An evolutionary ecological approach to the study of learning behaviour using a robot based model

Elio Tuci, Matt Quinn and Inman Harvey

Centre for Computational Neuroscience and Robotics, and

School of Cognitive and Computing Sciences

University of Sussex, Falmer - Brighton BN1 9QH, UK

phone: +44 (0)1273 872945 - fax +44 (0)1273 671 320

eliot, matthewq, inmanh@cogs.susx.ac.uk

Abstract

We are interested in the construction of ecological models of the evolution of learning behaviour using methodological tools developed in the field of evolutionary robotics. In this paper, we explore the applicability of *integrated* (i.e., non-modular) neural networks with fixed connection weights and simple "leaky-integrator" neurons as controllers for autonomous learning robots. In contrast to Yamauchi and Beer (1994a), we show that such a control system is capable of integrating reactive and learned behaviour without needing explicitly hand-designed modules, dedicated to particular behaviour, or an externally introduced reinforcement signal. In our model, evolutionary and ecological contingencies structure the controller and the behavioural responses of the robot. This allows us to concentrate on examining the conditions under which learning behaviour evolve.

Key Words: learning behaviour, dynamic neural networks, evolutionary robotics, genetic algorithms.

1 Introduction

Questions relating to how and why animals have come to evolve learning abilities are not always simple to answer. Both field and laboratory based studies *"...face formidable problems of disentangling the variables contributing to the behaviour which learning is assumed to be influencing"* (Plotkin, 1988). The classic cost/benefit modelling tools which biologists have at their disposal, generally require to make inevitable *a priori* assumptions which constraints the nature of those inquiries (see section 2.1).

Evolutionary robotics (ER), originally proposed an alternative methodology to automate the design of control systems for autonomous robots, has been recently employed to investigate or test hypotheses relating to the evolution of behaviour (see section 2.2). Due to its characteristics, ER complements the classic modelling tools like optimization models, Game theory, and population genetics models, allowing researchers to investigate phenomena which can be hardly investigated with the classic modelling tools. In contrast to a pure cost/benefit analytical approach, ER models of behaving agents allows investigation in which, by simply providing evolution with low-level "ingredients", it is possible to examine the selective pressures and the elements of the "ecology" of the model which are more likely to act in favour of the perceptuo-motor and cognitive structures that underpin the behaviour of interest.

In this paper, we apply ER methodology to the study of the *evolution* of *learning behaviour* from an *ecological* perspective. In particular, our interest is focused on Yamauchi and Beer's (1994a) model on the evolution of learning. This model, described in more details in section 3, represents one of the first attempts to develop an evolutionary ecological model of learning behaviour in artificial system. Yamauchi and Beer employed evolution to shape the low-level, dynamical properties of simulated agents' control systems in order to design the mechanisms capable of whichever combination of reactive and learning behaviour proved to be an effective solution to their learning task (Yamauchi and Beer, 1994a). However, their attempts to evolve an integrated (i.e., not modularised) control system was not successful. Despite the fact that learning was beneficial, the integrated controllers could not be shaped by evolution so as to underpin the agents' learning response. It seems thus that the factors which have prevented artificial evolution producing the desired result are likely to lie outside a cost/benefit analysis. However their nature has not been further investigated. In order to understand the nature of those factors, we decided to explore the problem by designing a similar task to the one employed by (Yamauchi and Beer, 1994a).

Our model, described in section 4, requires a robot to learn the relationship between the position of a light source and the location of its target (see section 4.1). The robot must be able to perform the task in an environment in which moving toward the light source will take it toward its target, but also in an environment in which this relationship is inverted. Only through interacting with its environment will the robot be able to learn and thus exploit the particular relationship between target and light. The robot is controlled by a single (i.e., not modularised) neural network, based on the same "low level" building blocks used by Yamauchi and Beer (1994a). Thus, the controller is provided with no explicit learning mechanism, such as automatic weight-changing algorithms (e.g., Floreano and Urzelai, 2001; Di

Paolo, 2000). Instead, evolution is left to shape the structure of the neural network that better ensures a correspondence between the agent's behavioural skills and the requirements of its environment.

We run two slightly different sets of simulations. In the first set of experiments (see section 5), agents are placed in an environment in which a non-learning agent has no selective benefit in paying attention to a cue, the light source. Consequently these agents did not evolve either the ability to distinguish the light source from other environmental stimuli, or the ability to learn. In the second set of experiments (see section 6), a modification to the previously employed evaluation function, creates a specific selective pressure to evolve the required mechanisms to make use of the light to navigate the arena. The modified evaluation function acts to make those perceptuo-motor abilities underlying learning, beneficial outside the learning context. Thus, they are more likely to evolve prior to and separately from the mechanisms required for learning.

The results of the second set of experiments (see section 6.1) demonstrate that it is possible to evolve an integrated neural network with fixed connection weights and "leaky integrator" neurons that successfully controls the behaviour of a Khepera mini-robot engaged in a simple learning task. The success of the second set of experiments can thus be attributed to the agents adapting to the specific requirements of an environment in which it was advantageous to evolve to pay attention to a potential learning cue before it had any value as a learning cue. Our results, discussed in section 7, demonstrate that our approach has been capable of illustrating and investigating conditions that influence, inhibit or facilitate the evolution of learning, but which are invisible to cost-benefit based analysis and modelling.

2 Background

In this section we look at the content of learning theory in biology, the ER, and then discuss some specific ER's models on the evolution of learning.

2.1 Learning Theory in Biology

During the first half of the twentieth century, research on learning was carried within the theoretical and methodological approach known as general process learning theory (Leahey, 1996). Although there were few dissenters (e.g., Tolman, 1949), during that time, learning theorists believed that some fairly small set of general principles, obtained regardless of any ecological consideration of the studied species, could explain all examples of learning (Hailman, 1985). These principles were tested by leaving the animals in an unnatural environment in which instinct was supposed to be of no use, and posing problems that only learning could solve. The conceptual justification for such practice was provided by firmly relying on the learning/instinct dichotomy and strongly believing in the absence of any qualitative differences between the learning abilities of different species.

During the Sixties and the Seventies a growing number of studies demonstrated biological predisposi-

tions and constraints on learning, which challenged the universality of learning principles across species. The general process view on learning came under attack from a series of studies which confuted the well-established laws and principles of the general learning theory (see Hinde and Stevenson-Hinde, 1973; Seligman, 1970; Bolles, 1985, for a review). Strong predispositions and constraints on learning, be it learning about kin phenotype, song, local environment features, or site, smell, and colour of flowers, have been identified in many different species (see Garcia and Koelling, 1966; Wilcoxon et al., 1971; Dukas, 1998; McNeely and Singer, 2001; Papaj and Prokopy, 1989; Craig, 1994). The specificity of the nature of learning (viz., evolved predisposition) led, at the beginning of the 80', at the development of the ecological approach to learning, which, although with few exceptions (see Bitterman, 2000; Macphail, 2001), has gathered a large consensus among learning theorist (see Johnston and Pietrewicz, 1985; Davey, 1989; Johnston and Turvey, 1980).

The ecological approach developed around the assumption that each instance of learning must be treated as a specialized capability shaped by selective pressures, and understandable only by reference to the ecology of the animal or its ancestor. Learning is seen as a process that assures a correspondence between an animal's behavioural skills and the requirements of its natural environment. Although the capacity to learn is ubiquitous among animals, the ecological approach asserts that learning has to be considered a behavioural response which fulfils species-specific and niche-specific biological functions. The ecological approach, thus, aims to construct a comprehensive and explicative network of more or less general statements on the nature of learning, by studying local principles of adaptation and the particularity of species-specific learning processes. By comparing local principles of adaptation, general principles of adaptation can be deduced by proving that different species of animals are faced with similar adaptational problems and appear to solve them in a similar way (see Johnston, 1981). However, the lack of "ecologically relevant" data and also methodological difficulties in collecting these data have been hindering the study of the adaptive significance of learning instances.

Despite the importance of studying learning as an adaptation, very little research has been done to study the adaptive significance of learning (Dukas, 1988). Although the relationship between evolution and learning has been the focus of the interest of naturalists since the eighteenth century, investigations have tended to focus more on the effects of learning on evolution (e.g., Baldwin, 1896; Morgan, 1896; Osborn, 1896; Waddington, 1942) rather than on the evolution of learning itself, as this latter kind of investigation has been hindered by the predominance of the general view of learning (see Johnston, 1996). Thus, as a consequence of the general paucity of ecologically motivated studies on the role that learning may play in species-typical development of behaviour under natural circumstances, more research needs to be done to produce any theories or to try to reveal anything about the selection pressures that may be involved in the evolution of learning (see Johnston, 1996; Papaj and Prokopy, 1989).

However, ecological investigations into the adaptive significance of learning seem to be facing several difficulties in gathering direct empirical evidence to support the claim that learning is an adaptation.

Plotkin (1988), in an essay about the relationship between evolution and learning, claimed that, *"although saying that learning is adaptive corresponds to the explication of the obvious, yet no published studies, have ever shown that learning in general, or some specific form of learning, increases reproductive competence"* (Plotkin, 1988). According to Plotkin this lack of evidence is mainly caused by the fact that both field and laboratory studies on learning seem to *"...face formidable problems of disentangling the variables contributing to the behaviour which learning is assumed to be influencing"* (Plotkin, 1988). It is a common practice in biology to simplify the analysis of a system by abstracting the trait(s) of interest from the adaptive whole that an organism is, to consider it/them in isolation. However, when the trait in question is learning, the ecological approach suggests that it is important not to lose sight of the numerous interactions that occur in evolution, or of the consequences of those interaction (see Johnston, 1996; Plotkin, 1988).

