

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 and ubiquitous role in living beings. Moreover, several behaviours, such as feeding and courtship, involve environmental exploration and exploitation, including local competition, and lead to a global benefit for the colony. This can be considered as a form of global cooperation, even if the individual agent is not aware of the overall effect. This paper aims to demonstrate that identical biorobots, endowed with simple neural controllers, can evolve diversified behaviours and roles when competing for the same resources in the same arena. These behaviours also produce a benefit in terms of time and energy spent by the whole group. The robots are tasked with a classical foraging task structured through the cyclic activation of resources. The result is that each individual 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, whereby the cumulative energy spent, in terms of the overall travelled distance and the time needed to complete the task, tends to be minimized. A series of simulation experiments is conducted using different numbers of robots and scenarios: the common emergent result obtained is the role-specialization of each robot. The description of the neural controller and the specialization mechanisms are reported in detail and discussed.

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

1. Introduction

Coordination and collaboration among robots are the result of self-organized behaviours: social insects provide brilliant solutions to foraging, migration, mating and other problems [1, 2]. Transferring these characteristics to future biorobotic systems will ensure both flexibility in space- and time-varying environments and high robustness to faults in individual agents [2, 3]. Nonetheless, 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 exhibit apparent cooperative capabilities but compete for food and mates [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 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 individual agent's behaviour and the environment co-evolve - during the agent life-cycle - to achieve a global equilibrium for the colony. The environment acts so as to shape the local

competitive behaviour of the individual agents to give rise to a global cooperative strategy, leading to an equilibrium state.

An open question in animal social behaviour relates 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 agents which simply compete for survival? We have tried to answer this question through a series of simulations considering 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 environmentally 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 the absence of 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 individuals is required 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 a specialization strategy was introduced and an initial collaborative algorithm was formalized to fulfil an overall mission. Many research activities tackle specialization and the relationships between a reward mechanism, behavioural diversity and the optimization of performance [9]. The aim of our work is to investigate and quantify the influence of environmental mediation on operant conditioning at the level of the individual 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 positions 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 in indirectly influencing individual agent behaviours depending on the other elements of the group. In the proposed work, task partitioning is addressed, allowing a decrease in 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 in each robot are presented here. Further details on specialization learning and the 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 us to obtain an environmentally dependent rule that guides the searching for the targets. At the end of the sequence, a reward signal (Rw) is activated and then the cycle begins again. The individual 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 towards it at 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 as opposed to another which is nearer but badly oriented. No direct communication is introduced among the agents: they compete to reach the same target once this is activated. When the cycle is concluded and the Rw is activated for all the robots, a learning phase is then performed. During the training phase, each robot increases its interest on the reached targets reducing - at the same time - the attractiveness of the others. 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 a form of 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 Rw contributes to generate diversity and specialization. Here, the targets are 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 individual 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 Rw, generated by the environment, triggers the learning process. This consists of - in our case - 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 *STM* permits the memorizing of the targets retrieved during the activation sequence and acts for every *Rw* activation; alternatively, the *LTM* retains the visited targets for a 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 by 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 proved useful in 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 entire 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 also receives information from the mechanoreceptors distributed along the body. Moreover, the 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) had already been applied, as shown in [10], to let all the robots exhibit tactic behaviour for all the targets present in the environment. In particular, the module devoted to visual and target recognition was split into as many S&SMs as the types of targets present in the environment (see Fig. 1).

The learning process obtained through the threshold adaptation neuron (T_aN) was applied to induce hyperpolarization or depolarization in 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 the neuron model used for a generic T_aN is reported here:

