , u are, respectively, the neuron membrane potential and the recovery variable. The pre-synaptic input current is composed of two terms: an adaptation parameter $g_a V_{th}$, that is voltage-dependent [15] and the external (sensory based) and internal (neuron synaptic connections) inputs I_i . The time unit is ms and $g_a=1$.

Variations of V_{th} produce neuron facilitation or hyperpolarization, depending on events occurred before the reward-signal activation (Rw) for the STM factor, and on visiting history for the LTM factor, according to the formula:

$$V_{th}(k+1) = \begin{cases} V_{th}(k) + \alpha \Delta V_{STM}(k) + (1-\alpha) \Delta V_{LTM}(k) & \text{if } Rw=1 \\ V_{th}(k) & \text{otherwise} \end{cases} \quad (3)$$

where k is the current simulation step and α is a parameter used to mediate the effect of the short- and long-term memory in the threshold adaptation mechanism. In particular, each S&SM block contains a series of *vision neurons*, each one emitting spikes only if a specific coloured target is detected in the scene. This block also contains *target neurons* which spike only when a specific target is reached by the robot. This mimics, for example, sugar sensors placed in the fly legs, responding only when reaching sweet surfaces. When $\alpha=1$ the Threshold adaptation is exclusively guided by the STM and it takes place when the Reward signal is triggered. In this case all the threshold adaptation neurons, corresponding to the targets retrieved and stored in the memory, are depolarized whereas all the others are hyperpolarized. The reward signal acts as a bias on $T_a N$, which adds a contribution to the threshold V_{th} of all the vision neurons within the block:

$$\Delta V_{STM}(k) = \begin{cases} \Delta V_D & \text{for depolarization.} \\ \Delta V_H & \text{for hyperpolarization.} \end{cases} \quad (4)$$

with $\Delta V_D=1.8$; $\Delta V_H=-0.6$.

$$\Delta V_{LTM}(k) = -H \left(-\frac{m}{N} \sum_{i=0}^N \left(\text{sign}[\Delta V_{STM}(k-i)] \right) \right) \quad (5)$$

where $N=5$ is the sliding window of Rw events, $m=0.2$ is a gain and H is the heaviside function. This is a key aspect of the learning procedure which acts according to the principle of local activation and global inhibition, as explained in the following.

Although all neurons start with the same value of $g_a V_{th}=20$, it can be modified within the range:

$0 \leq g_a V_{th} \leq 22$. Moreover, a target becomes no longer attractive when the bias goes below a lower bound that here corresponds to $I_{ina}=g_a V_{th}=14$. Below this value the vision neuron does no longer emit spikes, even if the corresponding target is within the visual field. Introducing the LTM factor ($\Delta V_{LTM}(k)$), the Threshold adaptation is modulated by the history of arena explorations during the cycles. In fact, this memory hyperpolarizes neurons depending on the number of cumulated target visits. It is useful especially in competitive scenarios where the wide competition among robots compromises performance.

3. Simulation Results and Performance Analyses

Experimental simulations were performed using our own software/hardware framework and Dynamic Simulator (SPARKRS4CS, details in [16, 17]). Each robot is a simulated version of TriBot I [18], a hybrid robot developed to investigate cognitive capabilities inspired by insects, with dimension $0.3m \times 0.2m$ moving in an arena of $3m \times 2m$ for square arenas and $3m$ diameter for circular arenas with targets on the floor (see Fig. 2). The use of an accurate dynamical simulation environment and the interest to study individual behaviour modulations implied to consider small groups, in order to analyse specific emergent behaviours in the single robots. For this reason, we limited the group to three and four robots. Moreover, the number of targets was fixed to four to evaluate the differences in Specialization and global performance. Scenarios with a perfect match between the number of robots and targets and scenarios where the resources exceed the number of robots were then analysed. Different target arrangements and various activation sequences are used for simulation tests. The reward signal is activated when the last target is reached by a robot.

A number of different simulations have been performed with three and four robots, for a total of more than 100 simulation tests and results are summarized in Table 1, Table 2 and Table 3. The campaign was conducted to investigate if the environment, merely imposing a fixed sequence, indirectly induces robots to select a subset of targets to minimise the travelled space to complete the sequence. This minimum energy solution to the task was investigated using only the *Short-term memory* ($\alpha=1$). In details, for each of the arenas reported in the following Tables, five different activation sequences are used (see captions for details) and the corresponding solutions are considered to statistically evaluate the overall distance travelled. Besides the evaluation of the energy spent by the group, the optimization in terms of time to complete the sequence is evaluated considering the time interval between two consecutive Rw activations. During the initial simulation trials the arrangement of targets was absolutely random, but the obtained results can be collected into three classes, depending on the target spatial configuration: *asymmetric*, *competitive* and *symmetric* arenas (see Table 1, 2,

3 respectively). The first row of each Table contains the arena description, for each configuration an amount of at least 10 simulations were performed starting from different initial robot positions. Results are splitted in sub-tables where robot-targets solutions, the time spent to obtain a reward and the covered space (defined as normalized distance travelled by the group of robots) are reported for three and four robots. From the analysis of the simulations, it can be noticed that the emerging solutions directly depend on the target position distribution. For example, in the first activation sequence related to the first arena (Table 1[A1]), the targets close one another will be covered by the same robot, with the emergence of a robust solution which does not depend on the activation sequence. Both trends, the time effort to obtain a reward activation and the global space covered by robots, are minimized.