The costs of learning interact with each other, and the nature of these interactions are unlikely to be linearly additive. The evolution of different learning abilities will not involve the same selective costs and benefits, nor will the same costs and benefits always be involved to the same extent. Different stages of the life cycle will be exposed to different selection context (that is to say, learning will change), probably in complex ways, as the organisms develop. The evolution of one kind of learning ability may make the evolution of others less expensive. Moreover, the costs of learning not only interact with each other but also with other elements of the organism's overall adaptation to the environment, in a complex web of causal interactions among genetic, epigenetic factors, cognitive and perceptuo-motor capabilities all underlying the learning response (Johnston, 1996).

Johnston (1996) made an explicit call for theoretical models of the evolution of learning, wishing that these models might help to approach the study of the adaptive significance of learning by *"stimulating the search for relevant data"* (Johnston, 1996). Optimization models, Game theory, and population genetics models have been the classic modelling tools of biologists, which are all potentially valuable methodologies to investigate the evolution of learning in natural systems (see Harley and Maynard-Smith, 1983; Anderson, 1995; Papaj and Prokopy, 1989; Dukas, 1999; Shettelworth, 1984; Stephens, 1993; Arnold, 1978; Stephens and Clements, 1998). However, the classic cost/benefit modelling tools which biologists have at their disposal generally require making inevitable *a priori* assumptions (e.g., about the nature of learning mechanisms and perceptual capabilities of the modelled agents) which constrain the nature of those inquiries and overshadow the important effects of the interactions among the factors we mentioned in the previous paragraph.

In recent years, evolutionary simulation models have begun to establish themselves as a useful complementary tool (reviewed in Di Paolo et al., 2000). Using computer simulations of evolving populations of behaving agents allows for more complex and dynamic models to be constructed than those afforded by other methods. The individuals in these evolving populations may be extremely abstract, as in the Hinton and Nowlan's classic simulation model of the Baldwin Effect (Hinton and Nowlan, 1987). At the

other extreme, the individuals may be situated agents, engaged in non-trivial interaction with an environment (reviewed in Belew and Mitchcel, 1996; Nolfi and Floreano, 1999b). It is the latter type of model we are concerned with here, due to our interest in an ecological approach to the evolution of learning. With this type of model, rather than providing the agents with the "ability to learn" and investigating the conditions under which such trait is beneficial (as in cost/benefit models), we can provide evolution with low-level "ingredients" and examine evolutionary and ecological contingencies which are more likely to favour the appearance of perceptuo-motor and cognitive structures that underpin learning behaviour. Thus, as explained in the next section, ER might be a profitable way to explore the role of genetic, epigenetic, and ecological factors and the effects of their interactions, by complementing the modelling tool kit of biologists.

2.2 Evolutionary Robotics (ER) — a new tool for investigating learning

ER was initially proposed as a design methodology; a way to automate the design of control systems for autonomous robots, using algorithms based on Darwinian evolution (see Meyer et al., 1998; Nolfi and Floreano, 2000, for reviews of the field). It has typically focussed on the evolution of controllers for agents with a tight sensory-motor coupling with their environment. Rather than providing a suite of pre-designed behaviours for the robots to employ, the approach has been to work with low-level "building blocks", such as neural networks, allowing evolution to explore regions of the design space that conventional design approaches are often constrained to ignore (Harvey et al., 1992).

More recently, the ER methodology has been employed to investigate or test hypotheses relating to the evolution of behaviour in natural systems (e.g., Dale and Collett, 2001; Seth, 1998; Quinn, 2001; Nolfi and Floreano, 1999a). In this paper, we show that ER can be a suitable method to investigate hypothesis regarding the evolution of learning behaviour from an ecological perspective. Its methodology facilitates the construction of ecological models in which an agent is situated in, and becomes adapted to, the requirements of a specific environment. In addition, it allows us to work with low-level control structures that are shaped by evolution to create more complex and adaptive behaviour. This is particularly useful since we are interested in the evolution of learning behaviour. Rather than providing the robots with the ability to learn, we can provide evolution with low-level "ingredients" and examine the conditions under which mechanisms for learning behaviour evolve. Moreover, given the low-level building blocks, we may expect behaviours to evolve gradually, potentially allowing us to investigate transitional stages between fully reactive behaviour and effective learning behaviour[1]. That is, we may simulate the progressive adaptation of an agent to its environment, as it has been shown by Todd and Miller (1991), in one of the first computer model of the evolution of associative learning.

Todd and Miller (1991) investigated conditions under which simple artificial creatures are more likely

[1]For an example of a model simulating gradual behavioural transitions, see Quinn's ER model of the evolution of communication, in which initially ballistic behaviour gives way to collision avoidance and following behaviours, and finally to a simple signal and response behaviours (Quinn, 2001).

to evolve learning mechanisms to distinguish edible from poisonous items. In their simple scenario, food — which increases the creature's internal energy store — smells sweet; poison — which decreases the creature's energy — smells sour. There is however, some probability of sensory error, referred to as "smell sense accuracy". This means that some items which smell sour may be food and those which smell sweet may be poisonous. However, food and poison have fixed colour characteristics. Using colour as a cue, the agent can identify items with complete accuracy. The information associated with each colour varies between locations, food being red in some locations and green in others. Thus, to benefit from colour information, these creatures must learn the particular association between colour and food-type that holds in their particular location. Todd and Miller (1991) run several simulations varying the level of smell sense accuracy, to see which level of noise facilitates and which inhibits the evolution of learning. The results of the simulations showed that at low level smell accuracies, although learning would be very beneficial, the difficulty of gaining information from the environment makes learning harder, leading to increased time for learning to evolve. At high smell accuracies, in contrast, there is little adaptive pressure to evolve colour learning, although colour learning would be easy to perform, since agents can do well if they only rely on their smell sense. Middle smell accuracy represent a happy medium between phylogenetic adaptive pressure and ontogenetic ease of learning, leading to the rapid evolution of colour learning and its spread through the population.

In 1994, Yamauchi and Beer published two papers detailing experiments in which they used minimal models to investigate the evolution of agents capable of combining reactive, sequential and learning behaviour (Yamauchi and Beer, 1994a,b). Although only one of their papers explicitly looks at the evolution of simple form of associative learning (Yamauchi and Beer, 1994a), the methodology employed in both works is particularly interesting in the context of this paper. In Todd and Miller's (1991) model, previously described, learning is explicitly reduced to a particular learnable connection — by a Hebbian mechanism — between a colour sensor and the single motor neuron of a neural network. The development of this particular structure which distinguishes learning from non learning creatures, is genetically specified and controlled by a single gene. In contrast to Todd and Miller's (1991) model, Yamauchi and Beer sought to avoid imposing any *a priori* distinctions between reactive and non-reactive components of the agents' control systems. Rather than employing dedicated learning algorithms or mechanisms, they evolved continuous time recurrent neural networks (CTRNNs) to control their agents. A CTRNN is a simple dynamical model incorporating fixed-weight synapses and "leaky-integrator" neurons (see section 4.4 for more details). In this way, they employed evolution to shape the low-level, dynamical properties of agents' control systems in order to design the mechanisms capable of whichever combination of reactive or learning behaviour proved to be an effective solution to the task. In (Yamauchi and Beer, 1994b), successful networks were observed to have evolved very specific learning abilities; the networks evolved, "just enough plasticity to be able to accomplish the particular tasks".

For the purposes of this paper, we are particularly interested in the experiment in which Yamauchi and

Beer attempted to evolve a controller capable of performing a simple associative learning task (Yamauchi and Beer, 1994a). This was the least successful of the experiments detailed in their two papers, since they were unsuccessful in evolving a single, integrated network capable of performing the task. Despite the fact that learning is beneficial to the agents, they only succeeded in evolving agents capable of performing the task when they abandoned the "single integrated" approach and instead adopted a modular one. The factors which have prevented artificial evolution from producing the desired result are likely to lie outside a cost/benefit analysis. In order to understand the nature of those factors, we decided to explore the problem by designing a similar task to the one employed by (Yamauchi and Beer, 1994a). Since the model we use for our experiments is derived from theirs, we shall now proceed to introduce their experiment in more detail.

[Insert Figure 1]

3 Yamauchi and Beer's model of learning agents

In Yamauchi and Beer's (1994a) model an artificial agent can move in either directions along a one-dimensional continuum environment (see Figure 1). The environment contains a goal and a landmark. The agent has only proximal sensors, which means that it must be in close to the landmark or to the goal in order to sense them, and cannot perceive both simultaneously. The agent is evaluated over a number of trials. At the start of each trial, the agent is positioned in the centre, and the goal is positioned randomly at either the left or right end. The relationship between the position of the goal and the position of the landmark changed occasionally within the sequence of searches. There are two possible way in which the landmark can be located within the environment. In the landmark-near environments 1A and 1B, the landmark is located on the same side of the agent as the goal. In landmark-far environments 1C and 1D, the landmark is located on the opposite side of the agent from the goal. The agent's task is to learn, over a series of successive trials, whether the current environment is landmark-far or landmark-near, and then reach the goal. Each trial lasts until either the agent reaches the goal, the agent reaches the non-goal end of the continuum, or the time limit is exceeded. An agent will benefit if it can learn the relationship between a specific environmental cue and the type of environment into which it is placed.