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

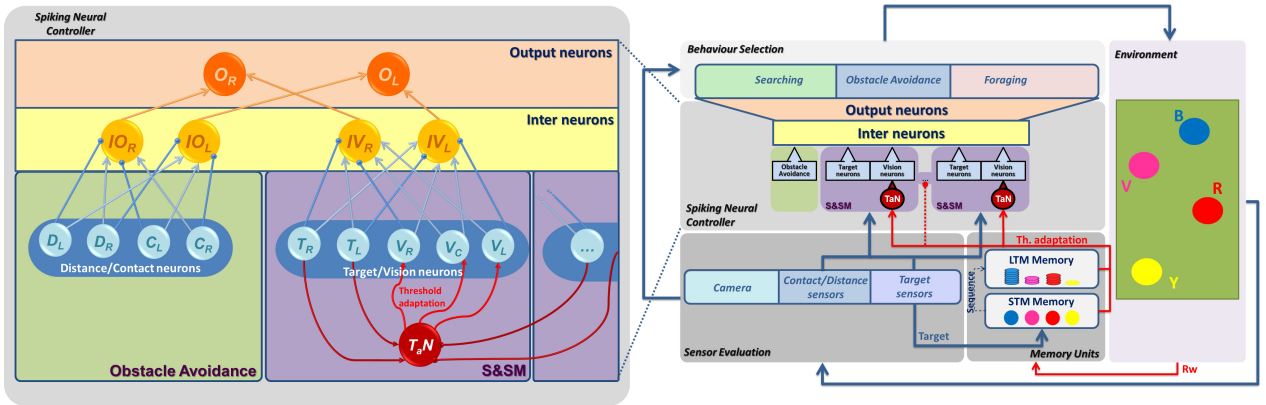


Figure 1. Block diagram of the control architecture. In the *Sensor Evaluation* block, data from the sensors are collected. The *Spiking Neural Controller* block describes the neural network model used for the foraging. The number of *Sensing and Specialization Modules* (S&SM) is related to the different targets to be recognized. T_aN is the threshold adaptation neuron, devoted to inducing specialization learning. The threshold adaptation, applied to the neural structure, is guided here 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 on the left-hand side: the three-layer structure permits the traducing of inputs arriving from the target sensors and the 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) and Distance (D_x), related to the *Obstacle Avoidance* block, and Target (T_x) and Vision (V_x), related to S&SM, where $x = L, R, C$ indicates the left-hand and right-hand sides and the central region, respectively. In the inter-layer, the sensory information is 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-hand and right-hand motors of the system (O_x).

with the spike-resetting

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

where $a=0.02, b=-0.1, c=-55, d=6, v$ and 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 the events occurring before the reward-signal activation (Rw) for the STM factor, as well as on the 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 on 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 Rw 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 Rw 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 the 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 ceases to be attractive when the bias goes below a lower bound which corresponds here to $I_{ina}=g_a V_{th}=14$. Below this value, the vision neuron no longer emits 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 visitations. This is especially useful in competitive scenarios where the broad 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 the dimensions $0.3m \times 0.2m$ 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 in studying individual behaviour modulations implied considering small groups in order to analyse specific emergent behaviours in the individual robots. For this reason, we limited the groups to three or four robots. Moreover, the number of targets was fixed as four so as 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 exceeded the number of robots were then analysed. Different target arrangements and various activation sequences are used for the simulation tests. The Rw is activated when the final target is reached by a robot.

A number of different simulations were performed with three and four robots, for a total of more than 100 simulation tests, and the results are summarized in Table 1, Table 2 and Table 3. The campaign was conducted to investigate whether the environment, merely imposing a fixed sequence, indirectly induces the robots to select a subset of targets to minimize the travelled space to complete the sequence. This minimal energy solution to the task was investigated using only the $STM(\alpha=1)$. In particular, for each of the arenas reported in the following tables, five different activation sequences are used (see the captions for details) and the corresponding solutions are considered in order 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 the targets was absolutely random but the obtained results can be collected into three classes, depending on the target's spatial configuration: namely *asymmetric*, *competitive* and *symmetric* arenas (see Table 1, 2, 3 respectively). The first row of each table

contains the arena description, and for each configuration at least 10 simulations were performed starting from different initial robot positions. The results are split into sub-tables where the robot-target solutions, the time spent to obtain a reward and the covered space (defined as the normalized distance travelled by the group of robots) are reported for three and four robots. From the analysis of the simulations, it might 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, namely the time spent to obtain a reward activation and the global space covered by the robots, are minimized.

