

Learning on a Continuum in Evolved Dynamical Node Networks

Eduardo Izquierdo-Torres and Inman Harvey
Centre for Computational Neuroscience and Robotics
University of Sussex
{e.j.izquierdo, inmanh}@sussex.ac.uk

Abstract

In artificial life, there has been much previous research using evolution to generate learning behaviour within dynamical system controllers without pre-defining the learning mechanisms; so far this research has focused exclusively on evolving agents that can behave differently in a discrete number of different scenarios, generally two. But many (arguably most) interesting discrimination tasks in real life are where the scenarios are over a continuum; one example would be parental imprinting in birds. Here we analyse a successfully evolved embodied and situated agent on an abstract model of this imprinting and give the first published example of such learning on a continuum.

Introduction

Learning and evolution are two of the most fundamental adaptive processes in nature. Evolution modifies the behaviour of populations of individuals over many generations, whilst learning modifies the behaviour of organisms during their lifetime. These are two processes that operate at very different time-scales but whose interaction is crucial for the generation of adaptive behaviour. The advantages of learning behaviour for organisms are manifold. In particular, the ability to learn enables an organism to cope with changes in the environment, even if their ancestors have never before experienced such changes.

Over the last decade, various people have evolved dynamical system controllers to perform different behaviours according to the environment in which they are in. All of this work has in practice been limited to a discrete set of environments, generally two; we cite some of this work in the following section. This paper focuses on demonstrating for the first time that agents can be evolved to learn to discriminate on some environmental variable that has a continuum of possible values.

When the changes that can happen to the environment are a finite known set then an agent which is evolutionary trained on each of these different scenarios can evolve a similar finite number of modes of interaction that enable it to cope with them. The learning behaviour in this case would be the swapping between the different modes of interaction

driven by the changes in the environment. However, organisms can also sometimes adapt to challenges and scenarios that have not previously been experienced. This is the case when, for example, the set of possible scenarios is infinite. In such challenges, it will always be possible for an agent to have to cope with a scenario which it had never trained on previously. The question that we are interested in is, can we evolve systems that can cope with unseen environments? This type of challenge would require a successful agent to be able to generalise from the set of training examples given. A good example of this is imprinting in birds (Lorenz, 1981).

We develop here a situated and embodied minimal version of an imprinting-like scenario, in order to start to answer for the first time (so far as we are aware) questions regarding learning phenomena on a continuum, using this methodology of evolving dynamical node networks. The idea of the proposed task and model agent follows the minimally cognitive behaviours approach (Beer, 1996), on the one hand simple enough to be analytically tractable, while at the same time bearing some resemblance to learning challenges faced in the real world by living organisms.

Related Work

Learning behaviour has traditionally been associated with the modification of neural network's parameters, especially with the changes of the synaptic connections or weights of the neural network during the lifetime of the individual. In (Yamauchi and Beer, 1994) a more integrated view of learning behaviour is presented where the agent's 'normal' behaviour and 'learning' behaviour are both generated by the same dynamical systems controller¹, without the need either for lifetime parameter modifications or for a clear-cut distinction between both mechanisms. More recently, in (Phattanasri, 2002) there is presented an in-depth analysis of the dynamics of agents that were evolved for an ab-

¹This system is also in the form of a network of nodes but it simply implements a continuous-time dynamical systems which need not be restricted only to 'neural networks' but can actually encompass a broader set of possibilities, including for example nodes as a population of neurons or a particular chemical in the body.

stract associative learning task as well as a comparison with weight-changing additions. In (Tuci et al., 2003) work was extended in this domain by using an even more integrated approach: (a) the system was evolved as a whole (as opposed to the approach in (Yamauchi and Beer, 1994) of first evolving behavioural modules) and (b) embodied and situated simulations as well as robotic experiments were included (as opposed to the abstract scenarios in (Phattanasri, 2002)). However, all of this work involves learning to swap behaviours between, in practice, n different modes of interaction according to which scenario the agent finds itself in (e.g. going to the light in landmark-near scenarios or going away from the light in landmark-far ones). As far as the authors are aware there is no work in the literature to date on the evolution of dynamical systems able to discriminate over a continuum of features.