As we noted above, Yamauchi and Beer's attempts to evolve an integrated (i.e., single network) control system to perform this task were unsuccessful. They were, however, successful in incrementally evolving a modular control architecture. They divided the agent's controller into three separate CTRNN modules, each of which was independently capable of controlling the agent. Each module was evolved separately to accomplish a different aspect of the task. The first module was successfully evolved to classify an agent's environment, discriminating between a landmark-near and landmark-far environment. The second module was evolved to direct the agent to the goal exclusively in landmark-near environments, and the third was evolved do the same in landmark-far environments. Once each module had been successfully

evolved they were combined to produce an agent capable of completing the task. In an agent's first trial, only the classifier module was used; its classification of the environment, signified by the final output of a designated neuron, determined which of the other two modules, the landmark-near module or the landmark-far module, would be used for the subsequent trials. Given the lack of success with their first attempt, it is worth noting just how easily they were able to produce a successful controller with the modular approach—the authors report that each module required only a few generations of artificial evolution. It is not difficult to see why the modular architecture should have proven easier to evolve than an integrated network. The original task had been decomposed into three independent sub-tasks, each of less complexity. In addition, part of the learning mechanism had been "hard-wired" into the modular architecture: The controller did not have to evolve the capacity to base its future behaviour on the outcome of the classifier module, instead its "memory" of the classification is provided as a hand-designed feature of the control architecture. Whilst it is clear why the modular approach should have proven *easier* to evolve, it is less clear why it should have proven *impossible* to evolve a single integrated network capable of performing the task.

In its first incarnation, Yamauchi and Beer's experiment is particularly relevant to the approach we are interested in pursuing. It presents an interesting example of an attempt to use evolution to shape low-level dynamical mechanisms in order to produce an integrated control system, adapted to a specific environment, and capable of a combination of reactive, non-reactive and learning behaviour. It was however, unsuccessful. The approach taken in the second incarnation of the experiment, despite its obvious success, would clearly be of far less value as an ecological model of the evolution of learning, due both to the imposed decomposition of the controller and the independent evolution of separate modules. We are therefore particularly interested in accounting their lack of success with their first approach. There seem to be at least two interesting possibilities. On the one hand, it might be that associative learning is simply too complex a form of behaviour to evolve using such low-level control systems. However, CTRNNs has been proved to be able to potentially simulate any dynamical system (see Funahashi and Nakamura, 1993). Thus, it seems to be more likely that associative learning is not too complex to evolve with the integrated approach, but that the nature of Yamauchi and Beer's model imposes constraints which prevent its evolution. This latter possibility potentially raises a further interesting question about the nature of these constraints. Given that controllers are potentially capable of learning, and given that learning has significant selective value (i.e., the benefits outweigh the costs), then why does learning fail to evolve?

4 Our model of learning robots

We set out to evolve an *integrated* (i.e., not modularised) CTRNN, with fixed synaptic weights and "leaky integrator" neurons, as a control system for a robot engaged in a task requiring associative learning. The task is derived from that used in Yamauchi and Beer's experiment, but implemented in a less minimal

fashion—we have replaced their one dimensional agent and environment with a simulation of a mobile robot located in a walled arena. As in Yamauchi and Beer's experiment, the task requires the robot exploit the relationship between the position of a landmark—in our experiment this is a light source—and the location of its goal. Unlike Yamauchi and Beer's model, however, our agent has no dedicated sensor input for a reinforcement signal and neither do we provide any explicit reinforcement signal when it successfully finds the goal. In keeping with our intention to adopt an ecological approach to our evolutionary model, we require that agents must evolve their own conditions of reinforcement, relying on existing sensors. The evolved control structures must support the robot's ability to act and perceive within the environment.

4.1 Description of the task

For our experiments, we used a simulated Khepera mini-robot. The robot's task requires it to navigate within a rectangular walled arena in order to find a goal, which is a black stripe on the white arena floor. The goal can only be perceived by the robot when it is directly above the stripe. A light source is located at one end of the arena. If the robot can learn the relationship between the location of the light source and the location of the goal it can use the light source to navigate directly to the goal. As made clear in section 4.3, the robot is equipped with 2 motor-driven wheels, 8 infrared sensors, (distributed around its body), 3 ambient light sensors (two facing diagonally forwards and one facing backwards), and a binary "floor sensor", facing downwards (an ambient light sensor, thresholded to distinguish between black and white areas of floor).

[Insert Figure 2]

As in Yamauchi and Beer's model, the robot is tested over a sets of trials, with the environment-type remaining constant within a set of trials, but varying between them. At the start of each trial, the robot is positioned in the left or right side of the arena (see the black dots in fig 2A). Once the robot has reached the central section of the arena (delimited in fig 2A by dashed lines), a goal— the black stripe—is randomly positioned close to one end of the arena. Thus, the positioning of the goal is independent of the robot's current orientation. At the same time, a light source at end of the arena is switched on. Depending on the type of environment, the light source will either appear at the same end of the arena as the goal (landmark-near environment, illustrated by figures 2B and 2C), or at the opposite end (landmark-far environment, illustrated by figures 2D and 2E). The robot can perceive the position of the light-source from anywhere within the arena. Thus, if the robot can learn the current relationship between the landmark and the goal, the robot can use the position of the landmark to guide it towards the goal. If the robot does not exploit the relationship between the landmark and the goal, it can do no better than choosing a direction at random.

Whilst performing the task, the robot's behaviour will be evaluated according to an evaluation function that rewards or penalises certain actions. Note, that this evaluation determines the fitness of the agent. The fitness is subsequently used to determines how many offspring the agent has within the evolutionary

algorithm. The fitness score is not made available to the robot during its "lifetime"; it does not serve as a reinforcement. The robot increases its fitness by moving from its initial position to the central section of the arena, thereafter by progressing toward the goal. It receives an additional score if it finds the goal and stays over it. The exact details of the evaluation function are set out in section 4.2 below.

During evolution, each robot undergoes two sets of trials in the landmark-near and two sets in the landmark-far environment. After each set of trials, the robot's control system is reset, so that there is no internal state or "memory" remaining from the previous environment. Each individual trial within these test sessions is subject to time constraints. The robot has 18 seconds to reach the middle of the arena. Then the landmark and the goal are positioned, and the robot has another 18 seconds to find the goal. Each trial can be terminated earlier either because the first time limit to reach the middle of the arena is exceeded; or because the agent reaches and remains on the goal for 10.0 seconds; or because the robot crashes into the arena wall. At the beginning of the first trial, and for every trial following an unsuccessful one the robot is positioned on either the left or the right part of a rectangular arena close to the side wall. For every trial following a success, the robot normally starts from the position and orientation with which it ended the previous trial, but there is a small probability that it is randomly repositioned. The simulation is deliberately noisy, with noise added to sensors and motors; this is also extended to the environment dimensions. Every time the robot is replaced, the width and length of the arena, and the width of the central area that triggers the appearance of the landmark and the goal, are redefined randomly within certain limits. The robot undergoes 15 trials for each test session. During this time the relationship between landmark and the goal is kept fixed; the relationship remains unknown to the robot unless and until it can "discover" it through experience. The position of the goal within the arena (i.e., left or right), is randomly determined every single trial and the landmark is subsequently appropriately positioned (depending on whether it is currently landmark-near or landmark-far). Over a full set of trials, all the 4 landmark/goal combinations (fig 2B,fig 2C,fig 2D, fig 2E) occur with equal likelihood.

4.2 Evaluation function

The evaluation function determines the fitness score for each robot. It has been designed in consideration of the selection pressures that might facilitate the evolution of the learning behaviour. The robot increases its fitness by moving from its initial position to the central section of the arena, thereafter by progressing toward the goal. It receives an additional score if it finds the goal and stays over it. The evaluation function penalises the robot for colliding with the arena walls, for failing to reach the central section of the arena, and for moving into the non-goal end of the arena after it has reached the centre.

The total fitness score (ϕ) attributed to each individual was simply the average score that it achieved for each trial t of each test session s, after the first trial of each session had been excluded. Performance in the first trial was ignored on the grounds that a robot cannot "know" what kind of environment it is

situated within (i.e., landmark-near or landmark-far environment), and thus its choice of direction can only be random.

$$\phi = \frac{1}{56} \sum_{s=1}^{4} \sum_{t=2}^{15} F_{st} \qquad \text{where} \qquad F_{st} = C \left(A + B \right)$$

- $A = 3.0 \frac{(d_f - d_n)}{d_f}$ This is the main component of the evaluation function. It contributes to increase the robot's fitness score by evaluating how far it went in moving toward the goal. In details, d_f represents the furthest distance that the robot reaches from the goal after the light is on. At the time when the light goes on, d_f is fixed as the distance between the centre of the robot body and the nearest point of the goal. After this, d_f is updated every time step if the new d_f is bigger than the previous one. d_n represents the nearest distance that the robot reaches from the goal after the light is on. At the time when the light goes on, d_n is fixed as equal to d_f, and it is subsequently updated every time step both when the robot gets closer to the goal and when the robot goes away from the goal. d_n is also updated every time d_f is updated. In this case d_n is set up equal to the new d_f. Notice that A does not discriminate between robots that reach the goal successfully and robots that reach the goal by exploring both sides of the arena. Thus, the maximum contribution that A gives to the robot's fitness score is equal to 3 every time the robot reaches the goal, no matter how.

- $B = \frac{p}{p_{max}}$ This component increases the robot's fitness score by evaluating the time spent on the goal. In details, p represents the longest unbroken period of time spent on the goal, and p_{max}, the target length of time for the robot to remain on the goal (i.e., 10 seconds). B is a scalar value in the range [0:1], a function of p.

- $C = abc$ This component decreases the robot's fitness score if it incurs any of the following penalties: a, b or c. The value of each defaults to 1, however a is set to 0 if the robot fails to reach the centre of the arena; b is set to $\frac{1}{3}$ if the robot leaves the central area in the wrong direction (i.e., moves away from the goal); c is set to $\frac{1}{5}$ if the robot crashes into the arena wall. These values were used in initial experiments and proved to work; however there is no reason to believe that other choices may not also work as well or even better.