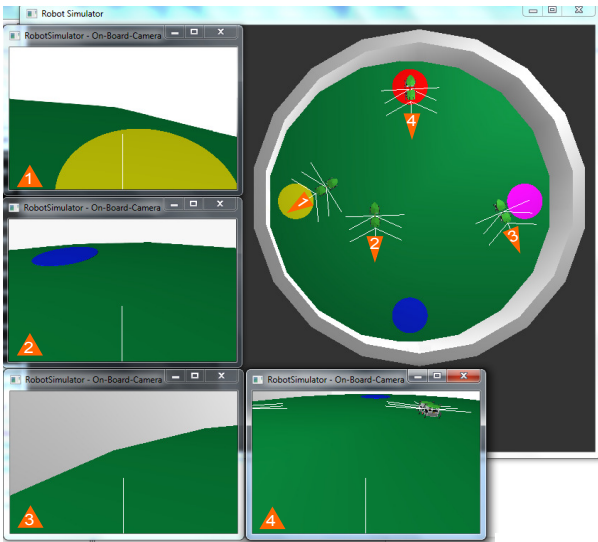


Figure 2. The dynamic simulator overview. In this scenario a group of four robots are placed in a circular arena where four targets are symmetrically distributed on the ground. The camera view of each robot is also reported. The white lines around the robots represent the distance sensors: the lowest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

A different situation arises in the scenario reported in Table 2 ([A3] and [A4]), where the yellow target position is shifted from the blue one towards the red one. In such cases, the solution is dependent - besides on the target position - on the activation sequence. Under these conditions, namely for situations where the number of targets exceeds the number of robots, competitive behaviours are massively promoted and it is clearly shown by the 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 tolerance issues; in fact, by sharing targets, the robots could follow the activation sequences after a fault in some robot; in contrast, the group does not achieve complete specialization and often competitive behaviours compromise the minimization of the travelled space and the time needed to obtain a R_w . Similar considerations can apply to the last two cases in Table 3,

which report two symmetric target configurations. In both cases, experiments with three robots show that the travelled space decreases, whereas the time interval between two R_w activations shows a disorderly trend until it converges to a more regular shape. In particular, the symmetric configuration and position of the targets in the arenas ([A5] and [A6]) induces obstacle avoidance reactions more often, due to competitive actions against other robots or else due to the 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 maximal; as such, the time passes 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 R_w 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 also be noted 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 globally covered path, whereas in the experiments with three robots a less clear trend can be noted. Furthermore, in some cases (i.e., symmetric arena [A5]), a larger number of R_w events is needed to achieve convergence in terms of the time spent to complete the task (about $300N_{Rw}^w$). The inherent asymmetry in the system generates solutions whereby some targets are shared by the robots, inducing a kind of disorder. Similarly, symmetric arenas largely emphasize 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. As such, these configurations facilitate the learning process avoiding becoming trapped in local minima. This happens - in contrast - in the symmetric situation, producing a minor performance enhancement. The competitive case can be considered in between these, since the presence of targets shared by the robots facilitates specialization with respect to the symmetric case but induces more competitive reactions in its asymmetric counterpart.

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

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

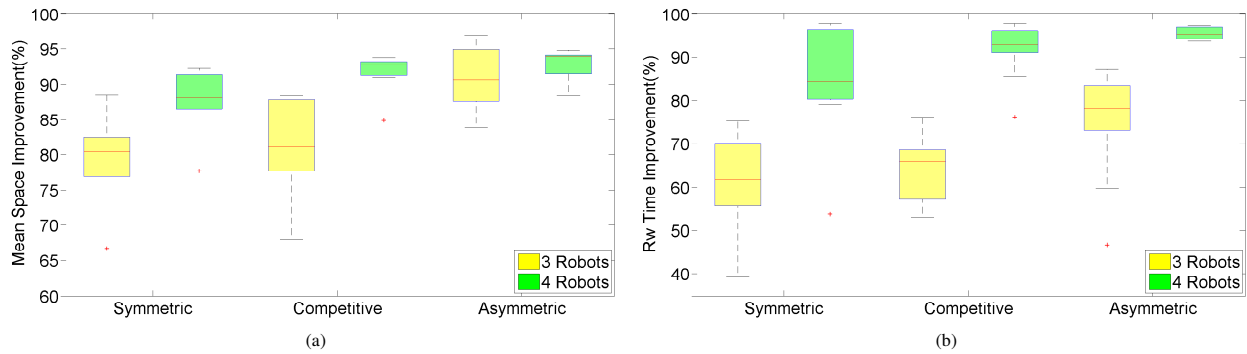


Figure 3. Evaluation of the improvement obtained through the learning process in terms of: (a) the mean space covered by each robot to obtain a R_w (i.e., when all the targets are visited following the specific activation sequence); (b) the time interval between two consecutive R_w signal activations.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