One type of learning behaviour which particularly lends itself this is that observed in imprinting in birds. Experiments by ethologist Konrad Lorenz (Lorenz, 1981) provide good examples of a form of learning behaviour which requires the subjects to be able to discriminate between different objects (i.e. who is the parent and who is not) using features on a continuum in their environment (i.e. what the parent looks like) and to adapt their behaviour according to this history of interaction. For example, in geese (also in antelopes and sheep), in order to keep up with their moving parents, they must first learn to recognise them. Other examples include song learning in birds, where for various reasons birds learn to sing their species-specific song from what they hear during their first days of life.

Here we develop a minimalistic learning task loosely inspired by this. In (Izquierdo-Torres and Harvey, 2005) a disembodied version of an imprinting scenario was first presented. However, the agent’s embodiment and situatedness is understood to play a fundamental role in the generation of behaviour and can sometimes ease the burden on the complexity of the internal dynamics (as for example in (Izquierdo-Torres and Di Paolo, 2005)). More concretely, in geese, for example, it is the physical act of following that triggers the learning (Lorenz, 1981) (i.e. a passively carried goose behind its parent will not learn). For these reasons, the work presented herein is an embodied and situated version of the imprinting-like learning scenario.

Methods

We use an agent and environment set-up similar to (Beer, 1996), using a real-valued genetic algorithm to evolve dynamical systems controllers with slight variations on the agent’s architecture and environment. We require the agents to be able to identify the first-presented object (their ‘parent’) by discriminating between it and other objects (‘strangers’) when a parent or stranger is subsequently presented, the identifying feature for a parent is changed between trials. The model has been chosen for two reasons:

its simplicity, and its potential for sufficiently interesting behaviours that could be called minimally cognitive.

Agent’s Body and Environment. The agent is situated and embodied, meaning that its sensory input is at all times co-determined by itself as well as by the interaction with its environment. This is so as to avoid constraining the solution to arise from the internal dynamics alone, but rather to ensure that it arises from the interactions as well. The agent has a circular body with a diameter of 30 units. The agent’s ‘eye’ consists of seven rays distributed uniformly between $\pm\pi/12$. An intersection between a ray and an object causes an input to be injected into the corresponding sensory node, with a magnitude inversely proportional to the distance. For distances over 220, the input is set to zero, and zero distance gives the maximum input. The agent can move horizontally under the influence of ‘left’ and ‘right’ motors as objects fall from above (see Figure 1A); the horizontal velocity is proportional to the sum of the opposing forces produced by these two motors. Circle-shaped objects of radii between [10, 20] fall straight down from the sky at constant speed -3; these represent possible parents or strangers.

Agent’s Controller. We would like to model the agent’s internal dynamics using the lowest level building blocks that can generate as many dynamics as possible. Hence we use a dynamical system implemented by a continuous-time recurrent node² network (CTRNN) with the following state equation (Beer, 1996):

$$\tau_i \dot{y}_i = -y_i + \sum_{j=i}^N w_{ji} \sigma(g_j (y_j + \theta_j)) + I_i \quad (1)$$

where y is the activation of each node, τ is its time constant, w_{ji} is the strength of the connection from the j^{th} to the i^{th} node, θ is a bias term, g is a gain, $\sigma(x) = 1/(1 + e^{-x})$ is the standard logistic activation function, I represents an external input (e.g. from a sensor) and N represents the number of nodes in the network. In simulation, node activations are calculated forward through time by straightforward time-slicing using Euler integration with a time-step of 0.1.