[Insert Figure 3]

4.3 Simulated robot

We used a minimal, 2-dimensional model of a Khepera robot and of its interactions with its environment. The implementation of our simulation closely follows Jakobi's minimal simulation of a Khepera operating in a corridor, both with respect to the way in which the robot's movement is modelled, and in the way in which the infrared and ambient light sensors are updated (see Jakobi, 1997, for a detailed description of this implementation). During the simulation, robot sensor values are extrapolated from a look-up

table. Noise was applied to each sensor reading. The robot has two motor-driven wheels, which are independently driven, and can move in forward or in reverse. The robot can make use of all of it's 8 infra red sensors (Ir_0 *to* Ir_7 in figure 3), these return a reading which is an inverse, non-linear function of the proximity of objects in the environment (i.e., the arena walls); they have a range of around 7 cm, the robot itself is 5.6 cm in diameter. We constrained the robot to use three of its ambient light sensors, two of these are positioned frontally at 45 degrees to the left and right, the third faces backwards (A_1, A_4, A_6 in figure 3). The light source is modelled as a bar that illuminates the whole arena with equal luminosity. Each ambient light sensor faces out from a point on the exterior of the robot and detects any light within the range ± 30 degrees from orientation of the sensor. Finally, the robot has a "floor sensor", this can be conceived of as an ambient light sensor, positioned facing downwards on the underside of the robot; it is thresholded so that it produces a binary output, 0 when the robot is positioned over white floor and 1 when it is over black floor. The simulation was updated the equivalent of 5 times a second.

4.4 The Network

The robot controller was a fully connected, 13 neuron CTRNN. Each neuron, indexed by i, is governed by the following state equation:

$$\frac{dy_i}{dt} = \frac{1}{\tau_i}\Big(- y_i + \sum_{j=1}^{13} \omega_{ji} z_j + gI_i \Big) \qquad \text{where, } z_j = \frac{1}{1 + exp\,(y_j + \beta_j)}$$

Here, by analogy with real neurons, y_i is the cell potential, τ_i the decay constant, β_j the bias term, z_j the firing rate, ω_{ji} is the strength of synaptic connections from the j^{th} neuron to the i^{th} neuron, I_i the intensity of the sensory perturbation on sensory neuron i. The neurons either receive direct sensor input or are used to set motor output. There are no interneurons. All but two of the neurons receive direct input from the robot sensors (for the remaining two, $gI_i = 0$). Each input neuron is associated with a single sensor, receiving a real value (in the range $[0.0 : 1.0]$), which is a simple linear scaling of the reading taken from its associated sensor[2]. The two remaining neurons are used to control the motors. Their cell potential y_i, is mapped onto the range $[0.0 : 1.0]$ by a sigmoid function, and then linearly scaled into $[-10.0 : 10.0]$, set the robot's wheel speeds. The strengths of the synaptic connections, the decay constants, bias terms and gain factors were all genetically encoded parameters. Cell potentials were initialised to 0 each time a network was initialised or reset. State equations were integrated using the forward Euler method with an integration step-size of 0.2.

4.5 The Genetic Algorithm

A simple generational genetic algorithm (GA) was employed (Goldberg, 1989). A population contained 100 genotypes. Each genotype was a vector comprising 196 real values (169 connections, 13 decay

[2] Specifically, neuron N_1 takes input from infra red sensor Ir_0, N_2 from Ir_1, N_3 from Ir_2, N_4 from Ir_3, N_5 from Ir_4, N_6 from Ir_5, N_7 takes the mean input of the two rear IR sensors, i.e., $\frac{Ir_6 + Ir_7}{2}$; N_8 takes input from ambient light sensor A_1, N_9 from A_4, N_{10} from A_6, and N_{11} takes input from floor sensor F

constants, 13 bias terms, and a gain factor). Initially, a random population of vectors was generated by initialising each component of each genotype to values chosen at uniform random from the range [0:1]. Subsequent generations were produced by a combination of selection with elitism, recombination and mutation. In each new generation, the two highest scoring individuals ("the elite") from the previous generation were retained unchanged. The remainder of the new population was generated by fitness-proportional selection from the 70 best individuals of the old population. New genotypes, except "the elite", were produced by applying recombination with a probability of 30% and mutation operator. Mutation entails that a random Gaussian offset is applied to each real-valued vector component encoded in the genotype, with a probability of 0.1. The mean of the Gaussian was 0, and its s.d was 0.1. During evolution, all vector component values were constrained to remain within the range [0:1].

Genotype parameters were linearly mapped to produce CTRNN parameters with the following ranges:

- biases $\beta_j \in [-2, 2]$;
- connection weights $\omega_{ji} \in [-4, 4]$;
- gain factor $g \in [1, 7]$.

Decay constants were firstly linearly mapped onto the range $\tau_i \in [-0.7, 1.3]$ and then exponentially mapped into $\tau_i \in [10^{-0.7}, 10^{1.3}]$

5 An Initial Experiment

Prior to summarising the results of our first attempt to evolve robots capable of associative learning, it is worth summarising the similarities and differences between our model and that used by Yamauchi and Beer (1994a). Like Yamauchi and Beer we present our robot with a task which requires it to learn, over a series of trials, the nature of the environment in which it has been placed. To be successful, the robot should evolve the ability to disambiguate between environments (i.e., landmark-near or landmark-far), and to predicate its behaviour on its previous interactions with its environment. As in their model, there is a 100% correlation between the location of the landmark and the goal within one test session, and the robot is exposed equally to both types of environment. As in Yamauchi and Beer's (1994a) first experiment, we are attempting to evolve a single, integrated CTRNN, incorporating no explicit learning modules or mechanisms.

There are also a number of differences between the two models. With possible exception of the lack of an explicit reinforcement signal, the differences in our model would seem to make the task easier to evolve. Firstly, in our model, the robots can rely on richer sensory-motor capabilities, which might help facilitate the job of the controller. Secondly, it is likely to be an advantage that the landmark is visible from all over the arena — recall that in Yamauchi and Beer's model the agent could not perceive the landmark and the goal at the same time. Thirdly, our trials are not terminated if the robot goes to the wrong end of the arena. Although the robot is penalised for this behaviour (i.e., this behaviour reduces

its fitness), it may still recover from its mistake and find the goal. Moreover, it may be able to learn relying on an evolved strategy which required to make and recover from one or more successive mistakes. Finally, our evaluation function has been designed to be incremental and, insofar as is possible, to provide a fitness gradient between "good" and "bad" behaviour. This contrasts strongly with the all-or-nothing evaluation function employed by Yamauchi and Beer (1994a), which discriminated only between agents which found the goal, and those which did not. Bearing these differences in mind, we now present the results of initial experiments.

[Insert Figure 4]

5.1 Results

Twenty evolutionary simulations, each using a different random seed, were run for 5000 generations. We examined the best (i.e., highest scoring) individual of the final generation of each of these simulation runs in order to establish whether it had evolved an appropriate learning strategy. In order to test the learning ability of the evolved robots, each of these final generation controllers was subjected to 500 test sessions in each of the two types of environment (i.e., landmark-near and landmark-far). As before, each test session comprised 15 consecutive trials, and controllers were reset before the start of each new session. During these evaluations we recorded the number of times the controller successfully navigated to the target (i.e., found the target directly, without first going to the wrong end of the arena). Note that a controller which does not learn, should be expected to average a 50% success rate. Controllers which consistently produce success rates high than this can only do so by learning. The results for the best controller of each of the twenty runs can be seen in figure 4, which shows the percentage of successes in each trial during a session under each environmental condition, averaged over 500 test sessions. None of these achieve a success rate significantly above random. Like Yamauchi and Beer, we were unable to evolve controllers to perform this associative learning task. This is in spite of the differences between our model and theirs, which might have made the successful behaviour easier to evolve.

6 A Hypothesis and a Second Experiment

How are we to explain the failure of our evolutionary simulation to produce the robots capable of associative learning? In each of the twenty runs, the robots failed to evolve to exploit the light source as a predictor of the location of the goal. Indeed, in observing the evolved controllers, it was evident that not only did they not evolve to use the light-source as cue, they simply failed to evolve any systematic response to the light-source at all. The light-source is effectively ignored. Indeed this can be noticed also looking at figure 4. If a robot either constantly followed or moved away from the light, its performance in respectively landmark-near or landmark-far environment would be significantly above the 50% success rate, which is what is expected by a robot that move randomly without exploiting the light. However, as

shown in figure 4, none of the robots did significantly better than 50% of successes in either landmark-near or landmark-far environment.

This observation seems to give a clue as to why the robots may have failed to evolve the ability to learn. It seems clear that unless the robot evolves to pay attention to the light[3], it will be unable to observe or exploit the relationship between the light and the goal. This leads us to a far more specific question: why does the robot fail to evolve to pay attention to the light source? If we can answer this question, it seems that we will be a long way to understanding why associative learning failed to evolve in our model.

Let us consider the adaptive value of paying attention to the light. Why is the light useful to the robot? The light is useful if the robot can learn the association between the light position and the goal position. That entails not only observing the current relationship between light and goal, but subsequently exploiting it, by modifying future behaviour based on this observed relationship. In short, paying attention to the light source is useful to a robot capable of learning. For a non-learning robot, however, the light presents no useful information because, on any given trial, there is no predictable relationship between the light and the goal. This seems to present a significant problem. Unless the robot has *already* evolved to pay some attention to the light, it will be very difficult for it to evolve to learn from the relative relationship of the light position and the goal position. However, unless the robot has already evolved the ability to learn, the light has no adaptive significance.