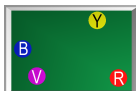
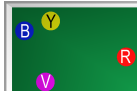
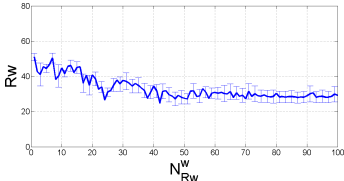
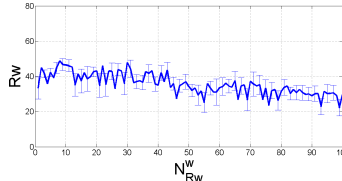
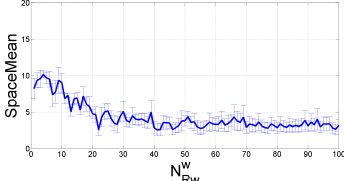
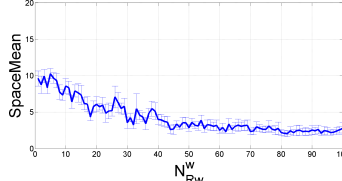
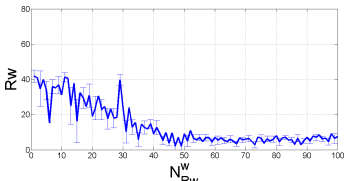
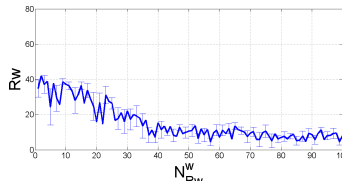
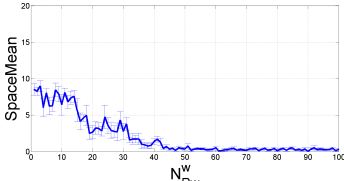
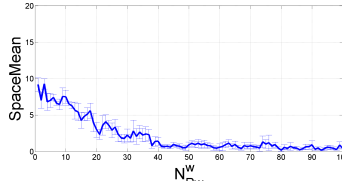
Arena Distances (m)	[A1]	[A2]
	 $YV = 2.05$ $RV = 2.03$ $BY = 1.98$ $BR = 2.49$ $BV = 0.76$ $YR = 1.53$	 $BY = 0.65$ $BR = 2.46$ $BV = 1.28$ $YR = 1.95$ $YV = 1.41$ $RV = 1.98$
3 Robots		
Solution	$R_{B,V} - R_Y - R_R$	$R_{B,Y} - R_R - R_V$
Rw Time		
Space		
4 Robots		
Solution	$R_B - R_V - R_Y - R_R$	
Rw Time		
Space		

Table 1. Summary of the simulation results with three and four robots in *asymmetric* arenas. Five different activation sequences are used in the series 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 R_w signal activations, **Space**: normalized distance travelled by the group of robots.