The network’s architecture consists of a layer of seven sensory nodes projecting to a layer of nine inter-nodes, which in turn projects to a layer of two motor nodes that control horizontal motion (see Figure 1B). The inter-node layer is fully interconnected, including self-connections, and it is only this inter-node layer that acts as a proper CTRNN, with time-parameters genetically specified with a minimum possible value 1; this allows the Euler integration, with a smaller time-step, to emulate continuous time dynamics within this layer. All the sensory nodes have their time-parameters set to the Euler time-step 0.1, which makes them stateless; they

²Neural is the term that is commonly used, we use the term nodes instead to denote a broader set of possibilities.

are in effect place-holders reflecting the instantaneous sensory input, with no memory. Within the sensory layer, the sensory inputs are mapped through the bias parameter, the gain parameter and a sigmoid function, before reaching the inter-node layer. Similarly, the motor node layer is stateless, and outputs from the internodes are mapped through bias, gain and sigmoid before reaching the motors. The sensory layer shares one genetically specified bias and one gain, the motor layer shares one genetically specified bias. The gain parameters for the motor layer and the inter-node layer are set to one. The network architecture is not constrained to be bilaterally symmetric.

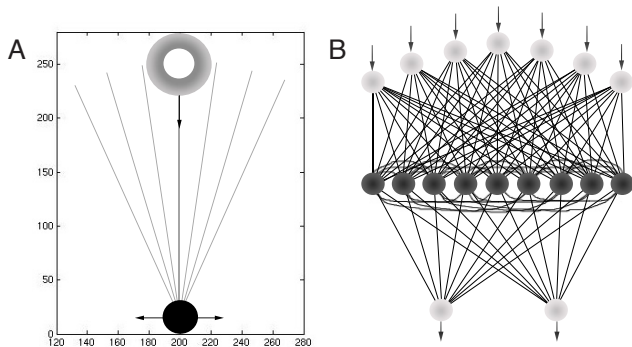


Figure 1: Environment and agent experimental set-up. [A] The agent (black circle) can move horizontally while objects (the grey disc) of varying diameters fall from above. The agent has an array of 7 distal sensors (black lines). [B] The network architecture has a layer of 7 sensory nodes connected to a layer of 9 inter-nodes, in turn connected to a layer of two motor nodes. The inter-node layer is fully interconnected, including self-connections.

Task. The task consists of two phases in each of which a circle-shaped object falls. The first circle presented is the ‘parent’, to be identified and recognised by its size. Without reinitialising the state of the agent, a second circle, or ‘test individual’ is then presented which may be the same or a different size (‘parent’ or ‘stranger’): the agent should centre in on a parent, but move away from a stranger. For each set of trials, the sizes of parents and strangers are drawn from a range of possible values. The task is loosely inspired on imprinting-like learning in that, for the case of geese, the first thing that it sees is its parent and it must remember it, with its particular features from a continuum of features, so that later, when presented again it can recognise it, or in the case of being a different individual from his parent, to avoid it (or get eaten!).

Artificial Evolution. All 183 parameters of the agent’s controller are evolved using a microbial genetic algorithm (Harvey, 2001). These are encoded in a genotype as

a vector of real numbers over the range $[0, 1]$. Offspring of microbial tournaments are generated as the product of uniform recombination between both members of the tournament, followed by mutation implemented as a random displacement with direction uniformly distributed on the N -dimensional hypersphere and with magnitude drawn from a Gaussian random variable with mean 0 and variance 0.01. The offspring replaces the loser of the tournament. The size of the population used was 50. Also, a minimal 1D wrap-around ‘trivial geography’ (Spector and Klein, 2005) has been introduced into this already minimal evolutionary technique with demes of size 10. Genotypes are mapped to network parameters using linear maps from $[0, 1]$ to $[-10, 10]$ for biases and connection weights and to $[1, 10]$ for the gain parameter while time-constants are exponentially mapped to $[e^0, e^4]$. Finally, because the fitness is noisy (described below), agents are re-evaluated every time they participate in a tournament.