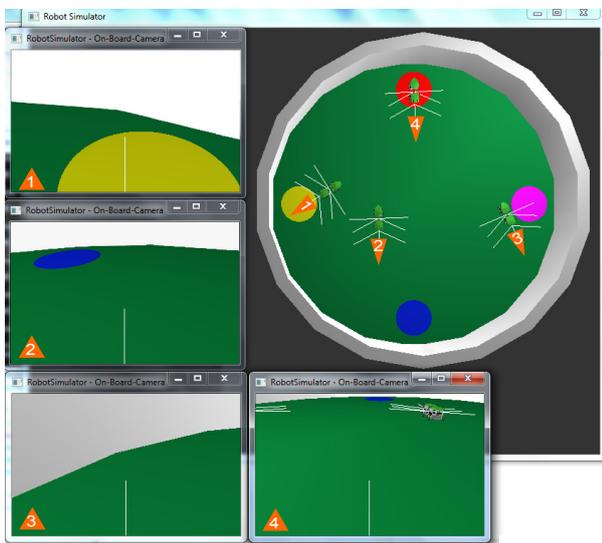


Figure 2. The Dynamic Simulator Overview. In this scenario a group of four robots is placed in a circular arena where four targets are symmetrically distributed on the ground. The camera view of each robot is also reported. The white lines around the robots represents the distance sensors: the lowest detected distance for each side is used as input to the neural network.

A different situation arises in the scenario reported in Table 2 ([A3] and [A4]), where the yellow target position is shifted from the blue toward the red one. In such cases the solution is dependent, besides on the target position, also on the activation sequence. Under these conditions, for situations where the number of targets exceeds the amount of robots, competitive behaviours are massively promoted and it is clearly shown by solutions: each robot is attracted by more than one target, and the targets are shared by more robots (see Table 2, case: 3 Robots). This situation may be useful for fault tolerances issues, in fact, sharing targets the robots could follow the activation sequences after a fault of some robot; on the contrary the group does not reach the complete specialization and often competitive behaviours compromise the minimization of travelled space and of the time needed to obtain a Rw. Similar considerations can be

carried out for the last two cases in Table 3 that report two symmetric target configurations. In both cases experiments with 3 robots show that the travelled space decreases, whereas the time interval between two Rw activation shows a disorder trend until it converges to a more regular shape. In particular, the symmetric configuration and position of targets in the arenas ([A5] and [A6]) induces more often obstacle avoidance reactions, due to competitive actions against other robots or arena walls. The results reveal that each graph intrinsically presents the same trend: at the beginning of the experiment, when the targets are shared, the distance covered by each robot is maximum; so the time passed between two reward-signal activations. As specialization emerges, the distance covered by robots is gradually minimized and the efforts are spread among individuals, as the period between two Rw activations becomes more regular. This is evident in any arena, for each sequence, even if the maximum and minimum limits and the convergence time vary according to the complexity of the environmental setup: it should be also noticed how the minimum asymptotic value reached by the curves is inversely proportional to the number of individuals in the group. In particular, simulations with four robots show a clear decrement of the global covered path, whereas in the experiments with three robots a more unclear trend can be seen. Furthermore, in some cases (i.e. symmetric arena [A5]) larger number of Rw events are needed to reach the convergence in terms of time spent to complete the task (about $300N_{Rw}^w$). The inherent asymmetry in the system generates solutions where some targets are shared by the robots, inducing a kind of disorder. Similarly, symmetric arenas largely emphasises competitive foraging, slowing down the specialization process. The performances of the robots in the different classes are shown in Fig. 3, where the improvement of the implemented learning strategy is reported in the different environmental configurations. The maximum improvement is obtained in the Asymmetric scenarios, where the absence of symmetric distribution generally boosts the dispersion of the individuals. So, these configurations facilitate the learning process avoiding to be trapped in local minima. This happens, by contrast, in the symmetric situation producing a minor performance enhancement. The competitive case can be considered in between, since the presence of targets shared by the robots facilitates the specialization with respect to the symmetric case, but induces more competitive reactions to the asymmetric counterpart.

Fig. 3 (b) shows also how the increment of the number of robots (from three to four) is relevant in terms of performances and strictly influences the learning process.

Hence, simulations with a number of robots (i.e. three robots) less than the number of targets (i.e. four targets) and with four robots, that match the number of targets in the arenas, let the opportunity to investigate different solutions in terms of robot specialization and global performance.