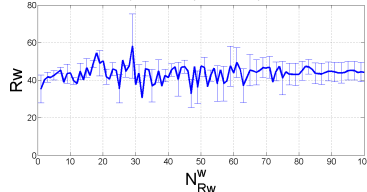
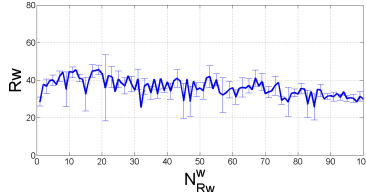
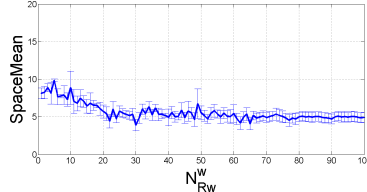
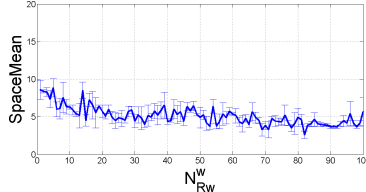
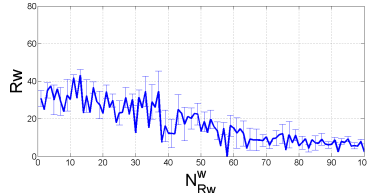
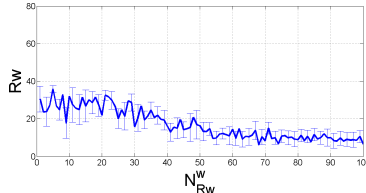
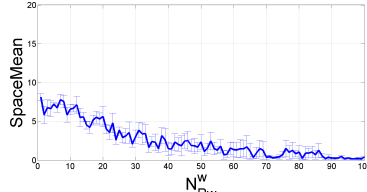
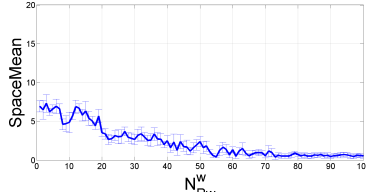
Arena Distances (m)			
3 Robots			
Solution	$R_{B,Y} - R_{Y,R} - R_V$ $R_{B,V} - R_{Y,R} - R_{Y,V}$ $R_{B,V} - R_Y - R_R$ $R_B - R_{Y,R} - R_V$ $R_{B,Y} - R_{Y,R} - R_{B,V}$	$R_{B,V} - R_Y - R_R$ $R_{B,V} - R_{Y,R} - R_{Y,V}$ $R_{B,Y} - R_R - R_V$ $R_{B,Y} - R_{B,V} - R_{Y,R}$	
Rw Time			
Space			
4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

Table 2. Summary of the simulation results with three and four robots in *competitive* arenas. Five different activation sequences are used in the series 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 activations, **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 interactions.

Even if the experiments involving four robots and four targets reach the most obvious solution (i.e., each robot specializes in relation to one target), it of interest 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 entire system, taking as a reference the works [19, 20]. In particular, it follows 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 from 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 .

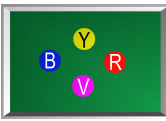
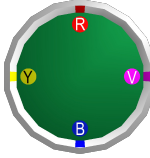
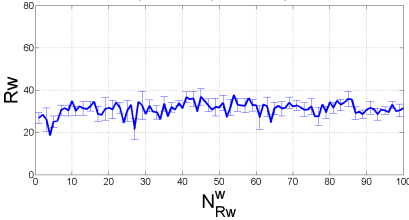
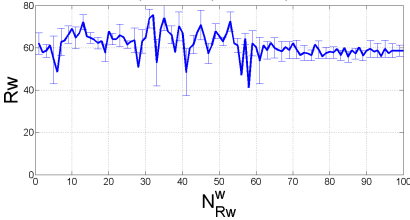
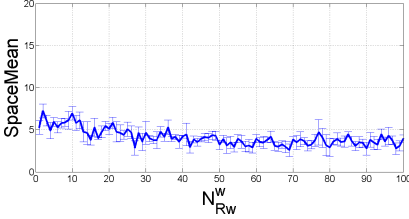
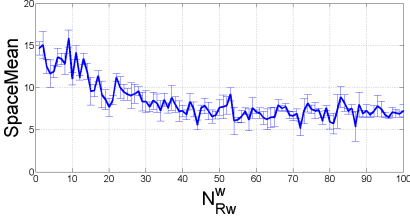
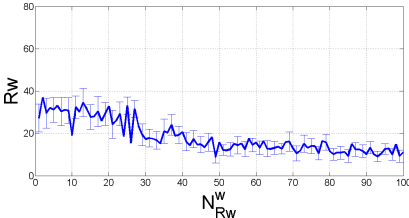
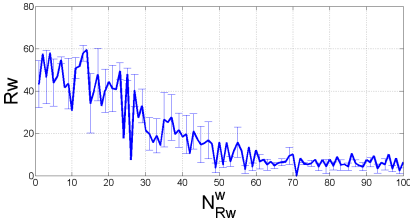
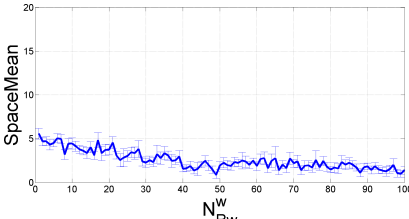
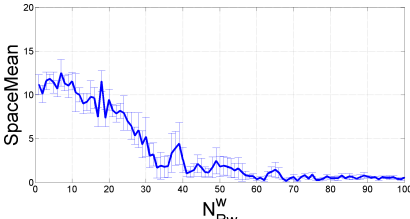
Arena Distances (m)		[A5]	[A6]
			