Fitness Evaluation. Agents are evaluated by performing 100 fitness trials. Each trial consists of 11 runs. A run is the presentation of the parent individual followed by the test individual. Within each trial the parent’s radius is chosen at random uniformly between $[10, 20]$, and this parent is used for all 11 runs. In one of the runs the test individual is the same as the parent, in the other 10 runs it has different sizes chosen uniformly random in the same range. The fitness for one run is determined by the relative horizontal distance, d , between the agent and the falling object by the time the test individual reaches the vertical position of the agent, clipped to a maximum of 25 and then normalised between 0 and 1. In the case where the test individual is the same as the parent then the fitness is given by $f_a = 1 - d$ while in the rest of the cases $f_b = d$. Accordingly, the fitness of the trial is given by $f = (10f_a + \sum_1^{10} f_b)/10$. Finally, this is averaged over all 100 trials to give a fitness between 0 and 1. The activations of the nodes in the agent are not initialised between the parent and the test individual’s presentation, but they are re-initialised between different runs to $y_i = -\theta_i$.

Results

We will describe and analyse the results obtained in two sets of experiments. In the first one, the circles fall always from the same absolute position, which allows a solution whereby the agent exploits its position after the parental presentation; it makes use of its situatedness as an ‘external memory’. We analyse such a successfully evolved agent. In the second set of experiments the spatial information available to the agent is constrained by making the test individual fall from directly above the agent’s current position. This scenario is harder because it constrains the agent to use its internal dynamics to solve the learning problem.

Part I: Spatial Information Available

Here both the parent and the test individuals fall from the same absolute position. The main interest is in the explanation of how it does it. Based on this we then test its performance on a set of predictions.

What the agent does? Figure 2A depicts the performance of the best evolved agent out of 10 evolutionary runs with different initial random seeds. The figure shows the relative distance between the test individual and the agent by the time the test individual ends its fall. White represents ‘caught’ while black represents avoided. Each point represents the parent and test individual used for that run. This generalisation test comprises 40401 runs (all combinations of parent individuals and test individuals in the range [10, 20] by steps of 0.05). At this level of granularity, the agent avoids correctly 80.3% of the strangers while it centres correctly 95.0% on the parents. In the figure we can observe that most of the roughly 20% of test-individuals that were misclassified as parents are actually very close in size to the parent for that scenario; so this corresponds to very small differences not being differentiated. We can also observe a square-like region in the top right corner. That is a sign of discrete discrimination dynamic: anthropomorphically speaking, “if the radius of the circle is bigger than 18 then I don’t care for the details, it will get caught when I see another one bigger than 18 as well”. It is important to note that an agent that would not generalise well would have n non overlapping square-regions along the diagonal, corresponding to n modes of interaction. It is also important to note that during evolution it was common to observe during early generations agents with two modes of interaction denoted by this two big white squares.

How does it do it? In order to answer this question we first turn to the internal dynamics of the agent. Figure 2B shows the sigmoided activation of 6 out of the 9 nodes responsible for the internal dynamics (the other 3 nodes are not shown because they remain saturated on or off during runs). The figure depicts 9 different runs given by 3 different circle with radii: 10 (lightest grey), 15 (medium grey) and 20 (black). These cases are shown with circles on figure 2A.

The dashed vertical line in 2B (and 2C) represents the beginning of the second phase. A closer look at the activations around this time clearly shows that the agent apparently keeps no internal state that varies according to the size of the parent. But, then, how is it able to ‘remember’ this if not by any internal state? The answer lies in some relationship with the environment, via the spatial offset between the agent and the falling object. Figure 2C depicts the motion of the agent in time for the same set of different individuals. Depending on the size of the parent, the agent moves to a certain distance from the position where the test individual is going to fall. This allows it to behave appropriately for each test individual. That is, it uses its interaction with the falling object (i.e. horizontal offset) as an external memory.