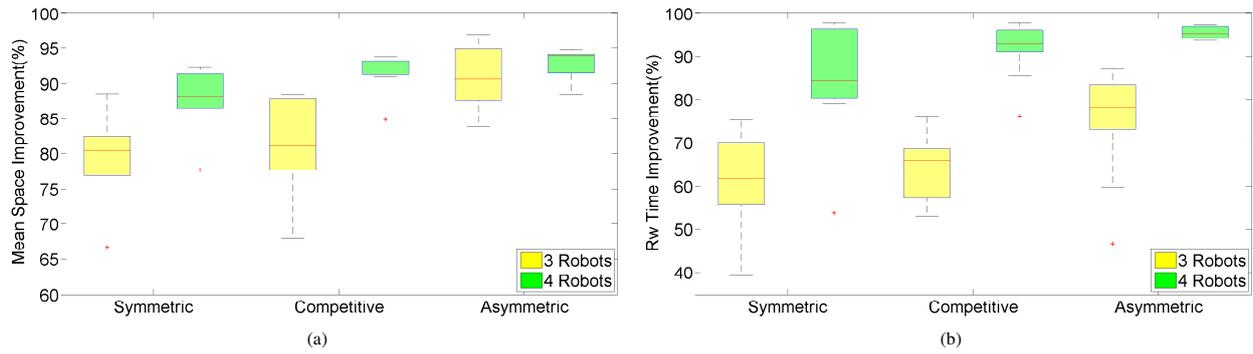


Figure 3. Evaluation of the improvement obtained through the learning process in terms of: (a) mean space covered by each robot to obtain a Reward signal (i.e. when all the targets are visited following the specific activation sequence); (b) time interval between two consecutive Rw signal activations. Results are classified on the basis of the number of robots and the class of scenarios: Symmetric, Competitive and Asymmetric.

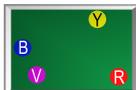
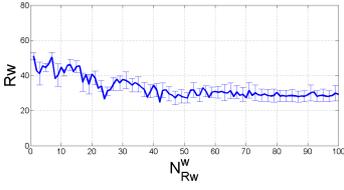
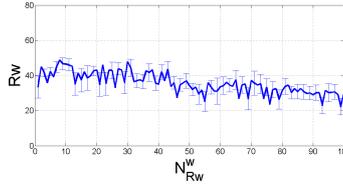
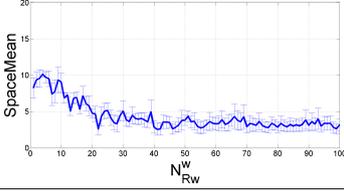
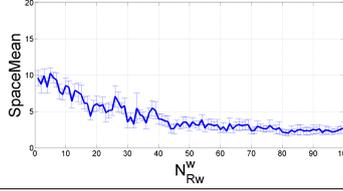
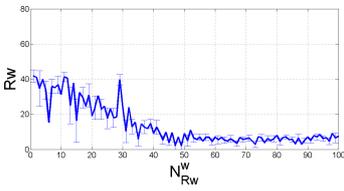
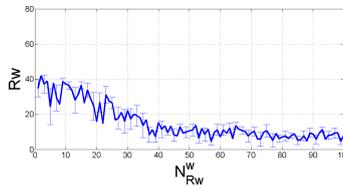
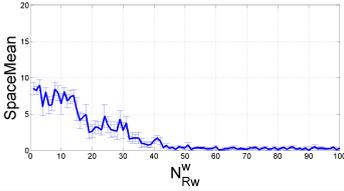
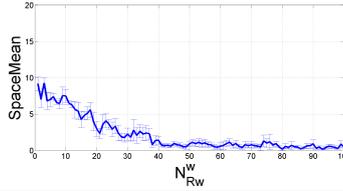
Arena Distances (m)	 $YV = 2.05$ $RV = 2.03$ $BY = 1.98$ $BR = 2.49$ $BV = 0.76$ $YR = 1.53$	 $BY = 0.65$ $BR = 2.46$ $BV = 1.28$ $YR = 1.95$ $YV = 1.41$ $RV = 1.98$
3 Robots		
Solution	$R_{B,V} - R_Y - R_R$	$R_{B,Y} - R_R - R_V$
Rw Time		
Space		
4 Robots		
Solution	$R_B - R_V - R_Y - R_R$	
Rw Time		
Space		

Table 1. Summary of simulation results with three and four robots in *asymmetric* arenas. Five different Activation Sequences are used in the campaign of experiments (B-Y-R-V, B-V-Y-R, B-V-R-Y, B-R-V-Y, B-Y-V-R). Legend: **Arena**: the environmental setup, **Solution**: the emerging solution, **Rw Time**: time interval between two consecutive Rw signal activation, **Space**: normalized distance travelled by the group of robots.