		BV=0.82 YR=0.76 BY=0.80 BR=1.29 YV=0.93 RV=0.79	BY= BV = 1.13 YR= RV = 1.13 BR= YV= 1.60
3 Robots			
Solution		$R_{B,Y} - R_{Y,R} - R_{B,V}$ $R_{B,V} - R_{Y,R} - R_{Y,V}$ $R_{B,Y} - R_{B,V} - R_{R,R}$ $R_{B,Y} - R_{Y,R} - R_V$ $R_{B,V} - R_{R,V} - R_{Y,R}$ $R_{B,Y} - R_{Y,V} - R_{R,V}$ $R_{B,Y} - R_{Y,R} - R_{R,V}$	$R_{B,Y} - R_{Y,R} - R_{R,V}$ $R_{B,V} - R_{Y,R} - R_{R,V}$ $R_{B,Y} - R_{B,V} - R_R$ $R_{B,Y} - R_{Y,R} - R_V$ $R_{B,V} - R_Y - R_{R,V}$ $R_{B,Y} - R_{B,V} - R_{R,V}$ $R_{B,Y} - R_{B,V} - R_{Y,R}$
Rw Time			
Space			
4 Robots			
Solution		$R_B - R_V - R_Y - R_R$	
Rw Time			
Space			

Table 3. Summary of the simulation results with three and four robots in *symmetric* arenas. Five different activation sequences are used in the series 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 activations, **Space**: normalized distance travelled by the group of robots.

Through $S(w)$, we can evaluate the evolution of the effect that learning induces in terms of the number of times the foraging task is properly accomplished with respect to 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 Rw-successful events (N_{Rw}^w). Comparing the figures related to the experiments with four robots, it is possible to notice how the arrangement of the targets strongly influences the dynamics of learning. In fact, the asymmetric configuration forces a very fast conver-

