Université Libre de Bruxelles

Institut de Recherches Interdisciplinaires
et de Développements en Intelligence Artificielle

On the Design of Neuro-controllers for Individual and Social Learning Behaviour in Autonomous Robots: an Evolutionary Approach

Giovanni Pini, Elio Tuci

IRIDIA – Technical Report Series

Technical Report No.
TR/IRIDIA/2007-020

November 2007

Submitted to the Special issue on Social Learning in Embodied Agents on Connection Science

IRIDIA – Technical Report Series
ISSN 1781-3794

Published by:
IRIDIA, *Institut de Recherches Interdisciplinaires*
et de Développements en Intelligence Artificielle
Université Libre de Bruxelles
Av F. D. Roosevelt 50, CP 194/6
1050 Bruxelles, Belgium

Technical report number TR/IRIDIA/2007-020

Revision history:
TR/IRIDIA/2007-020.001 November 2007

On the Design of Neuro-controllers for Individual and Social Learning Behaviour in Autonomous Robots: An Evolutionary Approach

Giovanni PINI[‡] gpini@iridia.ulb.ac.be
Elio TUCI[‡] etuci@ulb.ac.be
‡ IRIDIA, CoDE, Université Libre de Bruxelles, Bruxelles, Belgium

November 2007

Abstract

In Biology/Psychology the capability of natural organisms to learn from the observation/interaction with conspecifics is referred to as social learning. Roboticists have recently developed an interest on social learning, since it might represent an effective strategy to enhance the adaptivity of a team of autonomous robots. In this study, we show that a methodological approach based on artifcial neural networks shaped by evolutionary computation techniques can be successfully employed to synthesise individual and social learning mechanisms for robots required to learn a desired action (i.e., phototaxis or antiphototaxis).

1 Introduction

Social learning refers to a multitude of phenomena in which the "interactions" between individuals are directly responsible for the development of the behavioural repertoire of some of the actors of the interaction. Evidence of social learning has been found in several species in the domain of food choice, tool use, patterns of movement, predator avoidance, mate choice, and courtship. The ubiquitous nature of social learning in the life of living organisms might reside in the fact that, in a world in which making a mistake may cost the life, learning from the observation/interaction with conspecifics is probably safer and quicker than learning by trial-and-error (see Galef and Laland, 2005).

The interest on the concept of social learning has recently gone beyond the domain of life sciences by invading certain areas of engineering sciences and in particular those dedicated to the development of autonomous robots. The concept of autonomy in robotics is linked to the capacity of the agents to move in an unstructured and changing environment without continuous human guidance. One way to improve the autonomy of an agent is by providing it with a control system in which learning mechanisms allow the agent to adjust its behaviour to the characteristics of the environment. Social learning interest roboticists because it might represent a valid alternative to individual (trial-and-error) learning. For example, in a group of robots, if an agent can learn by observing other robots, the performance of the group might increase rapidly since it is not necessary that each individual learns by trial-and-error strategies.

Evolutionary robotics (ER) might represent a valid design method to synthesise social learning mechanisms in autonomous robots. ER is a methodological tool to automate the design of robots' controllers (see Nolfi and Floreano, 2000, Harvey et al., 2005). ER is based on the use of artificial evolution to find sets of parameters for artificial neural networks that guide the robots to the accomplishment of their objectives. As far as it concerns the subject of this study, the ER methods allow us to develop individual and social learning mechanisms that are grounded on the perceptual and motor experiences of an autonomous agent and fully integrated with all the other underlying structures that underpin its behavioural repertoire. This is because, with respect to other methods, ER does not require the designer to make strong assumptions concerning which behavioural mechanisms are needed by the robots.

The objective of this study is to explore the possibility to integrate in a single controller individual and social learning mechanisms by using ER methods. In particular, we investigate a scenario in which a robot should be capable of (i) switching from phototaxis to antiphototaxis after the perception of a sound (this is the individual learning task, see also Di Paolo, 2003, for a similar scenario); (ii) learning the desired action to take with respect to a light source by imitating a demonstrator previously educated by individual learning. As far as we know, there is no work in literature concerning the design of a single integrated controller capable of guiding autonomous agents required to perform both individual and social learning task using artificial neural networks shaped by evolutionary computation techniques. Distinctive features of our model are: i) the demonstrator is an agent that individually and autonomously learns what action to take with respect to the light; ii) learner and demonstrator share the same genetic material. The difference between the two resides in the states of their controllers at the beginning of the interactions. In particular, the demonstrator differs from the learner in having the state of its controller set as defined by a preliminary training-phase in which it individually and autonomously learn what action to take with respect to a light source (i.e., phototaxis or antiphototaxis). The reader should also bear in mind that in our model, learning refers to changes in time of the behaviour of the agents. We talk about individual learning in those cases in which the behavioural changes are induced by the perception of an environmental stimulus (i.e., a tone). We talk about social learning in those cases in which the behavioural changes are induced in the learner by the interactions with a previously educated agent (i.e., the demonstrator).

The results of this study are a “proof-of-concept”: they show that dynamic artificial neural networks can be successfully synthesised by artificial evolution to design the mechanisms required to underpin individual and social learning in autonomous robots.

1.1 Structure of the paper

In what follows, we first present a review of previous research works on social learning in autonomous robots (see section 2). In section 3, we describe the simulation scenario investigated in this research work. In sections 4, 5, 6, and 7 we describe methodological issues of our study. In section 8, we illustrate the results and post-evaluation analysis of our simulations. Discussion and conclusions are presented in section 9.

2 State of the art

Social learning is the subject of an enormous corpus of research work both in biology and psychology. For many living organisms, observing and imitating conspecifics rather than pursuing a trial-end-error strategy is a way to reduce the risk of making a wrong choice. This is because the decision of those that “walk” on territories unexplored yet by others, might have already been shaped by the consequences of the taken action, and therefore being safe. From an evolutionary perspective, biologists are interested in understanding what are the selective pressures which favour the evolution of social learning with respect to other learning processes that do not rely on the observation of the behaviour of conspecifics (see Galef and Zentall, 1988, Boyd and Richerson, 1989, 1996). Ethologists are interested in observing and possibly quantifying the effects of social learning on the dynamics of certain animal societies, such as those of monkeys and apes (see Fragaszy and Visalberghi, 1999, Tomasello and Call, 1997, Tomasello, 1999). Psychologists focus more on the mechanisms which lead an individual to imitate another. In particular, they try to understand whether social learning in animals requires cognitive mechanisms (e.g., perspective taking) that empirical evidence seems to limit to humans (see Thorpe, 1963, Heyes and Galef, 1996, Flinn, 1997, Zentall, 2006)).

More recently, social learning has become the subject of debate in disciplines which do not strictly belong to the domain of life sciences. In particular, robotics has been showing to be not just “permeable” to controversies on social learning but also capable of providing constructive elements that have structured our understanding of these biological phenomena (see Dautenhahn and Nehaniv, 2002, Nehaniv and Dautenhahn, 2007). Among the multiple reasons which might have contributed to the interests of roboticists on issues concerning social learning, we mention just two: i) the development of humanoid robots; and ii) the development of multi-robot systems.

Sophisticated humanoid platforms make possible for humans to interact with them in alternative ways based on structural and functional similarities between the human body and the morphological structure of the machine (see Breazeal and Scassellati, 2002, Dautenhahn, 2007).

These alternative ways of interaction have been considered the basis for the development of learning processes by which the machine acquires new skills or adapts to the specific needs of the user by imitating actions shown by a human being. In other words, the human and the machine are considered entities of a social context in which the latter benefits from the observation of the behaviour of the former by employing mechanisms based on the principles that regulate social learning in living organisms. From an engineering point of view, showing a task to a robot capable of social learning in order to teach the robot how to perform the desired action, might be easier than defining the low level actions the robot must take to accomplish the same task. Examples of this approach can be found in Billard and Matarić (2001), Pollard and Hodgins (2002). In the former work a simulated humanoid, controlled by a hierarchy of neural networks, must learn arm movements by "observing" recorded arm movements of a person. In the work of Pollard and Hodgins a humanoid robot, controlled by PD controllers, learns how to manipulate objects by observing a human teacher.

Recently, there has been a growing interest in multi-robot systems since, with respect to a single robot system, they provide increased robustness by taking advantage of inherent parallelism and redundancy. Moreover, the versatility of a multi-robot system can provide the heterogeneity of structures and functions required to undertake different missions in unknown environmental conditions. The growing interest that roboticists have shown in multi-robot systems, has stimulated research works on issues concerning the design of control mechanisms to allow autonomous robots to learn from conspecifics by means of social interactions (Matarić, 2002). This is because social learning provides a way to improve the performance of a group of artificial agents without necessarily that every agent of the group learns by trial-and-error strategies. Examples in which the robots learn autonomously by social interaction are the work of Demiris and Hayes (1994), and Billard (2002). In the work of Demiris and Hayes (1994), a learner robot must follow an expert in a maze. The learner robot is controlled through a set of modules, which makes it follow the teacher in the maze and learn some rules, which can then be used to navigate other mazes. In the work of Billard (2002), a recurrent neural network makes a mobile robot follow a teacher—that can be a human or another robot—and learn a vocabulary, to identify objects. Learning is obtained through network's synaptic weight changes, based on hebbian rules.