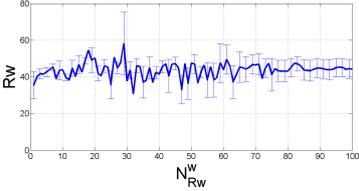
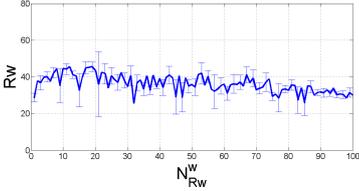
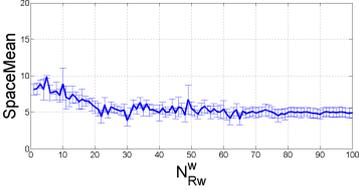
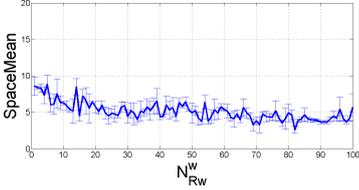
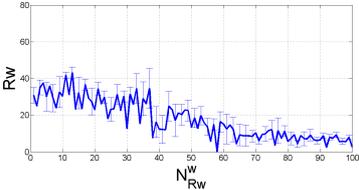
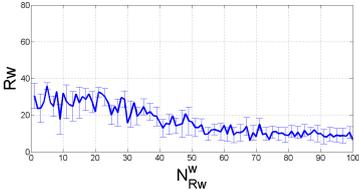
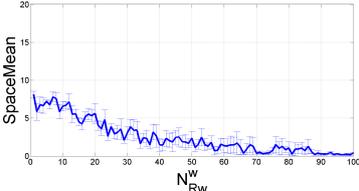
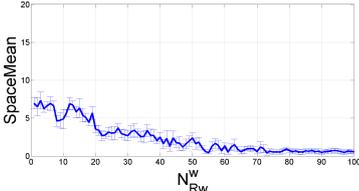
Arena Distances (m)	 [A3]	 [A4]
3 Robots		
Solution	$R_{B,Y} - R_{Y,R} - R_V$ $R_{B,V} - R_{Y,R} - R_{Y,V}$ $R_{B,V} - R_Y - R_R$ $R_B - R_{Y,R} - R_V$ $R_{B,Y} - R_{Y,R} - R_{B,V}$	$R_{B,V} - R_Y - R_R$ $R_{B,V} - R_{Y,R} - R_{Y,V}$ $R_{B,Y} - R_R - R_V$ $R_{B,Y} - R_{B,V} - R_{Y,R}$
Rw Time		
Space		
4 Robots		
Solution	$R_B - R_V - R_Y - R_R$	
Rw Time		
Space		

Table 2. Summary of simulation results with three and four robots in *competitive* arenas. Five different Activation Sequences are used in the campaign of experiments (B-Y-R-V, B-V-Y-R, B-V-R-Y, B-R-V-Y, B-Y-V-R). Legend: **Arena**: the environmental setup, **Solution**: the emerging solution, **Rw Time**: time interval between two consecutive Rw signal activation, **Space**: normalized distance travelled by the group of robots.

Regarding the evolution of specialization with three robots: initially two robots try to reach one target each, whereas the third robot is unable to specialize until the convergence of the other robots is almost reached. Thus, the emergence of specialization can be seen as a dynamic optimization process, constrained by the environment and by the robot interaction.

Even if the experiments involving four robots and four targets reach the most obvious solution (i.e. each robot specialises in one target), it is interesting to analyse the dynamics leading to this steady state condition. In [7] we introduced an index to quantify the dynamics of the global

performance of the whole system, taking as a reference the works [19, 20]. In particular, following the formula:

$$S(w) = \frac{N_s(w)}{N_i(w)} \quad (6)$$

This index is evaluated a posteriori. At the end of the simulation, the final division of labour is built of a series of subtasks s_i performed by each robot i . Within a given time window w , $N_s(w)$ represents the total number of events where the labour division task is composed of the sequences s_i . $N_i(w)$ is the total number of reward events in w .