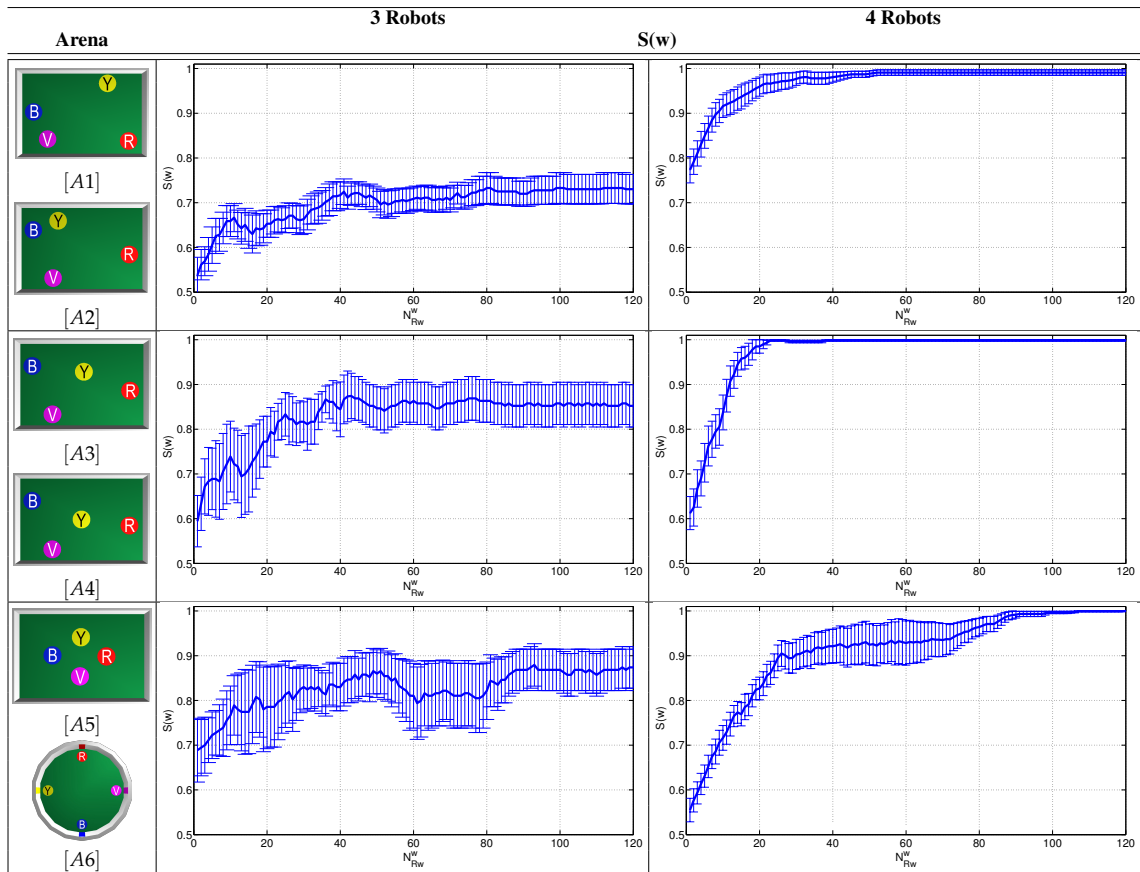


Table 4. Summary of specialization index ($S(w)$) trends grouped by the three arena classes. The error bar shows the deviation along the mean evolution of the $S(w)$. The mean specialization and the standard error of the mean (SEM) are evaluated using a sliding window of w reward events ($w=10Rw$) in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

gence and polarizes the index to high values (75% success) from the beginning. In the symmetric situations, the convergence trend is slower. The index shows how the initial situation is quite balanced (55% 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 maximizes $S(w)$ in eq.(6), competitive behaviours are strongly present here. 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 the deviation suggests that specialization and - accordingly - convergence depend on the activation sequence.

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 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 the 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 a robot i whose adaptation currents are above the threshold I_{ina} , and N_t denotes the total number of vision neurons in robot i that depend on the number of different types of targets in the arena (i.e., different colours).

According to this definition, comparisons among the diversity indices related to the arenas used in the campaign are reported in Table 5. In every condition, the system exhibits the same evolution: it starts with complete evenness and converges to a certain level of diversity; the differences involve the time trends and the