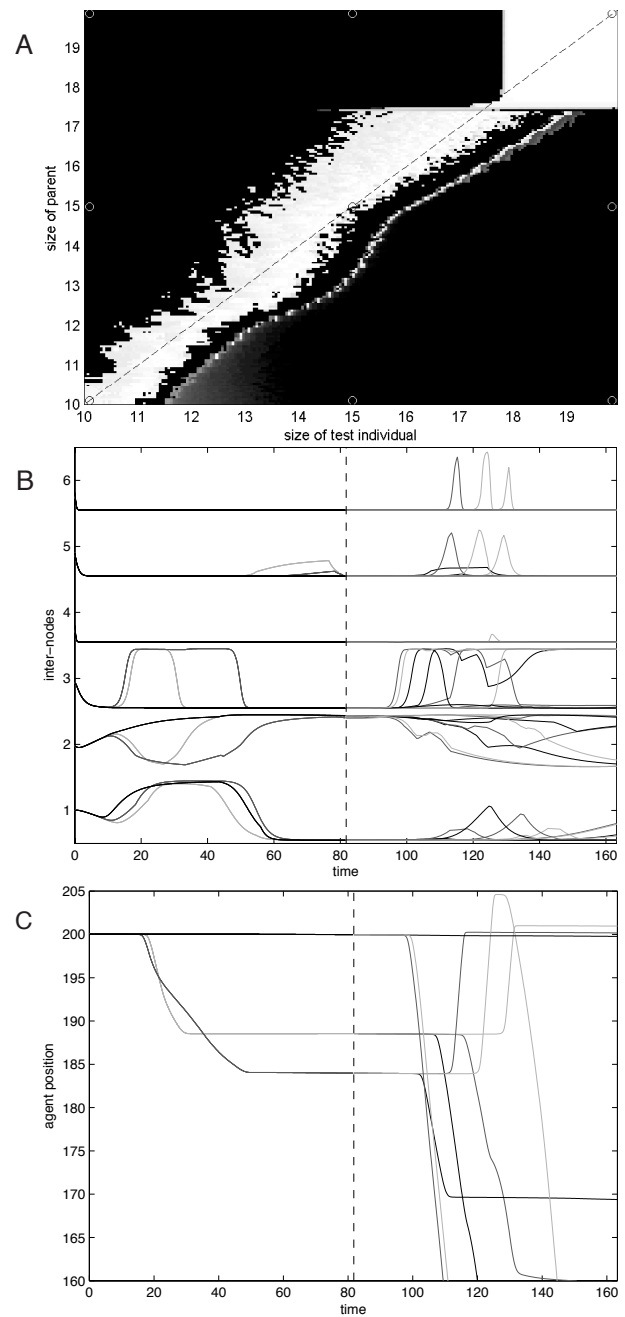


Figure 2: [A] Performance of best evolved agent. Shade of grey represents catching (white) or avoiding (black) the test individual. The vertical axis represents the radius of the parent, the horizontal axis represents the radii of the test individuals. The dashed diagonal line corresponds to those test individuals that are parents, strangers will lie off this diagonal. [B] Activation of inter-nodes over time. Vertical dashed line corresponds to the end of the parent presentation and the start of the test individual presentation. [C] Movement of agent in time.

What can we predict from this explanation? From this explanation given above, we can here make three predictions, and a further observation. Firstly, we should be able to set to some suitable common set of values the internal activations of the agents after presentation of the parents, whilst retaining their offset-positions, and expect the appropriate behaviours to remain. As a suitable set of values, for each node we took the average over all the runs of its activation at the change of phase. We then evaluated the agent’s performance using the same technique applied for Figure 2A, but resetting each node to its corresponding average value. As a result, the structure of the generalisation map was almost indistinguishable from the original (figure not shown). There was a small decrease to 91.86% observed in the catching of correct parents, while avoidance of strangers actually increased slightly to 81.29%. Hence despite internal state, at the change of phase, being now uniform across all the trials with differing parents, there was almost no change in the performance.

Secondly, memory behaviour is most useful when it is not too time-sensitive. We should be able to introduce any waiting period of time between the presentation of the parent and test individual and again obtain good discrimination performance. Using the same technique as before, we tested this prediction by measuring the performance when 10 extra units of simulated time were introduced here, and again for 100 extra units (the original task lasts almost 165). The performance, again, remained almost the same as the original: with 94.31% and 94.95% catching correct parents performance; and 80.36% and 79.25% avoiding stranger individuals for 10 and 100 units of time respectively.