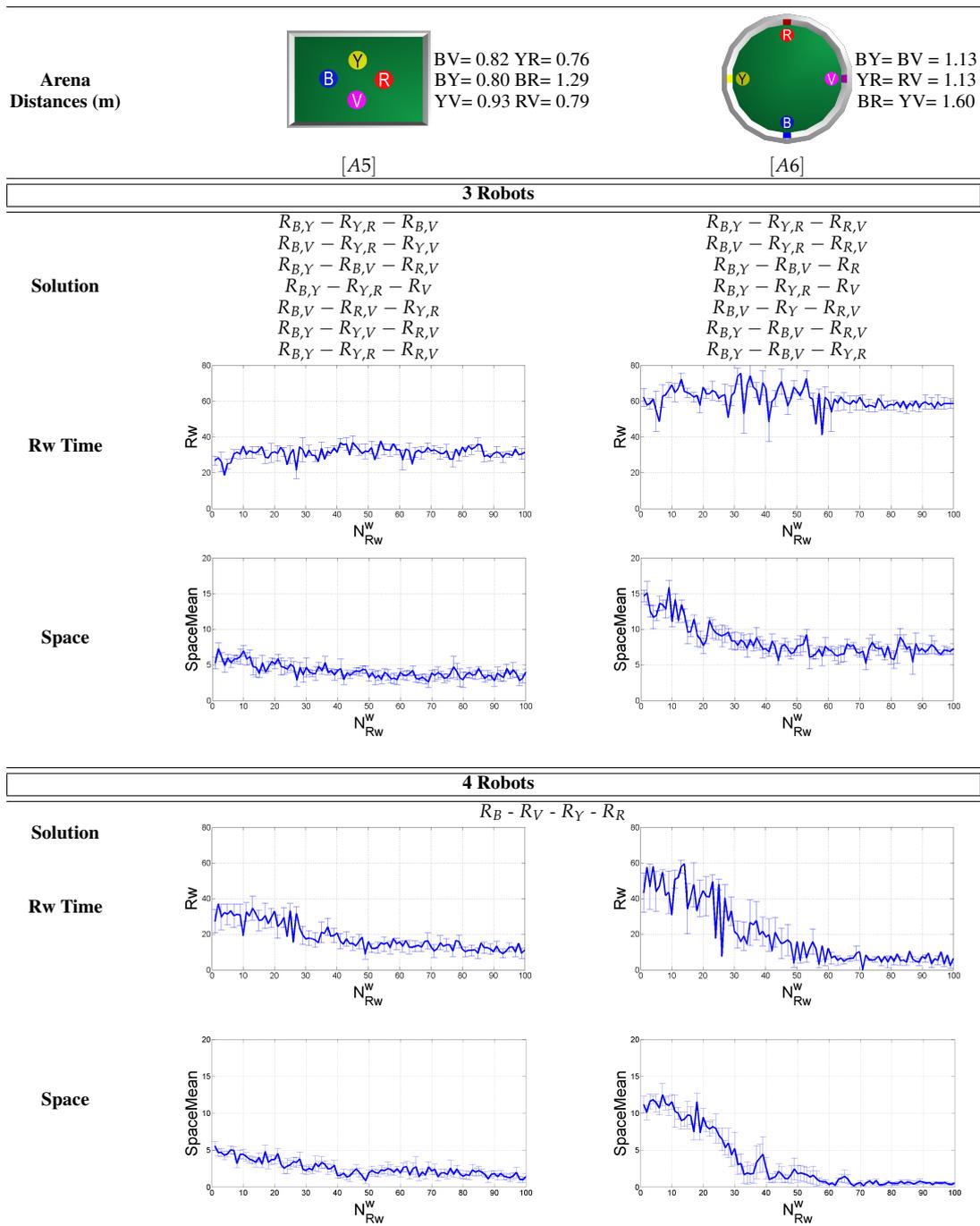


Table 3. Summary of simulation results with three and four robots in *symmetric* arenas. Five different Activation Sequences are used in the campaign of experiments (B-Y-R-V, B-V-Y-R, B-V-R-Y, B-R-V-Y, B-Y-V-R). Legend: **Arena**: the environmental setup, **Solution**: the emerging solution, **Rw Time**: time interval between two consecutive Rw signal activation, **Space**: normalized distance travelled by the group of robots.

Through $S(w)$ we can evaluate the evolution of the effect that learning induces in the number of times the foraging task is properly accomplished in respect of the final obtained solution. In Table 4 the different evolutions of the Specialization index $S(w)$ are shown for all simulations, referring to asymmetric, competitive and symmetric arenas with three and four robots. The index is evaluated using sliding windows of R_w successful events (N_{Rw}^w). Comparing the figures related to experiments with

four robots, it is possible to notice how the arrangement of the targets strongly influences the dynamic of learning. In fact the asymmetric configuration forces a very fast convergence, and polarizes the index to high values (75% of success) since the beginning. In the symmetric situations the convergence trend is slower. The index shows how the initial situation is quite balanced (55% of success): all robots visit all the targets before converging to a specific solution. Moreover from $20N_{Rw}^w$ to $80N_{Rw}^w$

it is possible to see how the symmetric arrangement affects the convergence; in fact, although the whole group behaviour always maximises $S(w)$ in eq.(6), competitive behaviours are here strongly present. Instead, comparing the index for the experiments with three robots the

trend shows a more balanced evolution, when robots continue to visit and share more targets until the convergence. The high-value of deviation suggests that specialization and accordingly the convergence depends on activation sequence.