The research work described in this paper aims to contribute to the development of social learning mechanisms to allow autonomous robots to interact among each other on the basis of their own experience of the world they inhabit. One of the main differences between the research work quoted in this section and the one illustrated in the paper, concerns the status of the demonstrator. In the above mentioned studies, the demonstrator is an agent programmed to show a specific action, while in our model is an agent that autonomously and individually learn what action to show to the learner. To the best of our knowledge, this is the first research work in which artificial dynamic neural networks shaped by artificial evolution are used as building blocks for the design of social learning mechanisms. Recent research work proved that these methods can be successfully employed to design control mechanisms to allow single robots to solve individual learning tasks (see Tuci et al., 2003). We believe that these methodological tools represent a powerful means to obtain prejudice-free controllers capable of underpinning social learning behavioural strategies grounded on the sensory-motor experience that the agents have of their world.

3 Description of the task

We consider a task in which populations of autonomous robots are required to show individual and social learning capabilities. Learning refers to the behaviour that the robots should perform with respect to a light source (i.e., phototaxis or antiphototaxis). In this section, we describe the individual and social learning task.

Individual learning refers to the capability of an agent to switch from phototaxis to antiphototaxis after the perception of a tone. During individual learning, at the beginning of each trial (i.e., t=0), a robot is positioned in a boundless arena at a randomly chosen distance from the light $d_t \in [55cm, 65cm]$. In those trials which precede the emission of a tone, the robot is evaluated for its ability to get as close as possible to the light. These trials are considered to be successful when the robot's distance from the light source is less than 5cm. The robot is required to perform phototaxis until a tone is emitted in the environment. The tone can potentially be emitted in any trial including the first one, and it is emitted only once in the robot's life-time, which corresponds to a sequence of 10 trials. The emission starts when the robot reaches half of the initial distance

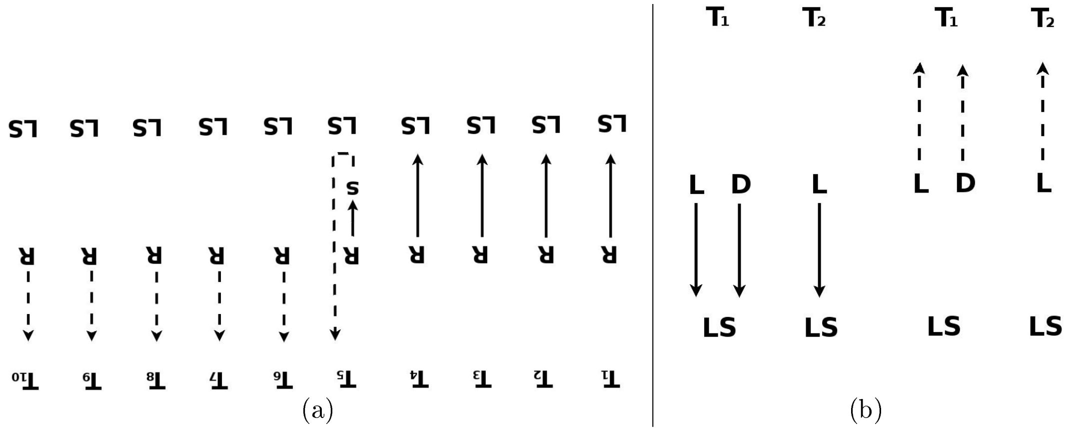


Figure 1: Depiction of the (a) individual and (b) social learning task. In figure (a), R indicates the robot's starting position at the beginning of each trial T_i. LS indicates the position of the light source. S indicates the emission of a tone. Before the emission of the tone, a successful robot should perform phototaxis (see continuous arrows). After the emission of the tone, the robot should perform antiphototaxis (see dashed arrows). In figure (b), L and D refers to the learner and demonstrator starting position respectively. During the first trial T_1, learner and demonstrator are placed close to each other. In the second trial T_2, the demonstrator is removed and a successful learner should imitate the behaviour shown by the demonstrator in the previous trial by performing either phototaxis or antiphototaxis.

from the light (i.e., $0.5d_0$) and it lasts between 3 and 4 seconds. After the emission of the tone, the robot is evaluated for its ability to move as far as possible from the light. The "after-tone" trials are successfully terminated whenever the robot's distance from the light source is $1.5d_0$ (see Figure 1a).

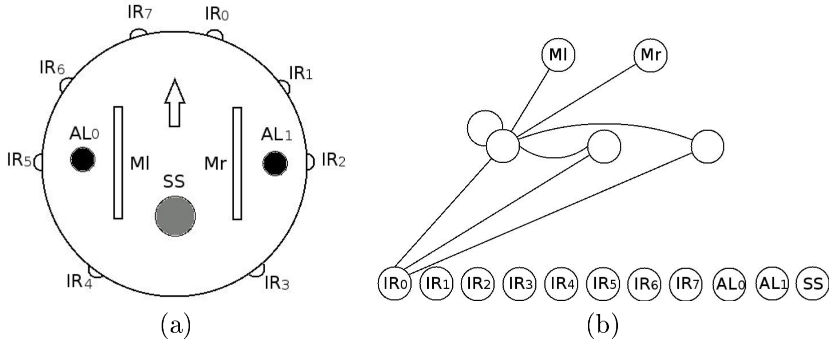
Social learning refers to the capability of a learner to imitate the behaviour of a demonstrator by directly interaction with it. The demonstrator is a robot that has already (individually) learnt the appropriated response with respect to the light (i.e., phototaxis or antiphototaxis). A robot (socially) learns if it proves to be capable of imitating in subsequent trials the correct response to the light performed by a demonstrator. Two trials are required to establish whether or not the learner is capable of imitating the actions of the demonstrator. At the beginning of the first trial (i.e., t=0), the learner is placed at a randomly chosen distance from the light $dll_0 \in [55cm, 65cm]$; and close to a demonstrator so that the two robots can perceive each other through the infrared proximity sensors. The demonstrator acts according to what it has already (individually) learnt during previous trials. That is, it performs phototaxis if, in previous trials, it has not perceived any sound. Otherwise, it performs antiphototaxis. The learner is required to remain close to the demonstrator. At the end of the first trial, the demonstrator is removed. The light is randomly repositioned in the arena at a distance from the learner $dll_0 \in [55cm, 65cm]$. In the second trial, the learner is required to imitate the behaviour of the demonstrator by either approaching or going away from the light. In other words, the learner is said to have socially learnt if it behaves in the second trial as shown by the demonstrator in the first one (see Figure 1b).

4 Characteristics of the simulated agent

The controllers are evolved in a two-dimensional simulation environment which models the kinematics of simple geometries and functional properties of an e-puck robot[1]. Our simulated robot is modelled as a circular object of 3.5cm of radius. Differential drive kinematics equations, as presented in (Dudek and Jenkin, 2000), are used to update the position of the robots within the environment. The simulated robot is equipped with 8 infrared proximity sensors (IR_i) placed around the robot body, 2 ambient light sensors (AL_i) placed on the left and right sides of the robot, at ±90° with respect to the robot heading, and a sound sensor SS (see also figure 2a). The activation of the infrared proximity sensors are taken from a look-up table which contains data sampled from the real e-puck. At each time-step, the look-up table returns the agent's infrared

---

[1] Further details on the robot platform can be found at `www.e-puck.org`.



(a) (b)

Figure 2: (a) The simulated robot. IR_i with $i \in [0,7]$ are the infrared proximity sensors; AL_i with $i \in [0,1]$ the ambient light sensors; SS the sound sensor; Ml the left motor and Mr the right motor. (b) Network architecture. Only the efferent connections for the first node of each layer are drawn.

sensors readings based on distances and angles of the obstacles with respect to the position and heading of the robot. 20% uniform noise with respect to maximum activation value is added to these readings. Sound is modelled as an instantaneous, constant field of single frequency and amplitude. The sound sensor's readings are set to 0 when no sound is emitted, and 1 when sound is broadcast in the environment. No noise is added to the sound sensor. The ambient light sensors perceive a light source up to a distance of 150cm, and each sensor has a receptive field of 60°. The ambient light sensors readings are set to 1 when the light source is inside the sensor's receptive field, 0 otherwise. Before computing the readings, 5% uniform noise is added to the distance and orientation of the light with respect to the robot position and heading.