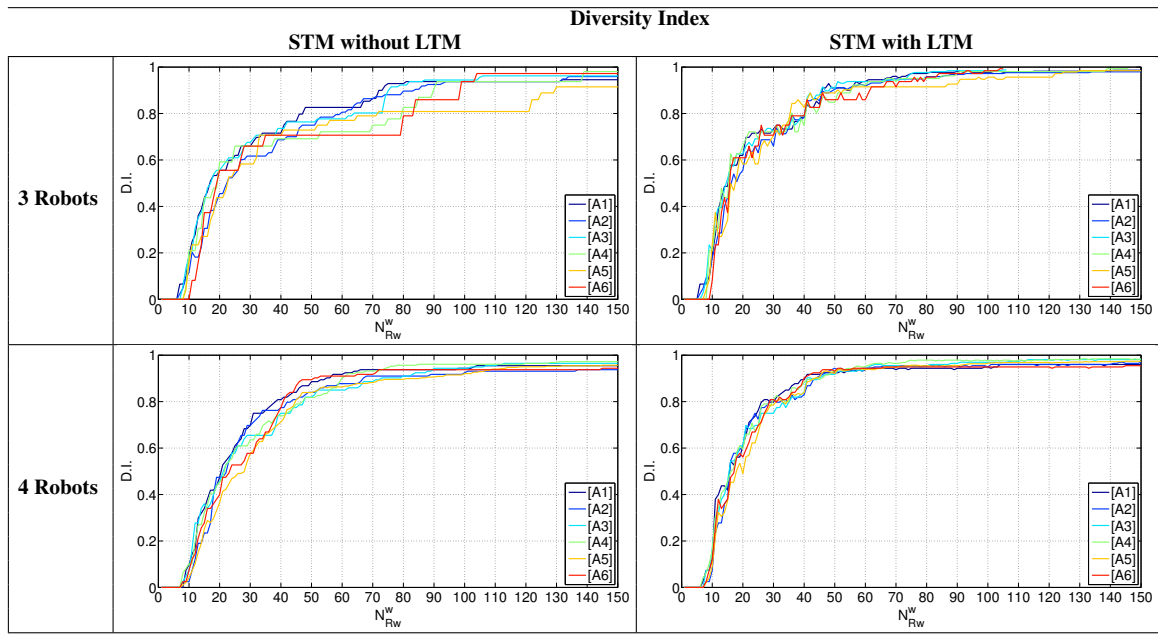


Table 5. Comparison of the evaluated diversity index for the different arenas (Tables 1, 2 and 3) where three and four robots are considered. In the first column, the index is evaluated using only the STM, whereas the second column shows the diversity index evaluation where the STM and the LTM are combined.

upper bound. In particular, in the experiments with three robots reported in the first column (the 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 the slow development of adaptation and weak diversity. To improve the performance of the system, a second-level of memory (LTM) can be considered working together with the STM, imposing $\alpha=0.5$ in eq. (3). The entire control system is now able to both speed up the learning process and maximize the diversity index, reducing the conflicts due to shared targets. Therefore, the addition of the LTM cause the robot - who by chance succeeds in visiting the contended target fewer times in the past - to strongly reduce its sensitivity to optimizing task partitioning. The diversity index evolution, with the addition of the LTM term, now shows a similar time-trend for all the scenarios considered, either independently 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 increased time in the diversity index for both groups of three and four robots are reported.

4. Remarks and Conclusions

The aim of this paper is to experimentally observe the emergence of labour divisions among robots in completely decentralized situations. In fact, no robot has information about the task to be globally pursued (in this case, the sequence of target activation) and, during the learning phase, they become more attracted to the frequently

retrieved targets whereas they lose interest in the other ones. This individual behaviour - the result of operant conditioning on the individual 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 individual individual behaviour and, indirectly, contribute to shape 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 individual robotic architectures.

The threshold adaptation mechanism was demonstrated to be an interesting approach to allow for the emergence of collective behaviours and labour division in the presence of distributed, environmentally mediated operant conditioning. From the extensive simulation campaign carried out, it was possible to outline the critical situations whereby 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 us to cope with such situations, obtaining an efficient adaptation which results quite independently 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 present 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 and decision-making) and certain others which have been recently proposed and which reproduce attention and expectation [13].

Starting from the simulation results presented here, 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 exhibit 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 the deriving of new strategies for cooperation that are basically dependent on 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 LTM block allows us to improve the convergence towards a specialized solution.

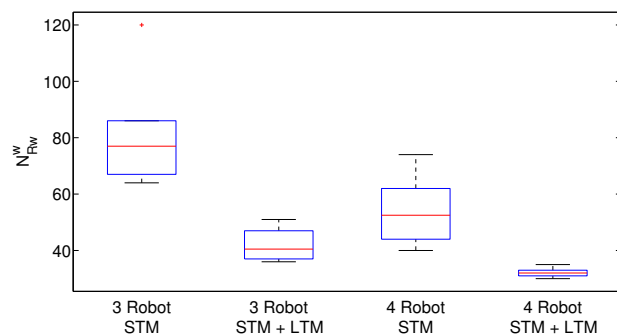


Figure 4. Statistical distribution of the increased time for the diversity index when the STM is used alone or in combination with the LTM. Both cases of groups of three and four robots are evaluated. In particular, on each box the central mark is the median, the edges of the box are the 25th and 75th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

5. Acknowledgements

The authors would like to 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.