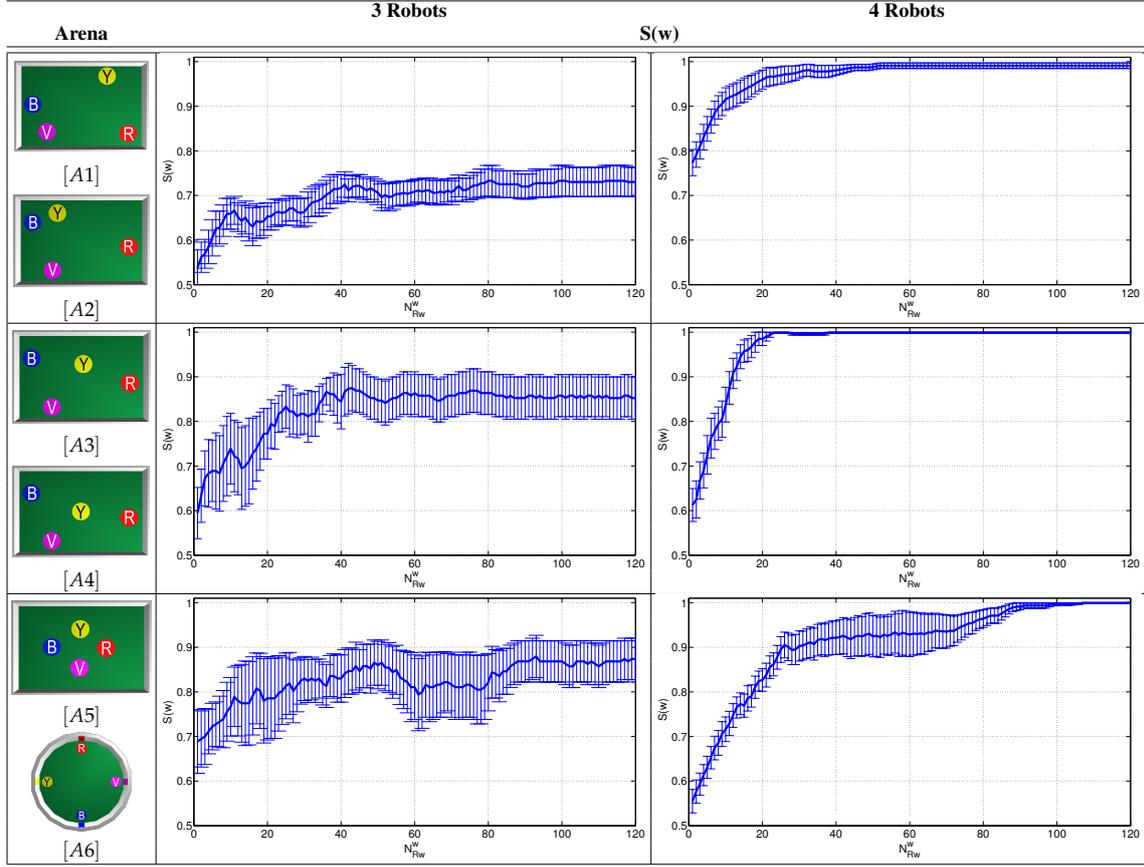


Table 4. Summary of Specialization index ($S(w)$) trends grouped by the three arena classes. The error bar shows the deviation along the mean evolution of the $S(w)$. The mean of Specialization and the standard error of the mean (SEM) are evaluated using a sliding window of w Reward events ($w = 10Rw$) in a set of 20 experiments each.

Starting from the evenness index defined by Simpson in [21] and the social entropy based on Shannon’s information entropy [22], it is possible to analyse the dynamic of specialization related to the diversity evolution among the group, using the Balch’s Behavioural Diversity provided in [23]. In our case the increase of the index is directly correlated with the diversity arising from adaptation in the neural controller of robots, according to the formula:

$$D(k) = 1 - \sum_{i=1}^R \left(\frac{q_i(k)}{R} \right)^2 \quad (7)$$

where $R = \{3, 4\}$ is the size of the group, the number of involved robots and

$$q_i(k) = \frac{N_i}{N_t} \quad (8)$$

where N_i is the total number of vision neurons in robot i whose adaptation currents are above the threshold I_{ima} ; N_t is the total number of vision neurons in robot i that depends on the number of different types of targets in the arena (i.e. different colours).

Accordingly with this definition, comparisons among Diversity indices related to arenas used in the campaign are reported in Table 5. In every condition the system exhibits the same evolution: it starts with a complete evenness and converges to a certain level of diversity; the differences involve the time trends and the upper bound. In particular, in experiments with three robots reported in the first column (upper side), the index does not converge to the maximum value since some targets are shared by robots and so a residual similarity in the group is still present: the use of STM without LTM induces slow development of adaptation and a weak diversity. To improve the performance of the system, a second level of

memory (LTM) can be considered working together with STM, imposing $\alpha=0.5$ in the eq. (3). The whole control system is now able both to speed up the learning process and to maximize the Diversity index reducing the conflicts due to shared targets. Therefore, the addition of LTM term makes the robot who by chance succeeds in visiting the contended target less times in the past to strongly reduce its sensitivity to optimize task partition-

ing. The diversity index evolution, with the addition of the LTM term, shows now similar time-trend for all the scenarios considered, independently either from symmetry in the target distribution, or from the robot number. The comparison between the two control configurations is shown in Fig. 4, where statistics related to the rise time of the Diversity index for both groups of three and four robots are reported.