5 The agents' controller

The aim of this study is to design single integrated (i.e., not modularised) controllers capable of guiding the robots both in the individual and in the social learning task. Given the nature of the tasks, we decided to work with a 16 neuron Continuous Time Recurrent Neural Networks (CTRNNs, (see also Beer and Gallagher, 1992)) with a layered topological structure as shown in figure 2b. In particular, each network is made of three inter-neurons and an arrangement of eleven sensory neurons and two output neurons. The sensory neurons receive input from the agent sensory apparatus (i.e., infrared, ambient light, and sound sensors). The inter-neuron network is fully connected. Additionally, each inter-neuron receives one incoming synapse from each sensory neuron. Each output neuron receives one incoming synapse from each inter-neuron. There are no direct connections between sensory and output neurons. The states of the output neurons are used to set the speed of the robot wheels. The network neurons are governed by the following equation:

$$\frac{dy_i}{dt} = \begin{cases} \frac{1}{\tau_i}\left(-y_i + gI_i\right); & i \in [0,10] \\ \frac{1}{\tau_i}\left(-y_i + \sum_{j=1}^{k} \omega_{ji}\sigma(y_j + \beta_j) + gI_i\right); & i \in [11,15]; \sigma(x) = \frac{1}{1+e^{-x}} \end{cases} \tag{1}$$

where, using terms derived from an analogy with real neurons, y_i represents the cell potential, τ_i the decay constant, g is a gain factor, I_i the intensity of the perturbation on sensory neuron i, ω_{ji} the strength of the synaptic connection from neuron j to neuron i, β_j the bias term, $\sigma(y_j + \beta_j)$ the firing rate. The cell potentials y_i of the 15^{th} and the 16^{th} neuron, mapped into [0.0,1.0] by a sigmoid function σ and then linearly scaled into [-1000, 1000], set the robot motors output.

6 The evolutionary algorithm

A simple generational genetic algorithm is employed to set the parameters of the networks (see Goldberg, 1989). The population contains 80 genotypes. Generations following the first one are produced by a combination of selection with elitism, recombination and mutation. For each new generation, the three highest scoring individuals ("the elite") from the previous generation are

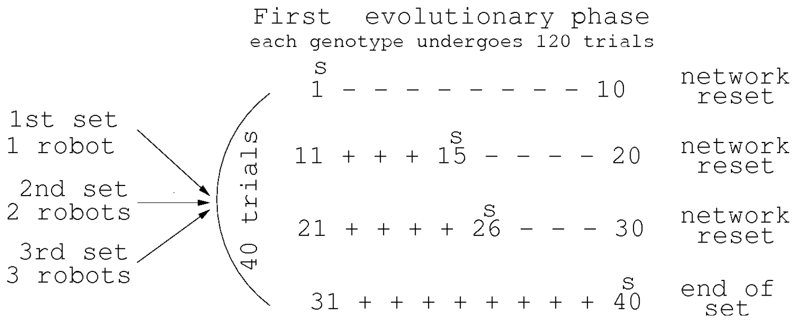


Figure 3: First evolutionary phase. Each genotype is evaluated for 120 trials. That is, 40 trials in the single robot case (i.e., first set), 40 trials in the two robots case (i.e., second set), and 40 trials in the three robots case (i.e., third set). s indicates the trials in which a tone is emitted. In those trials which precede the emission of the tone the robot/s are rewarded for performing phototaxis—these trials are indicated with the sign +. In those trials which follow the emission of the tone the robot/s are rewarded for performing antiphototaxis—these trials are indicated with the sign -.

retained unchanged. The remainder of the new population is generated by fitness-proportional selection (also known as roulette wheel selection) from the individuals of the old population. Each genotype is a vector comprising 69 real values (i.e., 48 connection weights, 16 decay constants, 4 bias terms, and a gain factor). Initially, a random population of vectors is generated by initialising each component of each genotype to values randomly chosen from an uniform distribution in the range [0,1]. New genotypes, except “the elite”, are produced by applying recombination with a probability of 0.03 and mutation. Mutation entails that a random Gaussian offset is applied to each real-valued vector component encoded in the genotype, with a probability of 0.08. The mean of the Gaussian is 0, and its standard deviation is 0.1. During evolution, all vector component values are constrained to remain within the range [0,1]. Genotype parameters are linearly mapped to produce network parameters with the following ranges: biases $\beta_i \in [-4, -2]$ with $i \in$ [0,10][2], biases $\beta_i \in [-5, 5]$ with $i \in$ [11, 15]; weights $\omega_{ij} \in [-8, 8]$ with $i \in [0, 10]$ and $j \in [11, 13]$, weights $\omega_{ij} \in [-10, 10]$ with $i \in [11, 13]$ and $j \in [14, 15]$; gain factor $g \in [1, 13]$ for all the input nodes. g is set to 1 for the other neurons; decay constants τ_i with $i \in$ [0,15] are exponentially mapped into $[10^{-1}, 10^{1.6}]$ with the lower bound corresponding to the integration step-size used to update the controller and the upper bound, arbitrarily chosen, corresponds to about $\frac{2}{3}$ of the maximum length of a trial (i.e. 60s). Cell potentials are set to 0 any time the network is initialised or reset, and circuits are integrated using the forward Euler method with an integration step-size of $dt = 0.1$.

7 The fitness function

In this section we provide the details of the fitness function and of the evolutionary process we employed to design the control structure for a robot capable of learning both individually and socially the correct response to the light. To achieve our goal, we employed an incremental approach made of two evolutionary phases. The first evolutionary phase is meant to produce population of agents which are subsequently used to create the populations at generation zero of the second phase evolutionary runs. Robots of the first evolutionary phase are selected for being capable of performing individual learning and group motion. Robots of the second evolutionary phase are selected for being capable of performing individual and social learning. In the following we detail the characteristics of both evolutionary phases.

7.1 The first evolutionary phase

The first evolutionary phase consists of twelve randomly seeded evolutionary processes. Each genotype is evaluated for 120 trials composed of three sets of 40 trials. In the first set, a single robot is scored for its capability to perform individual learning. In the second and third sets, homogeneous groups of two and three robots are respectively considered. The social environment

---

[2] The same bias is used for all the input nodes.

(i.e., groups of two/three robots) does not have any significance with respect to the individual learning task. Groups are employed to evaluate the robots for their capabilities to remain close to each other while individually performing the learning task. That is, selective pressures favour the evolution of group of agents capable of both individual learning and group motion.

At the beginning of each trial the light is randomly repositioned in the arena at a distance $d_0 \in [55cm, 65cm]$ from the robot/s. In the second and third set of trials, the robots are randomly positioned at less than 1cm from each other with randomly chosen orientations. A trial refers to a sequence of 600 simulation cycles (i.e., 60 simulated seconds) in which the light is not moved. Every 10 trials the robot/s controllers are reset (see section 6). Sound is broadcast at trials 1, 15, 26, 40 of each set (see figure 3). By drawing inspiration from the fitness function detailed in Quinn et al. (2003), the fitness of each genotype $F_i^{'}$ is computed in each trial i as follows:

$$F_i^{'} = \left(\sum_{t=1}^{T} \left[f\left(d_t, d_{t-1}\right) \left(1 - tanh(s_t)\right) \right] \right) c_p. \tag{2}$$

$T \in [1, 600]$ is the length of a trial in term of time-steps or simulation cycles. $f\left(d_t, d_{t-1}\right)$ is a component which rewards the correct movements of the robot/s with respect to the light source. In the single robot trials, d_t is the distance between the light and the robot centre of mass at time-step t. In the second and third set of trials—with groups of two/three robot—d_t is the distance between the centroid of the group—i.e., the geometric centroid of the line/triangle formed by the centres of the two/three robots—and the light at time-step t. In those trials in which phototaxis is required, $f\left(d_t, d_{t-1}\right)$ is computed at each time-step as follows:

$$f\left(d_t, d_{t-1}\right) = \begin{cases} \left(d_{t-1} - d_t\right) / \left(d_0 - d_{min}\right) & \text{if } d_{t-1} > d_t;\ d_{min} = 5cm \\ 0 & \text{otherwise} \end{cases}. \tag{3}$$

In those trials in which antiphototaxis is required, $f\left(d_t, d_{t-1}\right)$ is computed at each time-step as follows:

$$f\left(d_t, d_{t-1}\right) = \begin{cases} \left(d_t - d_{t-1}\right) / \left(d_{max} - d_0\right) & \text{if } d_{t-1} < d_t;\ d_{max} = 1.5 d_0 \\ 0 & \text{otherwise} \end{cases}. \tag{4}$$

The components c_p and s_t are considered only in sets of trials with more than one robot. In particular, c_p is a collision penalty component, such that $c_p = 1 - c/\left(2c_{max}\right)$, with c corresponding to the mean number of collisions between robots and $c_{max} = 20$ the maximum number of collisions allowed. s_t, is a penalty for team's dispersal at time-step t. If each robot is closer than the infrared sensors' range to at least another robot then s_t is 0. Otherwise, for groups of two, s_t corresponds to the amount of which the robot-robot distance exceeds the infrared sensors' range. For groups of three robots, the 2 shortest lines connecting the robots are found and s_t is the amount by which the longest of these lines exceeds the infrared sensors' range. $F_i^{'}$ is not computed in those trials in which sound is played (i.e., trial 1, 15, 26, and 40).