Thirdly, behaviour of the agent at the start of the second phase should vary if we change its physical location then. To test this prediction, generalisation maps following the same technique as for figure 2A were generated for a new set of trials where changes were made in the position of the agent before the presentation of the test individual. A displacement of -1 was observed to lead to the catching of objects roughly 1 unit bigger in radius - which means that the diagonal line in the figure shifted to the right. Similarly, a displacement of +1 led to catching of objects smaller by roughly 1 unit of radius - a shift of the diagonal line to the left (figures not shown).

As a further observation, one interesting question that we can ask of this agent, of relevance to imprinting-like learning processes, is whether the learning is reversible, in other words whether the agent can learn and imprint on a new parent. With Lorenz’s birds, there was a sensitive period in infancy when they imprinted on a parent, and later changes in this imprinting were impossible. For the evolved agent in these experiments, there are different possibilities. If the agent is moved to directly under the dropping zone, it will imprint on whichever individual is dropped on it, regardless of any previous imprintings. This behaviour, again, arises

from the agent’s use of its interaction with the falling circle as opposed to an internal state.

Part II: Constrained to Use Internal Dynamics

The previous experiment emphasises the importance of the agent’s history of interaction with the falling object in the production of the learning behaviour. This points towards plasticity being a feature not only of the controller but of the interaction itself. Nevertheless, we are still interested in the question of plasticity in the agent’s internal dynamics. That is, whether we can evolve a dynamical node network, without added ‘learning’ rules, so as to be able to perform such imprinting-like behaviour (based on a feature which is on a continuum) by relying almost entirely on its internal dynamics. In order to answer this question we constrain the spatial information available to the agent by making the test individual fall directly on top of the agent, regardless of where the agent is at the end of the first phase.

Figure 3 depicts the performance of the best evolved agent out of 10 evolutionary runs with different initial random seeds using the same technique explained for figure 2A. This agent avoids correctly 84.1% of the stranger test individuals while it centres correctly on 82.6% on the parent-like test individuals. Although this evolved agent fails to recognise its parent when its radius is near 17, the agent’s generalisation for any smaller parent on the spectrum is actually better than the evolved agent from Part I. Similarly to the agent in Part I, the generalisation ability of this agent drops as the parent individual gets bigger, until it reaches a region (top right corner) where one simple mode of interaction sets in. Analysing *how* this agent performs this task is already a much harder endeavour and will not be tackled in this paper.

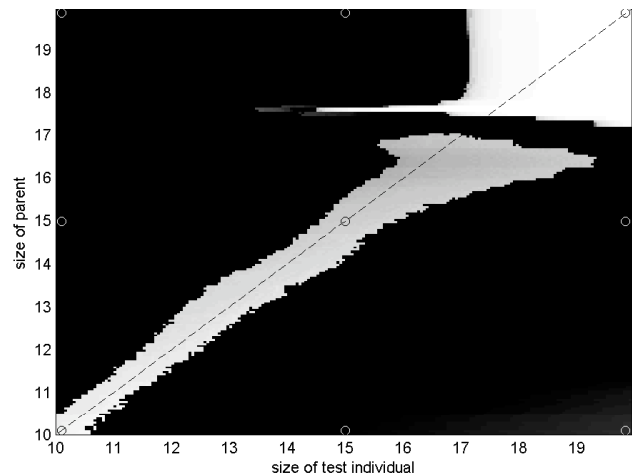


Figure 3: Performance of best evolved imprinting agent when spatial information is restricted.

Discussion

In artificial life, there has been research using evolution to generate learning behaviour within dynamical system controllers without pre-defining the learning mechanisms. However, all of this work involves learning to swap behaviours between, in practice, n different modes of interaction according to which scenario the agent finds itself in. We believe it is also important to look into learning behaviour where the features that are to be learned cannot be discretely determined and trained upon, i.e. learning behaviour where the agent must generalise. The task and evolved agents presented in this paper provide such an extension.