This problem is not insurmountable, it can be addressed by giving the light source some adaptive significance other than as a learning cue. One way of doing this is to bias the proportion of each type of environment that the robots experience. For example, if robots are presented with landmark-near environment more often than the landmark-far environment, then the goal will be close to the light more often than it is far from it. In this way, the light can serve as a cue which can be exploited by even a purely reactive robot, since a robot which moves toward the light will be correct more often than it is wrong. Fully successful behaviour will nevertheless still require that the robot can learn to distinguish each type of environment and act appropriately. In our second experiment, which will be described shortly, we adopted a slightly different, but effectively equivalent strategy. Rather than bias the proportion of times each environment is presented, we kept the proportion equal, but biased the relative weighting of the scores achieved in each environment, thereby achieving the same effect.

The following is the modified evaluation function with the introduction of a new component κ, which simply is an integer set to 3 in landmark-near environment, and to 1 in landmark-far environment. The values of (κ) have been systematically studied, and have been chosen to maximize the frequency of successful run out of a total of 20 randomly seeded simulations. All the other components of the evaluation function have been remained unchanged.

[3] "To pay attention to the light" could be seen as an anthropomorphism. But here we use this to mean no more than: a robot that does not pay attention to the light is simple a robot in which the states of its light sensors are irrelevant in determining its motor output.

$$F_{st} = \kappa C \left(A + B \right)$$

The total fitness score (ϕ) attributed to each individual was again simply the average score that it achieved for each trial t of each test session s, after the first trial of each session had been excluded.

It should be noted that our modified approach is not simply some an *ad hoc* solution to the problem we have identified. To explain this point we wish to describe a relatively simple example of associative learning found in nature, which refers to the the timing of egg laying by great tits in Switzerland (Nager and van Noordwijk, 1995). Breeding success of these great tits is considerably higher if the nestling stage coincides with peak caterpillar abundance. However, females must lay eggs a few weeks earlier than this if they are to capitalise on caterpillar abundance. This requires using some cue available earlier in the spring to predict when peak caterpillar abundance will occur. Great tits use temperature in early spring as a cue for determining when to lay eggs. By itself, however, this cue is only a weak predictor: there are significant differences between the date of peak caterpillar abundance in different localities. Female great tits apparently record the peak date of caterpillar abundance, as it relates to the nestling feeding period, and then adjust the time of egg laying in the same locality during subsequent years to increase synchronisation. In other words, learning allows the birds to strengthen the association between a specific cue (early spring temperature) and an environmental event that affects fitness (peak caterpillar abundance in a specific location). The tits benefit from learning the local correlation between early spring temperature and caterpillar abundance. Nevertheless, this cue provides a useful, albeit significantly less accurate, predictor of caterpillar abundance even in the absence of learning. That is, the cue would provide useful environmental information even to non-learning great-tit. Indeed, it seems reasonable to suggest that great tits would have evolved non-learning strategies to exploit this temperature cue before evolving learning strategies by which to improve its predictive value within a local environment. This separated evolution of the relevant cues upon which the learning response is built, is exactly what we aim to get with the introduction of the parameter (κ).

[Insert Figure 5]

6.1 Results

With the modified evaluation function, we run again for 5000 generations, twenty evolutionary simulations, each using a different random seed. Then for the best evolved individuals of the final generation of each of these simulation runs, we repeated the evaluation test described in 5.1. The results for the best controller of each of the twenty runs can be seen in figure 5, which shows the percentage of successes per trial during a session under each environmental condition.

It is clear that in 13 of the runs (i.e., run n. 2, 3, 4, 7, 8, 9, 10, 13, 14, 15, 16, 17 and 19) robots quickly come to successfully use the light to navigate toward the goal, irrespective of whether they are

in an environment which requires moving towards or away from the light source. All these controllers employ a default strategy of moving toward the light, and hence are always successful in the landmark-near environment (see figure 5 dashed lines). Consequently they are initially very unsuccessful in the landmark-far environment (see figure 5 continuous lines). Nevertheless, as can be seen from figure 5, when these controllers are placed in the landmark-far environment they are capable of adapting their behaviour over the course of very few trials and subsequently come to behave very differently with respect to the light. Each of these controllers, when placed in the landmark-far environment, initially navigates toward the light, fails to encounter the target and subsequently reaches the wrong end of the arena. At this point each of them turns around and proceeds back toward the correct end of the arena where they ultimately encounter the target stripe. The fact that in subsequent trials each of these robots move away from the light rather than towards it, demonstrates that these robots have learnt from some aspect, or aspects, of thier earlier experience of the current environment. More prosaically, they learn from their mistakes.

It should be noted that the controllers from the runs (i.e., 5, 12 and 20), whilst performing the task with varying degrees of success, all modify (viz., improve) their behaviour over the course of the initial trials. For example, the controller from run n. 20 initially has a 80% success rate in landmark-near environment and a 0% success rate in landmark-far environment. In the landmark-near environment its success rate climbs from around 80% to 100%. In the landmark-far environment its success rate climbs from 0% to around 80% within the course of a few trials. Similarly, with the controller from run n. 5 and n. 12, success in a landmark-far environment increases rapidly from 0% up to around 60% whilst success in the landmark-near remains consistently around 100%. Only in 3 runs out of 20 (i.e., 1, 6, 18) the evolution fails completely to properly structure the robot control systems. These three robots only evolved a strong disposition to follow the light unrespectful of the current environmental condition. Therefore, their behaviour results to be quite successful in landmark-near environment, very successful for run n. 6, but for all of them it ended in a completely failure within landmark-far environment.

[Insert Figure 6]

6.1.1 An analysis of the evolutionary transitions

In a further series of analyses, we repeated the evaluation test described in 5.1 for the best control system of each generation of one of our successful evolutionary runs (i.e., run n. 10). This analysis looks at how the performances of these control systems changed generation after generation, with respect to the percentage of successes, percentage of error, and percentage of detour behaviour recorded respectively in landmark-near and in landmark-far environments. Recall that the robot is successful every time it finds the target directly, without first going to the wrong end of the arena. The robot makes an error every time it employs a strategy which differs from the previously described: (a) if the robot does not find the target at all within the time limits, or (b) if it finds the target having previously explored the wrong end

of the arena. Detour behaviour refers to (b).

The results of this evolutionary analysis, shown in Figure 6, seem to suggest that the evolution of effective learning robots, starting from robots which behave according to a randomly initialised control system, is marked by 3 well distinguishable evolutionary stages.

Between generation 250 and generation 1250, the best evolved robot has the characteristics of a robot that always performs phototaxis regardless of the characteristics of the environment. Thus, this phylogenetically primitive robot is very successful in landmark near environment (see graph 6A); but its strategy is a complete failure within landmark-far environment (see graph 6C). An analysis of its typical behaviour shows that it set out from the central section of the arena moving systematically towards the light. It has a strong tendency to approach the light regardless of the position of the light with respect to the target. Despite also experiencing the landmark-far environment, it does not change its light seeking attitude. In landmark-far, having reached the wrong end of the arena, the robot remains until the end of the trial in the proximity of the wall close to the light (either the west or the east arena's wall). It does not perform detour behaviour in the landmark-far environment (see graph 6D).

Between generation 1250 and generation 3000, the best evolved robot has the characteristics of a robot whose behaviour appears to be more flexible and to some extent sensitive to the environmental conditions. The robot at this stage of evolution is still distinguishable by a tendency to approach the light, which makes it very successful in landmark-near environment (see graph 6A). However, an analysis of the behaviour of this robot shows that a small amount of experience in landmark-far environment can partially change its attitude towards the light. Every time it senses the light with the back ambient light sensor, it ceases light-approaching behaviour and keeps on moving straight forward until it finds the target. This plasticity accounts for its successes in the landmark-far environment (see graph 6C). It has also to be noticed that, almost all the time the robot fails to find the target properly in the landmark-far environment, it seems to be able to recover from its mistakes. This means that if approaching the light did not lead the robot to the target, it subsequently turns and explores the opposite end of the arena (see graph 6D percentage of detour behaviour).

Generation 3000 marks the evolution of an effectively learning robot, which proved to be successful both in landmark-near (see graph 6A) and in landmark-far environment (see graph 6C).

7 Discussion

We have suggested that ER methodology may be usefully applied to the construction of ecological models of the evolution of learning. We believe that one of the main potential benefits of employing an ER approach lies in the possibility of allowing evolution to shape low-level control structures to create more complex, higher-level adaptive behaviour; rather than providing the agents with a built-in ability to learn, we can provide low-level "building blocks" from which higher-level learning mechanisms can be constructed. In this context, we were particularly interested in Yamauchi and Beer's failure to evolve

an integrated low-level control system (a single CTRNN), capable of associative learning (Yamauchi and Beer, 1994a). Given the similarities between their approach and the one which we advocate, accounting for their lack of success was of particular interest to us, and has been one of the main aims of this paper. In particular Yamauchi and Beer's failure raised the possibility that associative learning might simply be too complex a behaviour to evolve using such low-level building blocks; a possibility that it was clearly important for us to investigate. Thus, one important result of this paper is simply that we have discounted this possibility. The second set of experiments reported in this paper shows that we can use artificial evolution to combine low-level building blocks, specifically fixed synaptic weights and leaky-integrator neurons, to produce controllers capable of associative learning.