Note also that, each trial can be terminated earlier either because (i) the robot/s reaches the proximity of the light source (i.e., $d_t < d_{min}$); (ii) the robot/s reaches the maximum distance from the light (i.e., $d_t > d_{max}$); (iii) in the social environment, the robots collide more than 20 times (i.e., $c > c_{max}$); (iv) the fitness score decreases for more than 25 consecutive time-steps.

7.2 The second evolutionary phase

The second evolutionary phase is also made of twelve evolutionary processes. However, contrary to the first phase, these processes do not start from scratch. That is, instead of being randomly seeded, the genotypes at generation 0 of the second phase evolutionary runs are generated by using the best three evolved genotypes of the first evolutionary phase (henceforth we refer to them as the seeding-genotypes). In details, each seeding-genotype is used to create four populations of 80 different genotypes by applying the mutation operator as described in section 6. In this second evolutionary phase, the fitness function has been designed in order to make sure that the robots do not lose the behavioural characteristics they evolved during the first evolutionary phase. In principle, the robots at generation 0, should be capable of individually learn and tolerate the presence of another robot[3]. Moreover, a selective advantage is given to those individuals which

---

[3] Note that, the capability of individual learning and group motion may be lost due to the effect of random mutations applied to the seeding-genotypes.

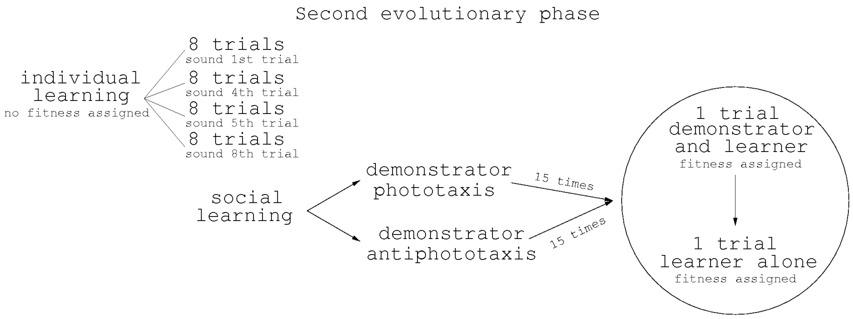


Figure 4: Second evolutionary phase. Each genotype undergoes a first set of 32 trials in which it is evaluated at the individual learning task (see top of the picture, and also caption figure 3). If the genotype managed to successfully complete 25 out of the 32 individual learning trials then it is allowed to undergo a subsequent sets of trials (16x2 trials for phototaxis and 16x2 trials for antiphototaxis) in which it is evaluated at the social learning task (see bottom of the picture). The demo-trial is the one in which the demonstrator and the learner are placed in the arena close to each other. The copy-trial is the one in which the learner (alone) is required to imitate the behaviour shown by the demonstrator in the demo-trial. Before the demo-trail, the demonstrator is taught what action to display to the learner.

prove capable of performing social learning. This methodological approach has been chosen to facilitate the design of a single controller capable of providing the robots the required neural structures to learn both individually and socially. In the remaining of this section, we detail how the genotypes are evaluated.

Each genotype is transformed into a controller which is evaluated for 96 trials, 32 trials for individual learning and 64 trials for social learning. In the first set of 32 trials, each controller guides a robot that is required to individually learn under the conditions previously illustrated for the first set of trials of the first evolutionary run. The behaviour of the corresponding robot is evaluated by using the fitness function detailed in the equation 2 (see section 7.1 for details). Those genotypes that “give birth” to robots that successfully complete 25 trials out of 32 (i.e., $\frac{1}{N}\sum_{i=1}^{N} F_i^{'} > 0.87$ with $N = 32$) at the individual learning task, undergo the second set of 64 evaluation trials of social learning (see figure 4). For the others, the evaluation terminates with fitness set to zero.

For those genotypes that are allowed to continue the evaluation, their corresponding controllers are cloned in two robots, one of which becomes the demonstrator, and the other becomes the learner. Learner and demonstrator are demanded to perform a trial together (hereafter referred to as the demo-trial). At the beginning of the demo-trial, the learner corresponds to the robot that has the neurons’ state of its controller set to zero; the demonstrator corresponds to the robot that has the neurons’ state of its controller set so that it performs either phototaxis or antiphototaxis. The values of the demonstrator’s neurons states are obtained from previous successful trials in which the controller guides a single robot required to switch from phototaxis to anti-phototaxis by reacting to the perception of a tone. In the demo-trial, the learner is evaluated for its capability to follow the demonstrator while approaching or moving away from the light by the following fitness function:

$$F_i^{''} = c_p\left(0.9 - ddl_t/ddl_{max} + k\right); \tag{5}$$

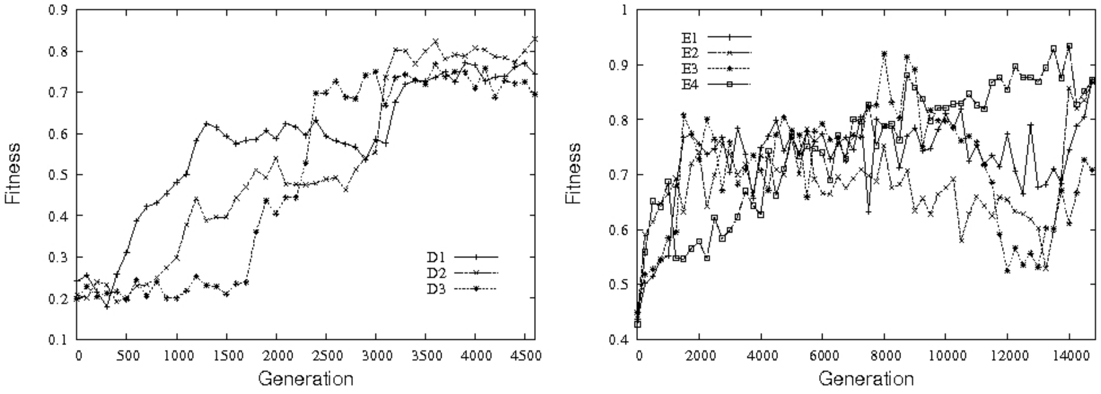
c_p is the collision penalty (see equation 2); ddl_t is the maximum demonstrator-learner distance reached during the trial. $ddl_{max} = 100cm$ is the maximum allowed demonstrator-learner distance. k is set equal to 0.1 if i) during phototaxis trials, $dll_t < dll_{min}$ with dll_t corresponding to the learner-light distance and $dll_{min} = 5cm$ to the minimum allowed learner-light distance; ii) during antiphototaxis, $dll_t > dll_{max}$ with $dll_{max} = 1.5dll_0$ corresponding to the maximum allowed learner-light distance; otherwise $k = 0$. If, for any reason, the demonstrator does not show to the learner the behaviour that is supposed to execute given the states of its neurons, the demo-trial is terminated, $F_i^{''}$ is set to zero, and the evaluation continues in the following way: 1) the demonstrator is taught again what action to display to the learner; 2) a new demo-trial is run. Demo-trial can be terminated earlier than the maximum length of a trial (i.e., $T = 600$ time-steps) due to the same circumstances listed in the previous section as well as in case $ddl_t > ddl_{max}$.

At the end of the demo-trial, the demonstrator is removed. The learner performs one more trial (hereafter referred to as the copy-trial), in which it is supposed to imitate whatever action the demonstrator performed with respect to the light in the demo-trial (i.e., phototaxis or antiphototaxis). In the copy-trial, the behaviour of the learner is evaluated according to the following fitness function:

$$F_i^{'''} = \begin{cases} \frac{(dll_0 - dll_{min})}{dll_0 - d_{min}} & \text{if phototaxis is required, } d_{min} = 5cm \\ \frac{(dll_{max} - dll_0)}{0.5 dll_0} & \text{if antiphototaxis is required} \end{cases} . \tag{6}$$

The sequence of 2 trials (i.e., demo-trial followed by the copy-trial) is repeated 16 times with the demonstrator instructed to perform phototaxis and 16 times with the demonstrator instructed to perform antiphototaxis. Within the set of 64 trials, half of the time the light is positioned in front of the robots and half of the time behind the robots. Moreover, half of the time the learner is placed on the right of the demonstrator and half of the time on its left. The demonstrator is taught again what action to display to the learner before each demo-trial. The learner's controller is reset at the end of each copy-trial.

8 Results



(a) (b)

Figure 5: (a) First evolutionary phase: fitness of the best genotypes at each generation of the best 3 out of 12 evolutionary runs. (b) Second evolutionary phase: fitness of the best genotypes at each generation of the best 4 out of 12 evolutionary runs.