The results shown in the first scenario emphasise the role of the agent's situatedness in the generation of behaviour, and in particular of learning behaviour. It is important to point out that the evolved agents were not at any point selected for maximising the use of its body and interaction with the environment, as opposed to its internal state. This suggests that (a) when available the plasticity arising from the interaction will be exploited and (b) that this plasticity may be a simpler way to solve 'memory-related' problems than using internal state.

Initial experiments (not further reported here) using a fixed set of training examples as the fitness evaluation were not as successful as the approach used in the work presented herein, where we provided a random set of training examples during evolution. Although we present no empirical evidence, the intuitive reason could be that the latter encourages a more generalised behaviour as opposed to a set of fixed modes of interactions.

Finally, imprinting phenomena has been used to study developmental processes. We believe the model presented in this work has the potential to investigate issues relating to developmental robotics. A proportion of work in this area consists of adding an *ad hoc* morphogenetic-like mechanism to the existing control systems and testing it on a range of different tasks to see whether performance or evolvability improves. Our work suggests a completely different approach: choosing a task that seems to require a developmental process and then evolving dynamical systems that perform the task. The interest in this case will be in the analysis of the evolved networks.

Conclusions and Future Work

In summary, this paper has: (a) developed a situated and embodied imprinting-like learning task; (b) demonstrated (for the first time) the ability to evolve dynamical system controllers (without additional plastic mechanisms) that can modify their behaviour according to features on a continuum from the environment, even when the agent is constrained to use only its internal dynamics; (c) emphasised the importance of the agent's interaction with the environment in the generation of learning behaviour, as a source of plasticity or 'memory' on a continuum.

There are a number of possible extensions to the work presented herein. Mainly, work is underway in making certain aspects of the task more random. For example, making the circles fall in a range of different offsets from the agent, at different velocities and introducing random waiting periods between the presentation of the first and second objects. Also, one of the main interests of this line of research is to deepen the analysis of such learning agents using the tools of dynamical systems theory in order to better understand the evolved internal mechanisms as well as the agent's sensorimotor coordination with the environment.

Acknowledgements. We would like to thank Thomas Buehrmann and Ezequiel Di Paolo for their extremely useful feedback and insight.

References

- Beer, R. (1996). Toward the evolution of dynamical neural networks for minimally cognitive behavior. In *Proc. of the Fourth International Conference on Simulation of Adaptive Behavior*, pages 421–429. MIT Press.
- Harvey, I. (2001). Artificial evolution: a continuing SAGA. In Gomi, T., editor, *Evolutionary Robotics: From Intelligent Robots to Artificial Life*. Springer-Verlag LNCS 2217.
- Izquierdo-Torres, E. and Di Paolo, E. (2005). Is an embodied system ever purely reactive? In *Proc. of the 8th European Conference of Artificial Life*, pages 252–261. Springer Verlag.
- Izquierdo-Torres, E. and Harvey, I. (2005). Learning to discriminate between multiple possible environments: an imprinting scenario. Memory and Learning Mechanisms in Autonomous Robots Workshop (ECAL 2005).
- Lorenz, K. (1981). *The Foundations of Ethology*. Springer-Verlag.
- Phattanasri, P. (2002). *Associative Learning in Evolved Dynamical Neural Networks*. PhD thesis, Case Western Reserve University.
- Spector, L. and Klein, J. (2005). Trivial geography in genetic programming. In Yu, T., Riolo, R., and Worzel, B., editors, *Genetic Programming Theory and Practice III*, pages 109–124. Kluwer Academic Publishers.
- Tuci, E., Quinn, M., and Harvey, I. (2003). An evolutionary ecological approach to evolving learning behavior using a robot based model. *Adaptive Behavior*, 10(3/4):201–221.
- Yamauchi, B. and Beer, R. (1994). Sequential behavior and learning in evolved dynamical neural networks. *Adaptive Behavior*, 2(3):219–246.