In addition to demonstrating that it is possible to use low-level control structures to evolve learning behaviour, this paper has also begun to demonstrate the potential usefulness of such an approach. We have presented two experimental conditions, both variations on the same, relatively simple, scenario. Under both conditions there existed a clear selective advantage to learning: Agents capable of learning the correlation between the position of the landmark and goal would be fitter (i.e., produce more offspring) than those which could not. In other words, from the perspective of a cost-benefit analysis, there is no difference between the two conditions. However, it is clear that the two conditions yielded significantly different outcomes. Under the first experimental condition, none of the 20 evolutionary runs produced agents capable of learning the relationship between the landmark and the target. In contrast, under the second condition agents evolved the ability to learn (to varying degrees) in all but three of the runs. Thus, it is clear that there was a significant difference between the two experimental conditions. This result demonstrates that the approach which we have employed is capable of illustrating and investigating conditions that influence, inhibit or facilitate the evolution of learning, but which are invisible to cost-benefit based analysis and modelling. It thus adds strength to our claim that ER models have the potential to complement existing analytic and modelling approaches.

How can the difference between the two conditions be explained? Ultimately, the explanation is an ecological one, that is, it requires reference to differences between the agents' environments under the two conditions. In both conditions, as we have noted above, agents would have benefited from being able to associate a cue stimulus (i.e., the light source) with an environmental state (i.e., the relative position of the target). One of the basic proximate requirements for an organism to be capable of associative learning is an ability to distinguish — or distinguish between — the relevant sensory stimuli (Dukas, 1998). This is simply to say that if an organism is not able to recognise a cue and distinguish it from other environmental stimuli, it will be unable to form an association between that cue and an environmental state. In the first set of experiments, agents were placed in an environment in which a non-learning agent had no reason to evolve any response to an important potential cue - the light source. Consequently these agents did not evolve the ability to distinguish the light source from other environmental stimuli. It is thus unsurprising that these agents failed to evolve the ability to learn and exploit the relationship between the light source

and the goal, given that they had no reason to pay attention to the light source in the first place. In the second set of experiments, the agents were placed in an environment in which it was advantageous for non-learning robots to evolve to pay attention to the light and its location; even an agent capable of a reactive strategy of phototaxis would be fitter than one which ignored the light source. Importantly however, even this simple reactive response to the light required that the agents could distinguish the light cue from other environmental stimuli. The success of the second set of experiments can thus be attributed to the agents adapting to the specific requirements of an environment in which it was advantageous to evolve to pay attention to a potential learning cue before it had any value as a learning cue.

The evolution of associative learning in the second set of experiments can be better understood if we consider the consequences that the second environment had for the evolution of the agents patterns of sensory-motor coordination. As was described in the analysis in section 6.1.1, agents in the second experiments first evolved the sensory-motor coordination necessary to approach the light source, irrespective of whether the environmental state was landmark-near or landmark-far. The agents subsequently evolved the ability to "recover" from mistakes, that is, if approaching the light did not lead the agent to the goal, it would subsequently turn and explore the opposite end of the arena. This was a non-reactive strategy, since once the agent had "recognised" its mistake, it altered its subsequent behaviour with respect to the light source. Note however, that this non-reactive behaviour required that the agents had already evolved to pay attention to the location light source. By this stage then, in most important respects, the agents were able to classify their current environment into landmark-near and landmark-far. From here it is not a significantly large step to the evolution of the capacity for associative learning (primarily it requires that an agent's memory of a "mistake" to be extended so that it comes to have an effect on subsequent trials). It should be clear that the differences between the two experimental conditions are not just to be explained in terms of differences between the agents' environments, but also in terms of the consequences that these environmental differences had for the agents' evolution of patterns of sensory-motor coordination. It should be emphasised that it was the use of low-level control structures that allowed for the gradual and progressive evolution of the higher-level behaviours, with learning behaviour building upon non-reactive behaviour (learning from mistakes) which in turn had built upon reactive behaviour (attending to the light source). It is extremely difficult to see how any high-level modelling technique could have offered an explanation of this type.

8 Conclusions

Yamauchi and Beer (1994a) proposed a model which certainly represents on of the first attempts to develop an evolutionary ecological model of learning behaviour in artificial systems. However, despite the fact that in their model learning is beneficial, their attempts to evolve an integrated (i.e., not modularised) control system for learning agents was not successful. The factors which have prevented Yamauchi and Beer's (1994a) model to produce the desired result were likely to lie outside a cost/benefit analysis (see

section 3). In order to understand the nature of those factors, we decided to explore the problem by designing a similar task to the one employed by Yamauchi and Beer (1994a).

By using an ER model, we isolate selective pressures and ecological contingencies which are more likely than others to favour the evolution of the desired "learning machinery". The results of our research support the claim that ER models have the potential to complement existing analytic and modelling approaches. Different other "low-level" building block are currently investigated as basic components for plastic neural structures (Floreano and Urzelai, 2001; Di Paolo, 2000). In isolation, our experiment gives us little indication of the relative merits of CTRNNs when compared to other approaches. Further investigations necessitate to be focused on comparative analysis of the functional properties of different neural structures for plastic behaviour. We hope that the evolutionary ecological approach undertaken in our model might represent both a further methodological tool to test specific hypotheses about the selection pressures involved in the evolution of specific learning skills possessed by particular species.

Acknowledgements

The authors would like to thank all the members of the Centre for Computational Neuroscience and Robotics (http://www.cogs.sussex.ac.uk/ccnr/) for constructive discussion, the Sussex High Performance Computing Initiative for computing support, and the three anonymous reviewers for their useful comments.

Figure 1

This picture is adopted from Yamauchi and Beer (1994a). Environments for the one-dimensional navigation task. The triangle represents the agent, the circle represents the goal and the rectangle represents the landmark. The landmark-near environments (A and B) are on the left, the landmark-far environments (C and D) are on the right.

Figure 2

Depiction of the task. The small circle represents the robot. The white oval represents the landmark (i.e., a light source) and the black stripe in the arena is the goal. Picture A at the top represents the arena at the beginning of each single trial without landmark and goal. The black filled circles represent two possible starting points. The curved dotted lines represent two possible routes to the central part of the arena which is delimited by dashed lines. Pictures B and C represent the two possible arrangement of the landmark and goal within the landmark-near environment. The pictures D and E represent the two possible arrangement of the landmark and goal within the landmark-far environment. The arrows, in pictures B, C, D, E represent the directions towards which a successful robot should move.

Figure 3

Plan of a Khepera mini-robot showing sensors and wheels. The robot is equipped with 6 infra red sensors (Ir_0 *to* Ir_5) and 3 ambient light sensors ($A_1,\ A_4,\ A_6$). It also has a floor sensor indicated by the central gray circle (F).

Figure 4

Experiment I: each single graph refers to the performance of the best control system of the final generation of a single run, evaluated over 500 test sessions in both types of environment. Dashed lines indicate the percentage of successes per trial in landmark-near environment. Continuous lines indicate the percentage of successes per trial in landmark-far environment.

Figure 5

Experiment II: each single graph refers to the performance of the best control system of the final generation of a single run, evaluated over 500 test sessions in both types of environment. Dashed lines indicate the percentage of successes per trial in landmark-near environment. Continuous lines indicate the percentage of successes per trial in landmark-far environment.

Figure 6

Experiment II: The figure shows different measures of the performance of the best control system of each generation of run n. 10, evaluated over 500 test sessions in both types of environment. The performances per generation are given by averaging their frequencies recorded by the best controller during the evaluation in all the trials except the first one. Graph A refers to the percentage of successes in landmark-near environment. In graph B, the black area refers to the percentage of error in landmark-near environment, and the grey area underneath refers to the proportion of errors which are due to detour behaviour. Graph C refers to the percentage of successes in landmark-far environment. In Graph D the black area refers to the percent of error in landmark-far environment and the grey area underneath refers to the proportion of errors which are due to detour behaviour.

References

Anderson, R. W. (1995). Learning and Evolution: A Quantitative Approach. *Journal of Theoretical Biology*, 175:89–101.

Arnold, S. J. (1978). The evolution of a special class of modifiable behaviors in relation to environmental pattern. *The American Naturalist*, 984(112):415–427.

Baldwin, J. M. (1896). A New Factor in Evolution. *The American Naturalist*, 30:441–451.

Belew, R. K. and Mitchcel, M., editors (1996). *Adaptive Individuals in Evolving Populations: Models and Algorithms*, volume XXVI. Addison-Wesley Publishing Company, Inc., Santa Fe.

Bitterman, M. E. (2000). Cognitive Evolution: a psychological perspective. In Heyes, C. M., editor, *The Evolution of Cognition*, pages 61–79. MIT Press, Cambridge, MA.

Bolles, R. C. (1985). The Slaying of Goliath: What Happened to Reinforcement Theory. In Johnston, T. D. and Pietrewicz, A. T., editors, *Issues in the Ecological Study of Learning*, chapter 14, pages 387–399. Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey.

Craig, C. L. (1994). Limits to learning: effects of predator pattern and colour on perception and avoidance-learning by prey. *Animal behavior*, 47:1087–1099.

Dale, K. and Collett, T. S. (2001). Using Artificial Evolution and Selection to Model Insect Navigation. *Current Biology*, 11:17:1305–1316.

Davey, G. (1989). *Ecological learning theory.* Routledge, London, UK.

Di Paolo, E. (2000). Homeostatic adaptation to inversion of the visual field and other sensorimotor disruptions. In Meyer, J.-A., Berthoz, A., Floreano, D., Roitblat, H., and Wilson, S., editors, *From Animals to Animats VI: Proceedings of the* 6^{th} *Interntional Conference on Simulation of Adaptive Behavior*, pages 440–449. Cambridge, MA: MIT Press.

Di Paolo, E. A., Noble, J., and Bullock, S. (2000). Simulation Models as Opaque Thought Experiments. In *Artificial Life VII: The* 7^{th} *International Conference on the Simulation and Synthesis of Living Systems*, Portland, OR, USA.

Dukas, R. (1988). Evolutionary Ecology of Learning. In Dukas, R., editor, *Cognitive Ecology. The Evolutionary Ecology of Information Processing and Decision Making*, chapter 4, pages 129–174. The University of Chicago Press.