As explained in section 7, in order to achieve our goal we employed an incremental approach made of two subsequent evolutionary phases. The first evolutionary phase aims at the evolution of agents capable of performing the individual learning task—that is, agents capable of switching from phototaxis to antiphototaxis at the perception of a tone. Figure 5a shows the fitness of the best agent at each generation of the best three first phase evolutionary runs[4]. Given the characteristics of $F_i^{'}$—detailed in equation 2—the maximum fitness score an agent can obtain during evolution corresponds to 0.9. Since the maximum score has not been reached by any of the best robots, we run a series of post-evaluations tests. These tests proved that robots with a score higher than 0.8 during evolution are perfectly capable of switching from phototaxis to antiphototaxis after the perception of a tone (data not shown[4]). We inferred that the small fitness loss during evolution are mostly due to penalties for group dispersal—controlled by the factor $(1 - tanh(s_t)$ in equation 2—in the second and third sets of trials. On the basis of the results of these post-evaluation tests, we selected the best three agents taken from different evolutionary runs. Subsequently we used the genetic material of these three agents to create the genome at generation zero of the second evolutionary phase. Owing to this seeding procedure (explained in details in section 7.2), the second phase evolutionary runs do not start from scratch. That is, by inheriting the genetic material of the best evolved agents of the first evolutionary runs, the

---

[4] For the sake of conciseness, the paper illustrates only a subset of all the evaluations/post-evaluations employed to describe the performances of the robotic system. Graphs and tables show only data/curves referring to a subset of all the individuals that have been evaluated. The reader can find an exhaustive illustration of all the results on `http://iridia.ulb.ac.be/supp/IridiaSupp2007-008`.

generation zero agents represent a favorable "starting point" for evolutionary processes that aim at the design of the mechanisms for social learning. This is because the generation zero agents are potentially capable of individually learn the response to the light and at the same time tolerate the presence of another robot. Consequently they can play the role of the demonstrator in a social context.

Figure 5b shows the fitness of the best agent at each generation of the best four second phase evolutionary runs. Given the way in which the fitness is computed (i.e., $F_i^{''}$ and $F_i^{'''}$ see equation 5 and 6 respectively), the maximum fitness score an agent can obtain during evolution corresponds to 0.97. The graph indicates that none of the best agents of the best four runs managed to get the maximum score during evolution. The graph also indicates that the fitness of these agents oscillates quite a lot throughout the evolution. The tuning of the parameters of the evolutionary algorithm—mainly those associated to the recombination, mutation, and selection operators—did not help to reduce the magnitude of the fluctuations of the fitness curves (data not shown). We think that, these fluctuations are probably determined by stochastic phenomena which bear upon the starting positions and relative orientation of the robots during social learning (i.e., in the demo-trial) as well as the nature of the training phase of the demonstrator before being placed close to the learner. In order to have a better estimate of the learning capabilities of the evolved agents we run a series of post-evaluation tests. The aim of these tests is to evaluate how good the best agents of each second phase evolutionary runs are in the individual and social learning tasks, under circumstances in which some of the stochastic phenomena are experimentally varied. The results of these post-evaluation tests are illustrated in the following sections.

8.1 Individual learning

Table 1: Results of post-evaluation tests aimed at evaluating the individual learning capabilities of the best four evolved agents (i.e., I1, I2, I3, and I4) taken from four different evolutionary runs (i.e., E1, E2, E3, and E4 see figure 5b). The agents are evaluated in condition F (the light is placed in front of the agent) and condition B (the light is placed behind the agent). Each condition refers to 8 different groups of 8000 trails in which the emission of sound is systematically varied from trial 1 to trial 8. For each condition, the table shows: 1) the success rate (S); 2) the rate of error type E_1 (the agent does phototaxis instead of antiphototaxis); 3) the rate of error type E_2 (the agent does antiphototaxis instead of phototaxis).

	F			B		
	S	E_1	E_2	S	E_1	E_2
I1	1.00	0.00	0.00	1.00	0.00	0.00
I2	0.965	0.00	0.035	0.965	0.00	0.035
I3	1.00	0.00	0.00	1.00	0.00	0.00
I4	1.00	0.00	0.00	1.00	0.00	0.00

In this section, we illustrate the results of post-evaluation tests aimed at evaluating the individual learning capabilities of the best evolved agents (i.e., I1, I2, I3, and I4) of the best four second phase evolutionary runs[4] (i.e., E1, E2, E3, and E4 see figure 5b). In order to inspect individual learning ability, each agent undergoes a series of evaluations in 8 different evaluation-groups made of 8 trials each, in which the time of emission of a tone is systematically varied from the first trial (in evaluation-group n. 1) to the eighth trial (in evaluation-group n. 8). In other words, the evaluation-groups differ in term of the trial in which the sound is emitted. For example, in evaluation-group n. 1, a tone is emitted during the first trial, while in evaluation-group n. 8 a tone is emitted during the eighth trial. The agent repeats each evaluation-group 1000 times. This means that the evaluation in group n. 1 is composed of 1000 first trials in which the learner perceives the sound, and 7000 trials in which the agent should perform antiphototaxis. The evaluation in group n. 2 is composed of 1000 first trials in which the agent is required to perform phototaxis, 1000 second trials in which the tone is emitted and 6000 trials in which the agent should perform antiphototaxis. The other evaluation groups follow the same principle. The robot's controller is always reset at the end of a sequence of 8 trials of each evaluation-group. The following post-evaluation tests (8 groups of 8 trials each, repeated 1000 times) has been repeated twice: the first time with the robot placed in the arena oriented towards the light; the second time with the robot oriented in the opposite direction with respect to the light. Moreover, in each trial, the initial orientation of the robot is determined by applying an angular displacement randomly chosen in the interval $[-10°, 10°]$ with respect to the facing direction associated to the evaluation-groups (i.e., pointing the light or the opposite direction of the light).

In these tests, the behaviour of the robots is scored according to a binary criterion (successful/unsuccessful). The robot is successful if a) when phototaxis is required, the agent reaches a distance from the light smaller than 5cm; b) when antiphototaxis is required, the agent reaches a distance from the light bigger than a time and an half the initial distance. Unsuccessful trials are those in which the robot does not fulfill the conditions mentioned above. The robot is not scored in trials in which the tone is emitted.

Table 1 shows the results at these post-evaluation tests of the best four evolved agents by distinguishing between trials in which the robot is placed oriented towards the light (i.e., condition F in table 1) and trials in which the robot is oriented in the opposite direction of the light (i.e., condition B in table 1). Given the high success rate in both F and B conditions (see table 1 second and the fifth column), we can conclude that the robots are extremely good in individually learn the correct response to the light by properly reacting to the perception of a tone regardless the trial in which the tone is broadcast. By looking at the behaviour of the agents, we noticed that all the robots employed the same strategy: a) regardless of the motion with respect to the light, they always act so that the light impinges on the right ambient-light sensor (i.e., AL_1, see figure 2a); b) they move forward as long as phototaxis is required; c) they reverse the wheels motion after the perception of a tone. The backward movement continues in all the trials in which antiphototaxis is required. This is by no means a simple reactive strategy as it may appear. Indeed, the robots do not simply switch from forward to backward motion at the perception of the tone. Such robots would be systematically unsuccessful in those trials in which their initial orientation in combination with the preferential wheels motion is not compatible with the required action. Consequently, they could not be so successful as they proved to be at the post-evaluation tests whose results are shown in table 1. For example, a robot that at the beginning of an antiphototaxis trial is facing the opposite direction of the light, it can not simple move backward because this would bring it on the proximity of the light rather than far away. By looking at the behaviour of the robots we noticed that, when facing such circumstances the agents make a 180° turn to act with respect to the light as required by the task (i.e., phototaxis or antiphototaxis) and in accordance with their preferential direction of motion (i.e., forward or backward motion). The 180° turn is made in order to bring the light within the receptive field of the right ambient-light sensor.

The forward/backward motions is the simplest mechanisms the robots can employ to accomplish phototaxis/antiphototaxis. This is due to the fact that the readings of the ambient-light sensors are not affected by changes in light intensity related to variation of the agent-light distances. In other words, since the intensity of the light that impinges on the agents ambient-light sensors does not increase or decrease while the agents are moving, these cues are not available to the agents to find out whether they are approaching or moving away from the light. Therefore, the association between forward/backward motion with phototaxis/antiphototaxis while always keeping the light on the right ambient-light sensor seems to be the most effective way to behave as required by the learning task.

In the next section, we show the results of a further series of post-evaluation tests aimed at evaluating whether the genotypes (I1, I2, I3, and I4) can also "give birth" to agents capable of socially learn the correct response to the light. Subsequently, we analyse the behavioural strategies in the social learning task.