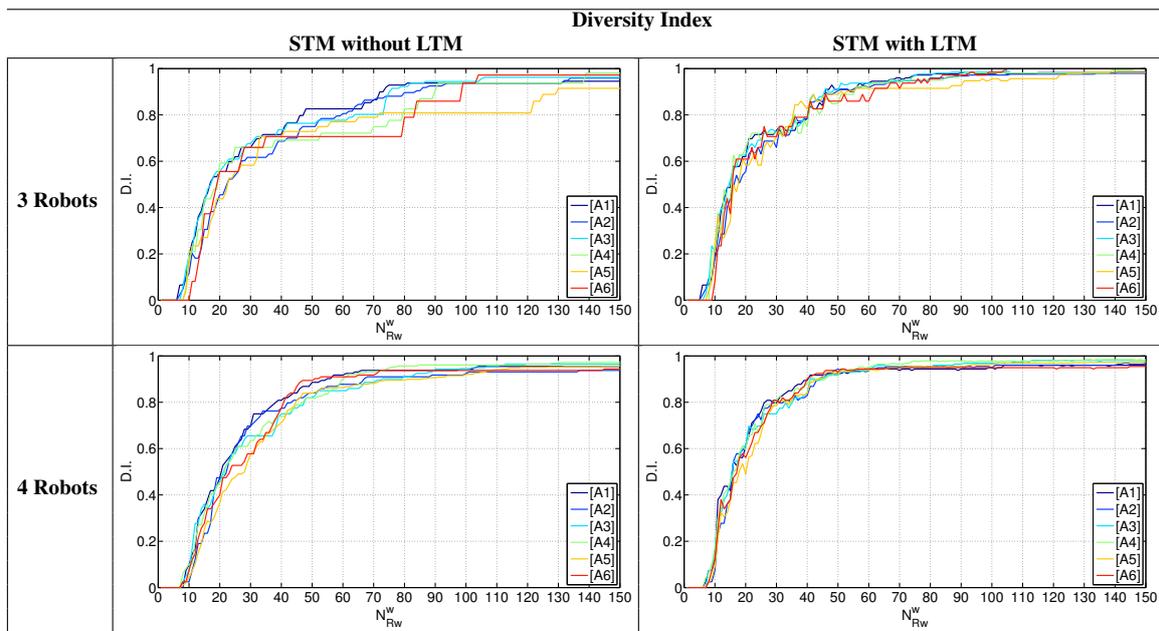


Table 5. Comparison of the evaluated Diversity index for the different arenas (Tables 1, 2, 3) where three and four robots are considered. In the first column the index is evaluated using only the STM, whereas the second column shows the Diversity index evaluation where the STM and LTM are combined.

4. Remarks and Conclusions

The aim of this paper is to experimentally observe the emergence of labour division among robots in completely decentralized situations. In fact each robot has no information about the task to be globally pursued (in this case the sequence of target activation) and, during the learning phase, it becomes more attracted to the frequently retrieved targets, whereas it loses interest in the other ones. This individual behaviour, result of an operant conditioning on the single agent, acts in a distributed manner, since the decisions of the individual bias the operant conditioning results of the others. This behaviour is common in insects: flies, for example, are initially attracted by all the targets in the environment and they learn positive or negative associations as a consequence of rewarding or punishing events. The same basic neural structure has been embedded into every robot, to evaluate how these simple but efficient plastic networks can bias the single individual behaviour and, indirectly, contribute to shape the collective capabilities. The presented results show that the emergence of collective behaviours, at least in these reported cases, can arise from very simple, egocentric and non-communicating single robotic architectures.

The threshold adaptation mechanism was demonstrated to be an interesting approach to allow the emergence of collective behaviours and labour division in presence of a distributed, environmentally mediated, operant conditioning. From the extensive simulation campaign carried out it was possible to outline critical situations where some targets are shared between robots and specialization slows down. The introduction of a two level memory, and in particular of the LTM in the adaptation rule, allows to cope with such situations, obtaining an efficient adaptation which results quite independent from the scenario considered. The key remark to be underlined is that in this approach no particular capabilities are ascribed to each agent within the group. Even in this case, a suitable task division among the agents is obtained, exploiting the mediation of the environment through the action of the reward function.

A spiking neural structure was derived after modelling the learning system in the fly. Indeed, such a structure is only a block of a more complex insect brain computational architecture endowed with other functionalities (e.g. orientation, path integration, decision making) and some others recently proposed which reproduce attention and expectation [13].

Starting from the simulation results here presented, the addition of the remaining blocks of the insect brain architecture within each robot is expected to give rise to the emergence of a surprising number of different collective behaviours, which will depend on the reward policy. In fact it was very recently discovered that fruit flies are indeed able to show also simple forms of imitation [24]. Even if they do not basically show evident social behaviours, nevertheless embedding their basic brain capabilities into a robotic population can lead to derive new strategies for cooperation, basically dependent on the individual capabilities rather than on the presence of a global social brain.

The simulation results demonstrate that the presence of a global reward and a simple memory structure are able to induce diversity in a team of homogeneous robots as also derived in general approaches to swarm intelligence [25, 26] but using different tools and methodologies. Moreover the introduction of a long-term memory block allows to improve the convergence towards a specialized solution.

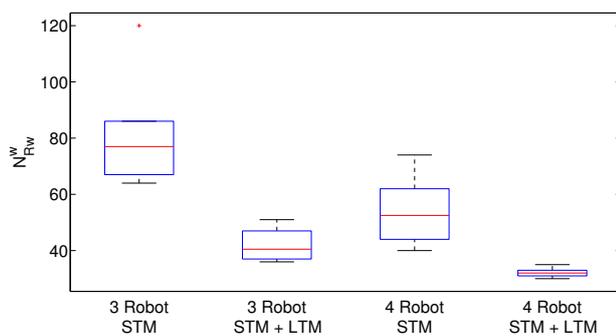


Figure 4. Statistical distribution of the rise time for the Diversity index when the STM is used alone or in combination with the LTM. Both cases of groups of three and four robots are evaluated. In details, on each box the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considering outliers, that are plotted individually (plus symbol).

5. Acknowledgements

The authors acknowledge the support of the European Commission under the project FP7-ICT-2009-6 270182 EMICAB and the MIUR project CLARA (CLoud pLatform and smart underground imaging for natural Risk Assessment). This article is a revised and expanded version of a paper entitled *Environmentally induced task partitioning in competing bio-robots* presented at BioInspired-Robotics Conferences 2014 14-15 May, ENEA-Frascati, Rome.