8.2 Social learning

In this section, we illustrate the results of post-evaluation tests aimed at evaluating the social learning capabilities of the best evolved agents (i.e., I1, I2, I3, and I4) of the best four second phase evolutionary runs^4 (i.e., E1, E2, E3, and E4 see figure 5b). In order to inspect social learning ability, each agent undergoes a series of evaluations in 4 different evaluation-groups made of 9 trials each (i.e., one demo-trial followed by eight copy-trials, see also section 7.2 for details). The evaluation-groups differ from each other in the position of the light with respect to the agents starting position and in the demonstrator-learner relative positions during the demo-trial. In particular, in evaluation-group n. 1 and n. 2, the agents are initialised always oriented towards the light, while in evaluation-groups n. 3 and n. 4, they are initialised always oriented in the opposite direction of the light. In evaluation-group n. 1 and n. 3 the learner is initialised on the right

Table 2: Results of post-evaluation tests, limited to the antiphototaxis response, aimed at evaluating the social learning capabilities of the best four evolved agents (i.e., I1, I2, I3, and I4) taken from four different evolutionary runs (i.e., E1, E2, E3, and E4 see figure 5b). The agents are evaluated in four different starting conditions, FL, FR, BL, and BR. F (i.e., front) and B (i.e., behind) refers to the position of the light with respect to the heading of the agents. L (i.e., the demonstrator on the left side of the learner) and R (i.e., the demonstrator on the right side of the learner) refers to the relative positions of the agents. Each condition refers to 2000 post-evaluation trials. For each condition, the table shows: 1) the success rate (S); 2) the rate of error type E_3 (the learner does not follow the demonstrator in the demo-trial); 3) the rate of error type E_4 (the demonstrator does not show the correct response to the light in the demo-trial); 3) the rate of error type E_5 (the learner does not replicate in the copy-trials what previously shown by the demonstrator).

		F				B			
		S	E_3	E_4	E_5	S	E_3	E_4	E_5
I1	L	0.99	0.01	0.00	0.00	0.70	0.30	0.00	0.00
	R	0.88	0.12	0.00	0.00	0.89	0.11	0.00	0.00
I2	L	1.00	0.00	0.00	0.00	0.99	0.01	0.00	0.00
	R	0.95	0.05	0.00	0.00	0.98	0.02	0.00	0.00
I3	L	0.65	0.35	0.00	0.00	0.70	0.30	0.00	0.00
	R	0.83	0.17	0.00	0.00	0.48	0.52	0.00	0.00
I4	L	1.00	0.00	0.00	0.00	0.769	0.215	0.016	0.00
	R	0.46	0.54	0.00	0.00	0.90	0.10	0.00	0.00

of the demonstrator while in evaluation-group n. 2 and n. 4 the learner is initialised on the left of the demonstrator. Moreover, the initial orientation of each robot is determined by applying an angular displacement randomly chosen in the interval $[-10°, 10°]$ with respect to the facing direction associated to the evaluation-groups (i.e., pointing the light or the opposite direction of the light). The agents repeat each evaluation-group 1000 times. The following post-evaluation tests (4 evaluation-groups of 9 trials each, repeated 1000 times) have been repeated twice: the first time with the demonstrator instructed to perform phototaxis, and a second time with the demonstrator instructed to perform antiphototaxis.

Each sequence of demo-trial/copy-trials is preceded by a training-phase, that is a series of trials in which the demonstrator is taught what action to show to the learner. Note that learner and demonstrator share the same genetic material. The difference between the two resides in the states of the neurons of their controllers at the beginning of the demo-trial. In particular, the learner corresponds to the robot that has the neurons' state of its controller set to zero. The demonstrator corresponds to the robot that has the neurons' state of its controller set as defined by the preliminary training-phase so that, once placed close to the learner, it performs either phototaxis or antiphototaxis.

In these tests, the behaviour of the learner during the copy-trials is scored according to a binary criterion (successful/unsuccessful). The learner is successful if a) it manages to reach a distance from the light smaller than 5cm, in copy-trials following a demo-trial in which the demonstrator performed phototaxis; b) it manages to reach a distance from the light bigger than a time and an half its initial distance, in copy-trials following a demo-trial in which the demonstrator performed antiphototaxis. Unsuccessful copy-trials are those in which the learner does not fulfill the conditions above mentioned. In order to have a better estimate of the performances of the robots in the social learning context, we decided to distinguish between the following three types of errors: a) error type E_3 which corresponds to the case in which the learner fails because it was not capable of following the demonstrator during the demo-trial; b) error type E_4 which corresponds to the case in which the learner fails because the demonstrator did not show the correct response to the light during the demo-trial; c) error type E_5 which corresponds to the case in which the learner fails because it is not capable of imitating in the copy-trials what previously shown by the demonstrator. No score is assigned in demo-trials.

Table 2 shows the results at the post-evaluation tests of the best four evolved agents (i.e., I1, I2, I3, and I4) by distinguishing among the 4 different evaluation-groups FR, FL, BR, and BL. Note that, F (i.e., front) and B (i.e., behind) refers to the position of the light with respect to the heading of the agents. L (i.e., the demonstrator on the left side of the learner) and R (i.e., the demonstrator on the right side of the learner) refers to the relative positions of the agents. The table reports the results of only those series of evaluation-groups in which the demonstrator shows the learner an antiphototaxis response. The table with the data concerning phototaxis is not shown[4]. It turned out that the learners are quite successful (0.9 success rate) in imitating the actions of the demonstrator when the behaviour to execute is phototaxis in conditions in which the demonstrator is initialised on the left of the learner. More problematic appears to be the conditions in which the the demonstrator is initialised on the right of the learner, since we observed a performance drop, probably due to the agents difficulties in spatially rearranging in order to pursue phototaxis. Indeed, for certain agents, the success rate drops to 0.6. However, we have noticed that, in general phototaxis is the default action that any agent does if not instructed to perform antiphototaxis through the perception of a tone (i.e., by individual learning) or through the influence of a demonstrator (i.e., by social learning).

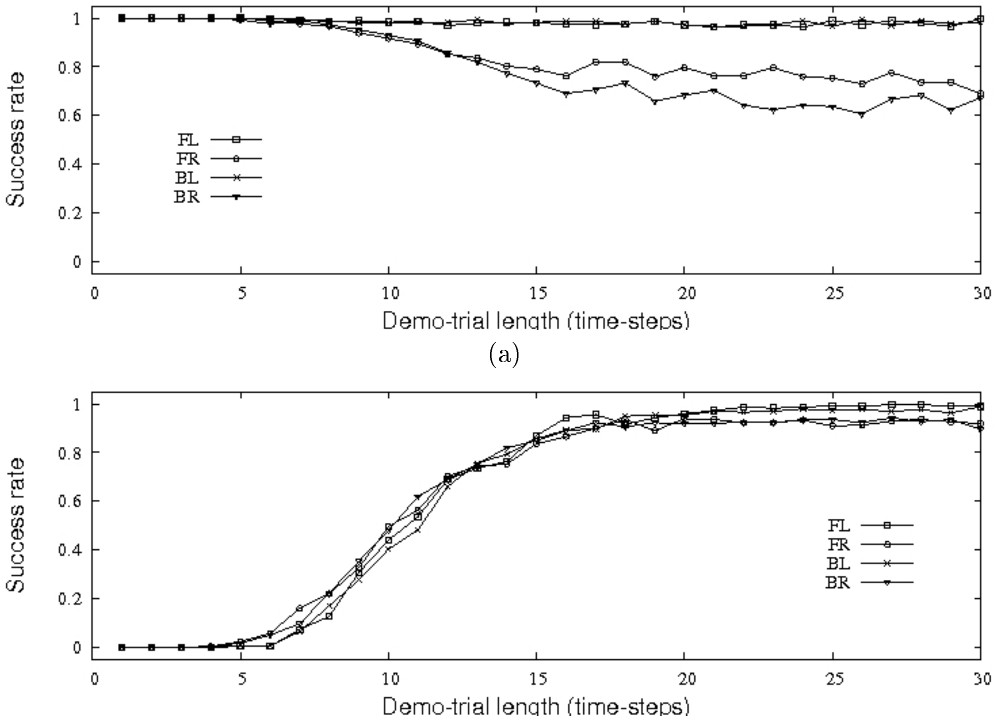
By looking at the success rate (S) of the learners (see table 2 third and seventh column), we can notice that the performances of the agents are quite good. In particular, learner I2 proved to by very successful in all the different evaluation-groups with a success rate higher than 0.95. With such a score, we can claim that I2 is capable of imitating the action (i.e, in this case antiphototaxis) shown by the demonstrator regardless of the two orientations of the light (i.e., front, back) and the relative position of the demonstrator in the demo-trial. Although quite successful, the results obtained by the other learners are not as homogeneous as the one obtained by I2. In particular, for I1, I3, and I4, the rate of success sensibly varies among the conditions. For example, the performance of learner I1 drops to 0.70 in the evaluation-group BL (i.e., both agents oriented facing the opposite direction of the light and the demonstrator initialised on the left side of the learner). The performance of learner I3 is quite good only in evaluation-group FR whereas the performance of learner I4 drops dramatically in FR (i.e., the demonstrator initialised on the right side of the learner and both agents oriented facing the light).

There are multiple reasons which can explain why the learners fail in certain conditions and not in others to imitate the behaviour shown by the demonstrator. However, before commenting these error rates, it is important to say that a) none of the learners has ever forgot what learnt during the demo-trial; b) none of the learners has ever recovered after an initial series of unsuccessful copy-trial. In other words, if successful at the first copy-trial, each learner continues to be successful in all the following copy-trials. On the contrary, if unsuccessful at the first copy-trial, each learner continues to be unsuccessful in all the following copy-trials. Since the learners do not show any forgetting, failure can be caused by either the incapacity of the demonstrator to show the learner the right response (i.e., error E_4 in table 2), or by the incapacity of the learner to learn during the demo-trial what shown by the demonstrator. In the latter case, we can further split the causes of failure between those circumstances in which the learner does not follow the demonstrator in the demo-trial (i.e., error E_3 in table 2), and those in which, despite a good demo-trial, the learner does not imitate the demonstrator in the first copy-trial (i.e., error E_5 in table 2). Data in table 2 clearly indicates that, in all the evaluation-groups and for all the learners, the cause of failure has to be attributed almost completely to the incapacity of the learner to follow the demonstrator during the demo-trial (see table 2 fourth and eighth column, rate of error E_3). By looking at the behaviour of the agents in the demo-trials preceding unsuccessful copy-trials, we noticed that almost always these trials end with the demonstrator far away from the light and the learner in the proximity of the light. In the next section, we look at the behavioural strategies employed by the agents in the social learning context to try to find out how the learner "learns" the appropriated response.

8.3 An initial analysis of the social learning strategies

What are the behavioural mechanisms which underpin social learning? In this section, we show the results of an analysis aimed at unveiling the strategies underpinning social learning behaviour in agents controlled by genotype I2—i.e., the best one at the social learning task (see table 2).

Clearly learning has to happen during the interactions—mediated by the infrared sensors—between the demonstrator and the learner, so that the learner, by following the demonstrator, manages to "learn" what action to take with respect to the light. Given what we said concerning the individual learning strategies (see section 8.1), we can exclude that the light is an important cue employed by the learner to find out whether the demonstrator is performing phototaxis or antiphototaxis. Recall that, the readings of the ambient-light sensors are not affected by changes in light intensity related to variation in the agent-light distances. Since the intensity of the light



(a)

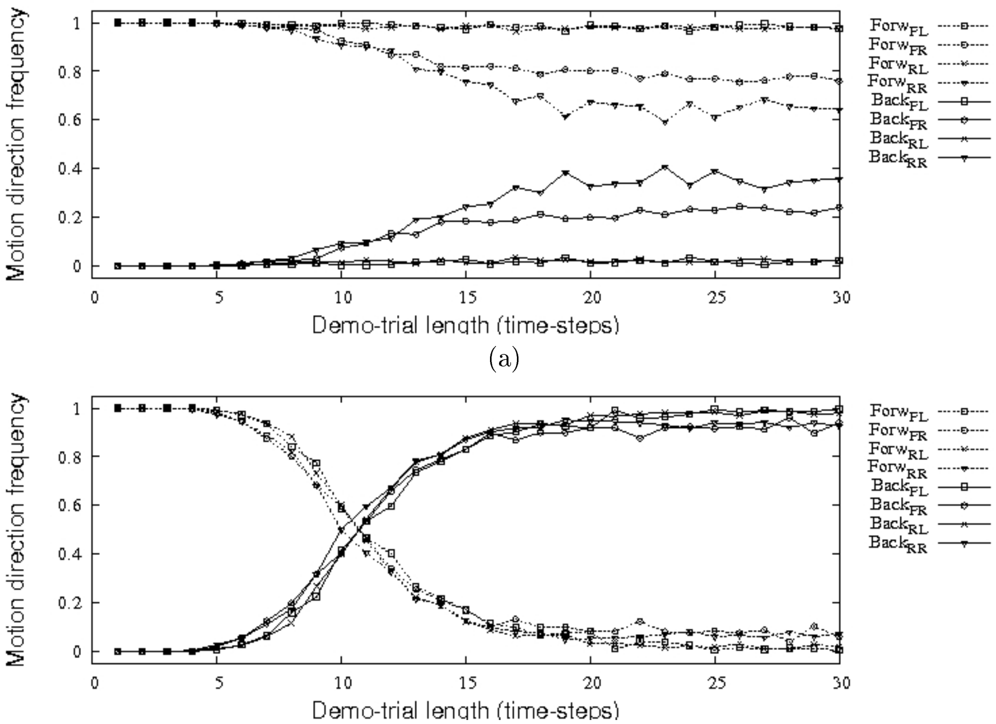
(b)

Figure 6: The graphs refer to post-evaluation tests in which the learner controlled by genotype I2, is evaluated in copy-trials following demo-trials of different length (from 1 time-step to 30 time-steps) in which the demonstrator performed phototaxis (see graph a) and antiphototaxis (see graph b). The graphs show the learner performances (i.e., success rate) in the copy-trials. For phototaxis and antiphototaxis, the learner is evaluated in four different starting conditions, FL, FR, BL, and BR (see caption table 2 for details).

that impinges on the learners ambient-light sensors does not increase or decrease while the agents are moving, these cues are not available to the learner to find out whether, while following the demonstrator, it is approaching or moving away from the light.

We already know that learner and demonstrator, by sharing the same genetic material, they share the same instinctive response as well. That is, both agents instinctively tend to bring the light within the receptive field of the right ambient light sensor, and they tend to perform phototaxis unless instructed to do otherwise. At the beginning of the demo-trial, the demonstrator has already undergone a training phase during which it has been taught what action to take by individual learning. The learner is obliged to interact with the demonstrator in order to find out whether the demonstrator is moving towards or away from the light. Our hypothesis is that social learning takes place through sensory-motor interactions, mediated by the infrared sensors, at the end of which the learner direction of motion matches the one of the demonstrator.

Figures 6a and 6b show the distribution of success rates in copy-trials of a learner controlled by the genotype I2, under conditions in which the length of the demo-trials is systematically varied. In particular, we repeated the post-evaluation tests described in the previous section, by systematically varying the length (i.e., the number of time-steps) of the demo-trials starting from 1 time-step up to 30 time-steps demo-trials. Figure 6a refers to tests in which the demonstrator performs phototaxis. As expected, interrupting the demo-trials before the 10^{th} time-step does not have much effect on the learner success rate in the following copy-trials. This is because, as already mentioned in section 8.2, the learner instinctively performs phototaxis unless taught to do otherwise. However, in condition FR and BR (see caption table 2 for details), the graph shows that interrupting the demo-trials after the 10^{th} time-step has a slight disruptive effect on the learner performances. Most probably, the starting positions and in particular the fact that the demonstrator is placed on the right of the learner, force the agents to initiate specific manoeuvres whose interruption is responsible for the disruptive phenomena observed. Further analysis on the behaviour of the agents under these conditions are required to understand what causes the slight



(a)

(b)

Figure 7: The graphs refer to post-evaluation tests in which the learner controlled by genotype I2, is evaluated in copy-trials following demo-trials of different length (from 1 time-step to 30 time-steps) in which the demonstrator performed phototaxis (see graph a) and antiphototaxis (see graph b). The graphs refer to the average frequency of forward and backward direction of movements of the learner in the corresponding copy-trials. For phototaxis and antiphototaxis, the learner is evaluated in four different starting conditions, FL, FR, BL, and BR (see caption table 2 for details).

drop in the learner performance.

Figure 6b refers to tests in which the demonstrator performs antiphototaxis. We clearly see that for demo-trials shorter than 10 time-steps, the learner is not capable of learning the correct response to the light. There is an interval between the 10^{th} and the 25^{th} time-step demo-trials in which the success rate of the learner in imitating what shown by the demonstrator increases with the increment of the length of the demo-trial. Any interruption of the demo-trial made after the 25^{th} time-step does not have any effect on the capability of the learner to imitate the demonstrator. The graphs clearly tell us that only few sensory-motor interactions between demonstrator and learner at the beginning of the demo-trials are sufficient to instruct the learner to switch from phototaxis (i.e., the “innate” response) to antiphototaxis.

Figure 7a and 7b provide further elements of the post-evaluation tests detailed above in this section. In particular, these figures show the distribution of the average frequency of forward and backward movements of the learner following different length demo-trials for different starting conditions (i.e., FL, FR, BL, BR) and with the demonstrator performing phototaxis (see figure 7a) and antiphototaxis (see figure 7b). By comparing figure 6a with figure 7a, and figure 6b with figure 7b, we clearly notice that while high success rate in phototaxis copy-trials corresponds to high frequency of forward movement, high success rate in antiphototaxis copy-trials corresponds to high frequency of backward movement. These correspondences confirm that, in our model social learning consists in the capability of the learner to copy the demonstrator direction of motion.