6. References

- [1] Y. Mohan and S.G. Ponnambalam. Swarm robotics: An extensive research review. *Nature & Biologically Inspired Computing*, pages 140–145, dec 2009.
- [2] E. Bonabeau, M. Dorigo, and G. Theraulaz. *Swarm Intelligence: From Natural to Artificial Systems*. Oxford Univ. Press, New York, 1999.

- [3] S. Camazine, J.-L. Deneubourg, N. R. Franks, J. Sneyd, G. Theraulaz, and E. Bonabeau. *Self-Organization in Biological Systems*. Princeton Univ. Press, 2001.
- [4] L. Chittka and J. Niven. Are Bigger Brains Better? *Current Biology*, 19(21):R995–R1008, November 2009.
- [5] J. G. Burns and A.G. Dyer. Diversity of speed-accuracy strategies benefits social insects. *Current biology*, 18:R953–R954, Oct 2008.
- [6] P. Arena, L. Patané, and A. Vitanza. Environmentally induced task partitioning in competing bio-robots. In *Proceedings of the Conference-Workshop Bio-inspired Robotics*, 14-15 May 2014.
- [7] P. Arena, L. Patané, and A. Vitanza. Spiking neural controllers in multi-agent competitive systems for adaptive targeted motor learning. *submitted to Special Issue on Advances in Nonlinear Dynamics and Control in Journal of the Franklin Institute*, 2014.
- [8] P. Arena, L. Patané, and A. Vitanza. Autonomous learning of collaboration among robots. In *Neural Networks (IJCNN), The 2012 International Joint Conference on*, June 2012.
- [9] T. Balch. Learning roles: Behavioral diversity in robot teams. pages 7–12. AAAI, 1997.
- [10] P. Arena, S. de Fiore, L. Patané, M. Pollino, and C. Ventura. STDP-based behavior learning on the TriBot robot. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 7365 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 736506, may 2009.
- [11] S. Scherer, R. F. Stocker, and B. Gerber. Olfactory learning in individually assayed drosophila larvae. *Learning Memory*, 10(3):217–225, 2003.
- [12] T. Nowotny, M. I. Rabinovich, R. Huerta, and H. D. I. Abarbanel. Decoding temporal information through slow lateral excitation in the olfactory system of insects. *Journal of Computational Neuroscience*, 15(2):271–281, 2003.
- [13] P. Arena and L. Patané. *Spatial Temporal Patterns for Action-Oriented Perception in Roving Robots II: an insect brain computational model*. Springer, Series: Cognitive Systems Monographs, Vol. 21, 2014.
- [14] E.M. Izhikevich. Simple model of spiking neurons. *IEEE Transactions on Neural Networks*, 14(6):1569–1572, Nov 2003.
- [15] J. Benda, L. Maler, and A. Longtin. Linear Versus Nonlinear Signal Transmission in Neuron Models With Adaptation Currents or Dynamic Thresholds. *Journal of Neurophysiology*, 104(5):2806–2820, Nov 2010.
- [16] P. Arena, M. Cosentino, L. Patané, and A. Vitanza. SPARKRS4CS: a software/hardware framework for cognitive architectures (invited paper). In *Proceedings of the SPIE - The International Society for Optical*

- Engineering*, volume 8068, pages 8068A–18, Prague, Czech Republic, 2011.
- [17] P. Arena, L. Patané, P.S. Termini, A. Vitanza, and R. Strauss. Software/hardware issues in modelling insect brain architecture. In *Intelligent Robotics and Applications*, pages 46–55. Springer Berlin Heidelberg, 2011.
- [18] P. Arena, L. Patané, M. Pollino, and C. Ventura. Tribot: a new prototype of bio-inspired hybrid robot. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, St.Louis, Missouri, USA, 2009.
- [19] L. Li, A. Martinoli, and Y. S. Abu-Mostafa. Learning and measuring specialization in collaborative swarm systems. *Adaptive Behavior - Special issue on Mathematics and Algorithms of Social Interactions*, 12(3–4):199–212, Dec 2004.
- [20] T. Balch. Hierarchic social entropy: An information theoretic measure of robot group diversity. *Autonomous Robots*, 8:209–237, 2000.
- [21] E.H. Simpson. Measurement of diversity. *Nature*, 163:688–688, 1949.
- [22] C.E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 1948.
- [23] T. Balch. *Behavioral Diversity in Learning Robot Teams*. PhD thesis, Atlanta, GA, USA, 1998.
- [24] M. B. Sokolowski. Social interactions in “simple” model systems. *Neuron*, 65:780,794, 2010.
- [25] T. Balch. Communication, diversity and learning: Cornerstones of swarm behavior. In Erol Sahin and William Spears, editors, *Swarm Robotics*, volume 3342 of *Lecture Notes in Computer Science*, pages 21–30. Springer Berlin / Heidelberg, 2005.
- [26] L. Bayindir and E. Sahin. A review of studies in swarm robotics. *Turkish Journal of Electrical Engineering*, 15(2), 2007.