9 Conclusions

In this paper, we described a model in which artificial evolution is employed to design neural mechanisms that control the behaviour of learning robots. In particular, our research work focused on circumstances in which behavioural changes (i.e., a switch from phototaxis to antiphototaxis) are induced by either an environmental stimulus (i.e., a tone) or by social interactions between individuals. We referred to the former case as individual learning, and to the latter one as social learning (see section 3). The results proved that dynamical neural networks shaped by evolutionary computation techniques can allow a robot to individually and autonomously learn to invert its behaviour from phototaxis to antiphototaxis at the perception of a tone as well as a learner to imitate the behaviour of a demonstrator that has autonomously and individually learnt in a previous training-phase what action to show to the learner (i.e., phototaxis or antiphototaxis).

An important contribution of the paper is in illustrating a methodological approach, based on the idea on incremental evolution, that significantly differs from previous approaches to the study of social learning in autonomous robots (see section 7). Contrary to other studies in which the demonstrator follows hand-coded instructions and behavioural plasticity concerns only the learner, this study detailed the successful design of single integrated neural controllers capable of underpinning individual and social learning mechanisms (see section 2). In this model, learner and demonstrator are genetically identical; the demonstrator differs from the learner in having the state of its controller set as defined by the preliminary training-phase. Moreover, individual and social learning mechanisms are grounded on the sensory-motor experience of the robots and fully integrated with all the other underlying structures that underpin the robots behavioural repertoire. We believe that, these results are a small but significant step toward the development of robots with a larger autonomy.

Post-evaluation tests highlighted operational aspects of the model. In particular, we found out that during both individual and social learning, phototaxis and antiphototaxis are associated to forward and backward movement. The best evolved genotypes "give birth" to robots that instinctively approach a light source by moving forward. In the individual learning task, the perception of a tone induces the robots to change their behaviour with respect to the light by switching from forward to backward movement (see section 8.1). In the social learning task, a sequence of sensor-motor interactions between the learner and the demonstrator at the beginning of the demo-trial, allows the learner to imitate the direction of movement of the demonstrator (see section 8.2). In both individual and social learning task, in order to accomplish the appropriated actions, the robots combine the switch mechanism to change the direction of movement with another mechanism to pay attention to the relative orientation of the light with respect to their heading. In particular, the robots tend to keep the light source within the receptive field of the right ambient-light sensor.

We noticed that, in the social learning task, the learning strategies of the best evolved robots, are not as robust as we expected (see section 8.3). In particular, there are circumstances associated to the learner-demonstrator initial relative positions in which the robots found hard to engaged themselves in those interactions at the end of which the learner should imitate the direction of movement of the demonstrator. Our hypothesis is that this difficulties are mainly determined by the model of the infrared sensors we used in our simulation environment. As explained in section 4, we used a lookup table model, which integrates samples taken from real e-puck[1] robots. During the sampling, we realized that due to the characteristics of the hardware, there was a huge variability among the readings of different sensors mounted on a single e-puck, and also huge differences among the readings of sensors mounted on different robots. In spite of our effort to "correct" the sampling in order to create a model general enough to represent all the sensors sampled, certain biases could not be avoided. These biases break the symmetry between infrared sensors readings corresponding to symmetrical demonstrator-learner spatial relationships. Therefore, they hide environmental structures upon which the neural controllers build successful strategies. Further post-evaluations are required to test our hypothesis and to develop alternative solutions to improve the robustness of the best evolved controllers.

For the future, we intend to further develop the methodological approach described in this paper to more complex individual and social learning tasks. In other words, without giving up with the idea of having genetically identical demonstrators and learners, we will investigate more complex scenarios in which the complexity is linked to the nature of the associations to individually and socially learn as well as by the mechanisms of the social interactions. In particular, we are thinking about scenarios in which social interactions are mediated by forms of communications based on compositional and recursive syntactical structures.

10 Acknowledgements

The authors thank Marco Montes, Christos Ampatzis, and their colleagues at IRIDIA for stimulating discussions and feedback during the preparation of this paper. E. Tuci acknowledge European Commission support via the*ECAgents* project, funded by the Future and Emerging Technologies programme (grant IST-1940).

References

Beer, R. D. and Gallagher, J. C. (1992). Evolving dynamic neural networks for adaptive behavior. *Adaptive Behavior*, 1(1):91–122.

Billard, A. (2002). Imitation: A means to enhance learning of a synthetic proto-language in an autonomous robot. In Dautenhahn, K. and Nehaniv, C. L., editors, *Imitation in Animals and Artifacs*. MIT Press.

Billard, A. and Matarić, M. J. (2001). Learning human arm movements by imitation: Evaluation of biologically inspired connectionist architecture. *Robotics and Autonomous Systems*, 37:145–160.

Boyd, R. and Richerson, P. J. (1989). Social learning as an adaptation. *Lectures on mathematics in the life sciences*, 20.

Boyd, R. and Richerson, P. J. (1996). Why culture is common, but cultural evolution is rare. *Proceedings of the british academy*, 88:77–93.

Breazeal, C. and Scassellati, B. (2002). Challenges in building robots that imitate people. In Dautenhahn, K. and Nehaniv, C. L., editors, *Imitation in animals and artifacts*, pages 363–390. MIT Press.

Dautenhahn, K. (2007). Methodology and themes of human-robot interaction: A growing research field. *International Journal of Adavance Robotic Systems*, 4(1):103–108.

Dautenhahn, K. and Nehaniv, C. L., editors (2002). *Imitation in animals and artifacts*. MIT Press.

Demiris, J. and Hayes, G. (1994). A robot controller using learning by imitation. In *Proceedings of the* 2^{nd} *international symposium on intelligent robotic systems*.

Di Paolo, E. A. (2003). Evolving spike-timing-dependent plasticity for single-trial learning in robots. *Phil. Trans. of the Royal Soc. of London, Series A*, 361:2299 – 2319.

Dudek, G. and Jenkin, M. (2000). *Computational principles of mobile robotics*. Cambridge, UK: Cambridge University Press.

Flinn, M. V. (1997). Culture and evolution of social learning. *Evolution and human behavior*, 18:23–67.

Fragaszy, D. M. and Visalberghi, E. (1999). Social processes affecting the appearance of innovative behaviors in capuchin monkeys. *Folia Primatologica*, 54:155–165.

Galef, B. G. and Laland, K. N. (2005). Social learning in animals: Empirical studies and theoretical models. *BioScience*, 55(6):489–499.

Galef, B. G. and Zentall, T. (1988). *Social learning: Psychological and biological perspective*. Hillsdale, NJ: LEA.

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization and machine learning*. Reading, MA: Addison-Wesley.

Harvey, I., Di Paolo, E. A., Wood, R., Quinn, M., and Tuci, E. (2005). Evolutionary robotics: A new scientific tool for studying cognition. *Artificial Life*, 11(1/2):79–98.

Heyes, C. M. and Galef, B. G., editors (1996). *Social learning in animals: the roots of culture*. Academic Press, London.

Matarić, M. J. (2002). Sensory-motor primitives as a basis for imitation: Linking perception to action and biology to robotics. In Dautenhahn, K. and Nehaniv, C. L., editors, *Imitation in animals and artifacts*, pages 391–422. MIT Press.

Nehaniv, C. L. and Dautenhahn, K., editors (2007). *Imitation and social learning in robots, humans and animals.* Cambridge University Press.

Nolfi, S. and Floreano, D. (2000). *Evolutionary robotics: The biology, intelligence, and technology of self-organizing machines.* Cambridge, MA: MIT Press.

Pollard, N. and Hodgins, J. (2002). Generalizing demonstrated manipulation tasks. In *Proceedings of the workshop on the algorithmic foundations of robotics (WAFR '02).*

Quinn, M., Smith, L., Mayley, G., and Husbands, P. (2003). Evolving controllers for a homogeneous system of physical robots: Structured cooperation with minimal sensors. *Philosophical Transactions of the Royal Society of London, Series A*, 361:2321–2344.

Thorpe, W. H. (1963). *Learning and instinct in animals.* London: Methuen.

Tomasello, M. (1999). *The cultural origins of human cognition.* Harward University Press.

Tomasello, M. and Call, J. (1997). *Primate cognition.* Oxford University Press.

Tuci, E., Quinn, M., and Harvey, I. (2003). An evolutionary ecological approach to the study of learning behaviour using robot-based model. *Adaptive Behavior*, 10(3-4):201–221.

Zentall, T. R. (2006). Imitation: denitions, evidence, and mechanisms. *Animal Cognition*, 9:335–353.