Dukas, R. (1998). Ecological relevance of associative learning in fruit fly larvae. *Behavioral Ecology and Sociobiology*, 19:195–200.

Dukas, R. (1999). Costs of memory: Ideas and predictions. *Journal of Theoretical Biology*, 197:41–50.

Floreano, D. and Urzelai, J. (2001). Neural Morphogenesis, Synaptic Plasticity, and Evolution. *Theory in Biosciences*, 120(3-4):223–238.

Funahashi, K. and Nakamura, Y. (1993). Approximation of Dynamical Systems by Continuous Time Recurrent Neural Networks. *Neural Networks*, 6:801–806.

Garcia, J. and Koelling, R. A. (1966). Relation of cue to consequence in avoidance learning. *Psychonimic Science*, 4:123–124.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning.* Reading, MA:Addison-Wesley.

Hailman, J. P. (1985). Historical Notes on the Biology of Learning. In Johnston, T. D. and Pietrewicz, A. T., editors, *Issues in the Ecological Study of Learning*, chapter 1, pages 27–57. Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey.

Harley, C. B. and Maynard-Smith, J. (1983). Learning - an evolutionary approach. *Trends in Neurosciences*, 6:204–208.

Harvey, I., Husbands, P., and Cliff, D. (1992). Issues in evolutionary robotics. In Meyer, J.-A., Roitblat, H., and Wilson, S., editors, *From Animals to Animats II: Proceedings of the* 2^{nd} *International Conference on Simulation of Adaptive Behavior*, pages 364–373, Cambridge MA. MIT Press/Bradford Books.

Hinde, R. A. and Stevenson-Hinde, J., editors (1973). *Constraints on Learning. Limitations and Predispositions.* ACADEMIC PRESS, London, UK.

Hinton, G. E. and Nowlan, S. J. (1987). How learning guides evolution. *Complex System*, 1:495–502.

Jakobi, N. (1997). Evolutionary robotics and the radical envelope of noise hypothesis. *Adaptive Behavior*, 6:325–368.

Johnston, T. D. (1981). Contrasting approaches to a theory of learning. *The Behavioral and Brain Sciences*, 4:125–173.

Johnston, T. D. (1996). Selective costs and benefits of learning. In Belew, R. K. and Mitchell, M., editors, *Adaptive Individuals in Evolving Populations: Models and Algorithms*, chapter 20, pages 315–358. Addison-Wesley Publishing Company, Inc.

Johnston, T. D. and Pietrewicz, A. T., editors (1985). *Issues in the Ecological Studies of Learning.* Erlbaum, Hillsdale, N.J., USA.

Johnston, T. D. and Turvey, M. T. (1980). A sketch of an ecological metatheory for theories of learning. *Psychology of Learning and Motivation*, 14:147–205.

Leahey, T. H. (1996). *A History of Psychology: Main Currents in Psychological Thought.* Prentice-Hall, Inc, New Jersey.

Macphail, E. M. (2001). The evolution of intelligence: adaptive specializations *versus* general process. *Biological Review*, 76:341–364.

McNeely, C. and Singer, M. C. (2001). Contrasting the roles of learning in butterflies foraging for nectar and oviposition sites. *Animal Behaviour*, 61:1–6.

Meyer, J.-A., Husbands, P., and Harvey, I. (1998). Evolutionary Robotics: A survey of Applications and Problems. In Husbands, P. and Meyer, J.-A., editors, *Evolutionary Robotics: Proceedings of the* 1^{st} *European Workshop, EvoRobot98.* Springer Verlag.

Morgan, C. L. (1896). On modification and variation. *Science*, 4:733–740.

Nager, R. G. and van Noordwijk, A. J. (1995). Proximate and ultimate aspects of phenotipic plasticity in timing of great tit breeding in a heterogeneous environment. *American Naturalist*, 146:454–474.

Nolfi, S. and Floreano, D. (1999a). Co-evolving predator and prey robots: Do arms races arise in artificial evolution? *Artificial Life*, 4(4):311–335.

Nolfi, S. and Floreano, D. (1999b). Learning and evolution. *Autunomous Robots*, 7(1):89–113.

Nolfi, S. and Floreano, D. (2000). *Evolutionary Robotics: The Biology, Intelligence, and Technology of Self-Organizing Machines.* Cambridge, MA: MIT Press/Bradford Books.

Osborn, H. F. (1896). Ontogenic and phylogenic variation. *Science*, 4:786–789.

Papaj, D. R. and Prokopy, R. J. (1989). Ecological and evolutionary aspects of learning in phytophagous insects. *Annual Review Entomology*, 34:315–350.

Plotkin, H. C. (1988). Learning and Evolution. In Plotkin, H. C., editor, *The Role of Behaviour in Evolution*, chapter 5. MIT Press.

Quinn, M. (2001). Evolving communication without dedicated communication channels. In *Advances in Artificial Life:* 6^{th} *European Conference on Artificial Life*, Prague, Czech Republic.

Seligman, M. E. P. (1970). On the generality of the laws of learning. *Psychological Review*, 77:406–418.

Seth, A. K. (1998). The evolution of complexity and the value of variability. In Adami, C., Belew, R., Kitano, H., and Taylor, C., editors, *Proceedings of the* 6^{th} *International Conference on Artificial Life*, pages 209–221, Los Angeles, California, US. MIT Press.

Shettelworth, S. J. (1984). Learning and Behavioural Ecology. In Krebs, J. and Davis, N., editors, *Behavioural Ecology: an evolutionary approach*, chapter 7, pages 170–194. Blackwell Scientif Publications, Oxford.

Stephens, D. W. (1993). Learning and Behavioral Ecology: Incomplete Information and Environmental Predictability. In Papaj, D. R. and Lewis, A. C., editors, *Insect Learning. Ecology and Evolutionary Perspectives*, chapter 8. Chapman & Hall, New York - London.

Stephens, D. W. and Clements, K. C. (1998). Game Theory and Learning. In Dugatkin, L. A. and Reeve, H. K., editors, *Game Theory and Animal Behavior*, pages 239–260. Oxford University Press, New York.

Todd, P. M. and Miller, G. F. (1991). Exploring adaptive agency II: Simulating the evolution of associative learning. In Meyer, J. A. and Wilson, S. A., editors, *From Animals to Animats I: Proceedings of the* 1^{st} *International Conference on Simulation of Adaptive Behavior*, pages 306–315. MIT Press.

Tolman, E. C. (1949). There is more than one kind of learning. *Psychological Review*, 56:144–155.

Waddington, C. H. (1942). Canalization of development and the inheritance of acquired characters. *Nature*, 150:563–565.

Wilcoxon, H. C., Dragoin, W. B., and Kral, P. A. (1971). Illness-induced aversions in rat and quail: relative salience of visual and gustatory cues. *Science*, 171:826–828.

Yamauchi, B. M. and Beer, R. D. (1994a). Integrating Reactive, Sequential, and Learning Behavior Using Dynamical Neural Network. In Cliff, D., Husbands, P., Meyer, J.-A., and Wilson, S. W., editors, *From Animals to Animats III: Proceedings of the* 3^{rd} *International Conference on Simulation of Adaptive Behavior*. MIT Press.

Yamauchi, B. M. and Beer, R. D. (1994b). Sequential behavior and learning in evolved dynamical neural networks. *Adaptive Behavior*, 2(3):219–246.

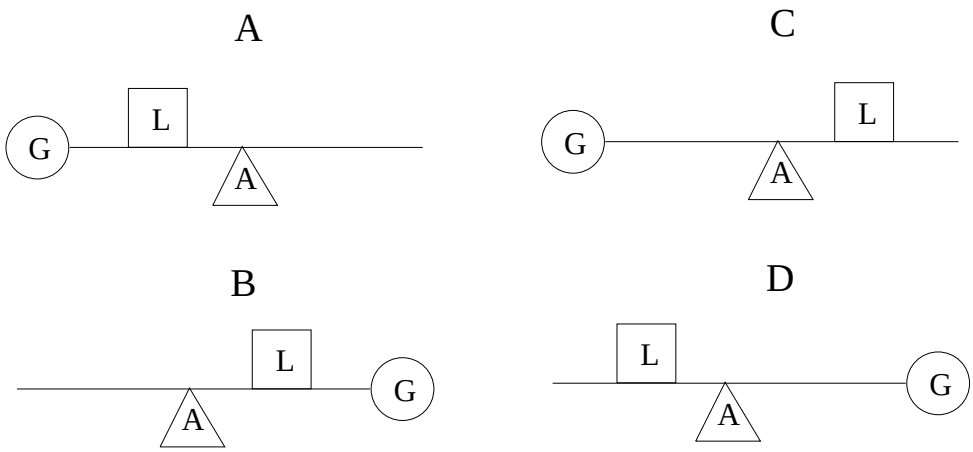

Figure 1:

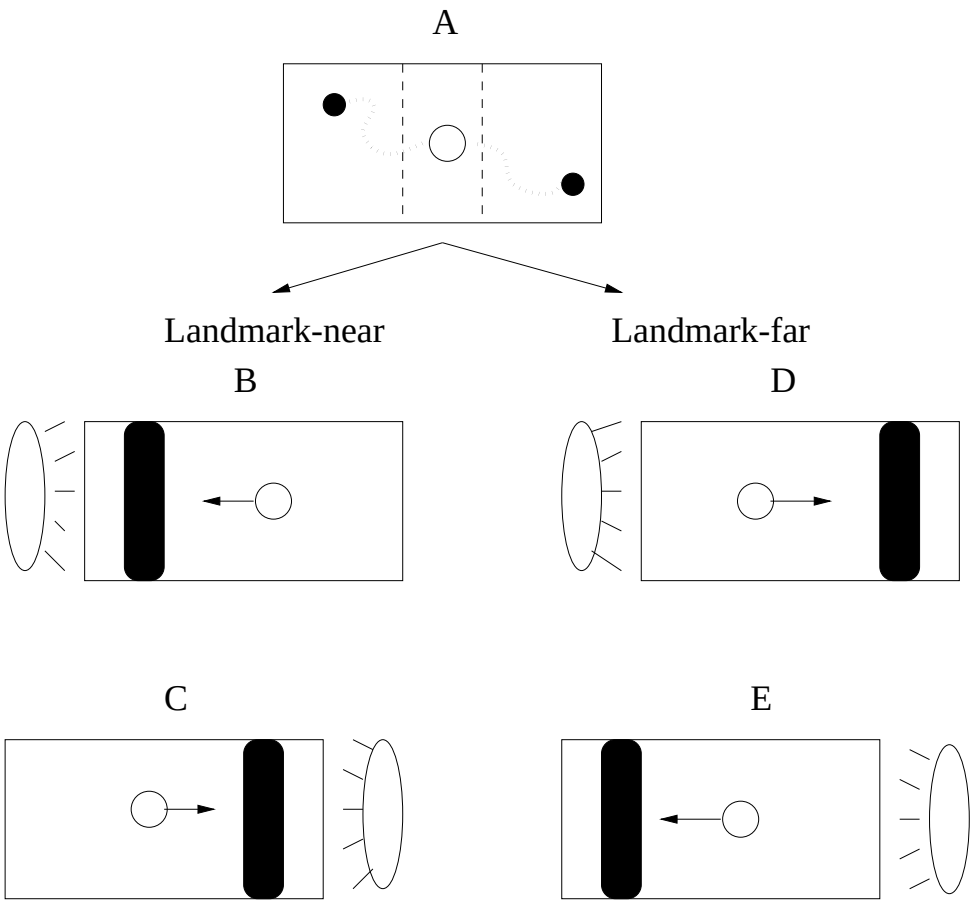

Figure 2:

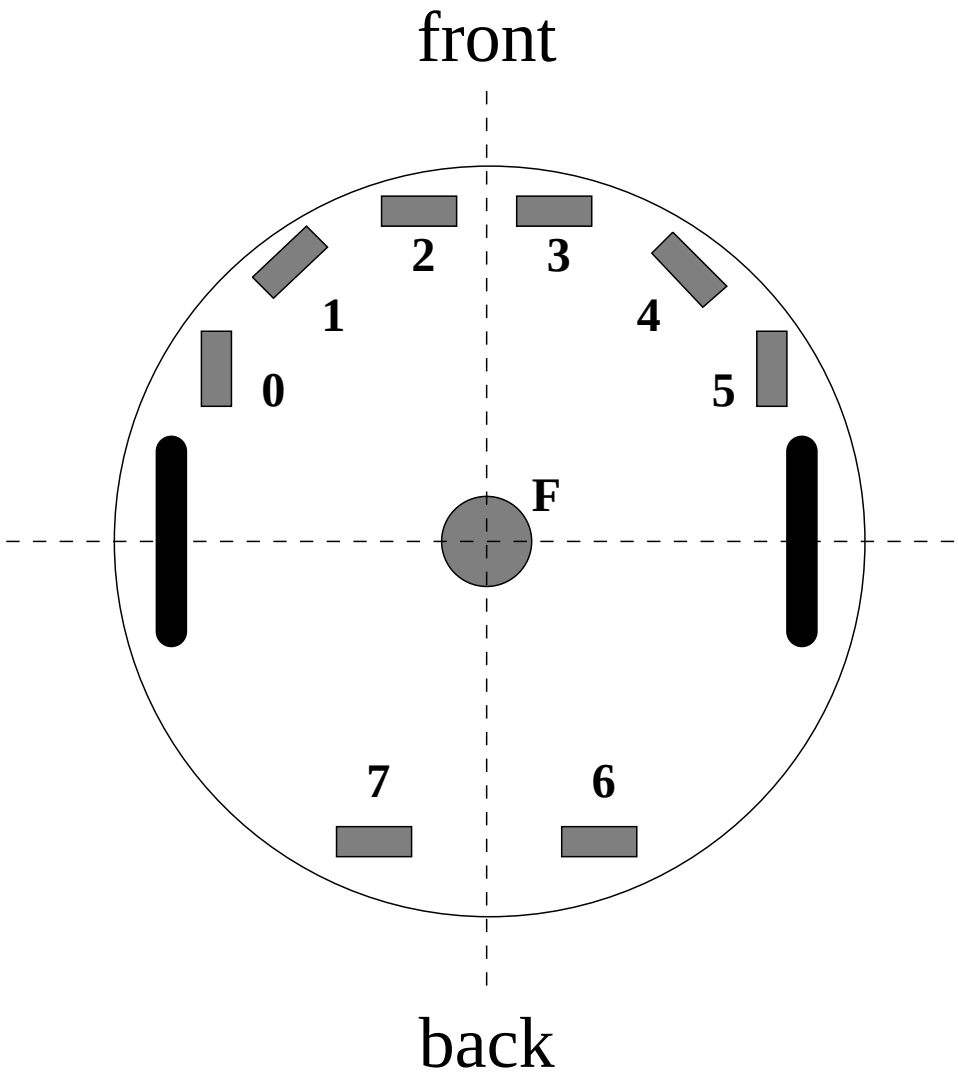

Figure 3:

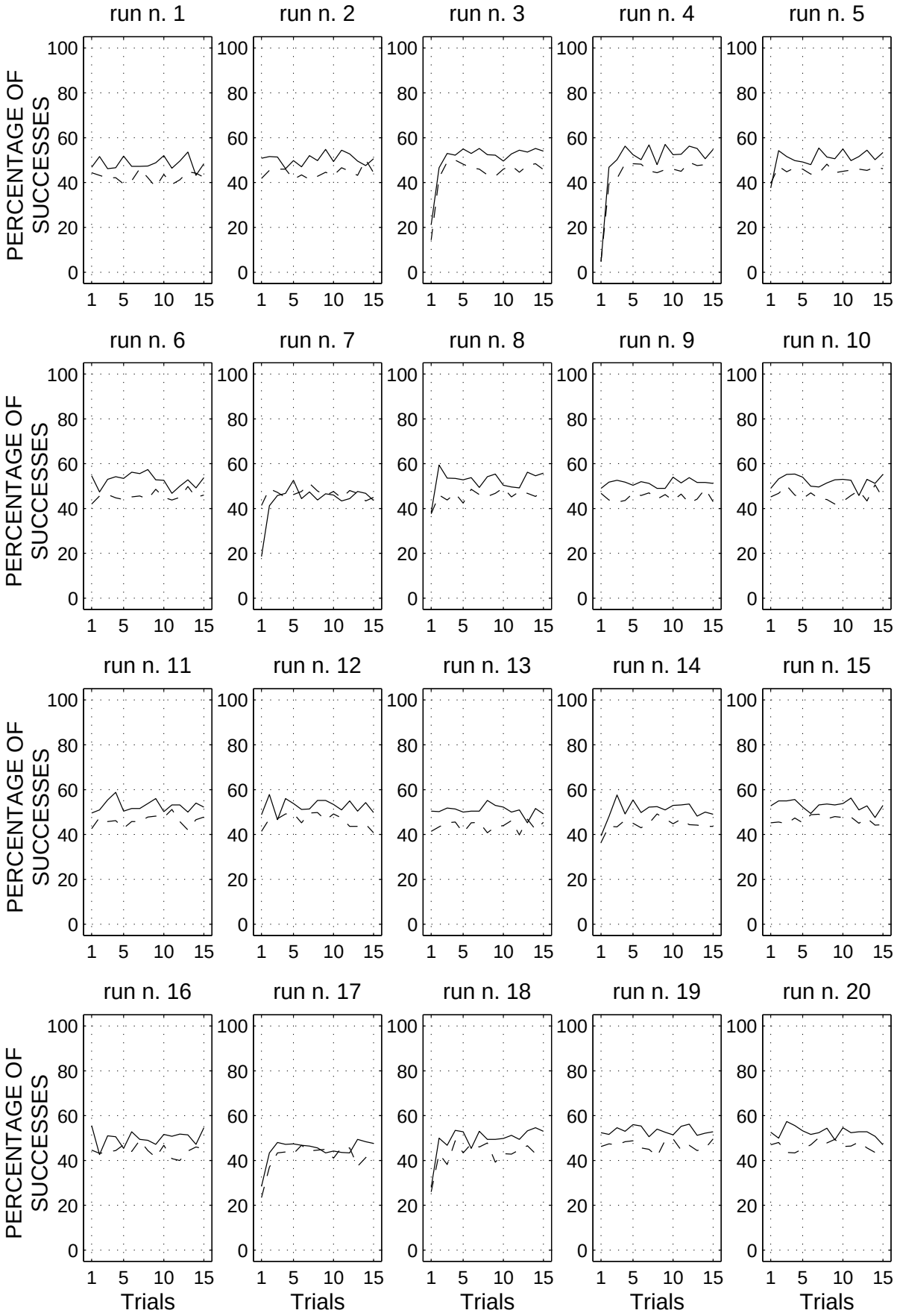
run n. 1
run n. 2
run n. 3
run n. 4
run n. 5
run n. 6
run n. 7
run n. 8
run n. 9
run n. 10
run n. 11
run n. 12
run n. 13
run n. 14
run n. 15
run n. 16
run n. 17
run n. 18
run n. 19
run n. 20
PERCENTAGE OF SUCCESSES
Trials
0
20
40
60
80
100
1
5
10
15

Figure 4:

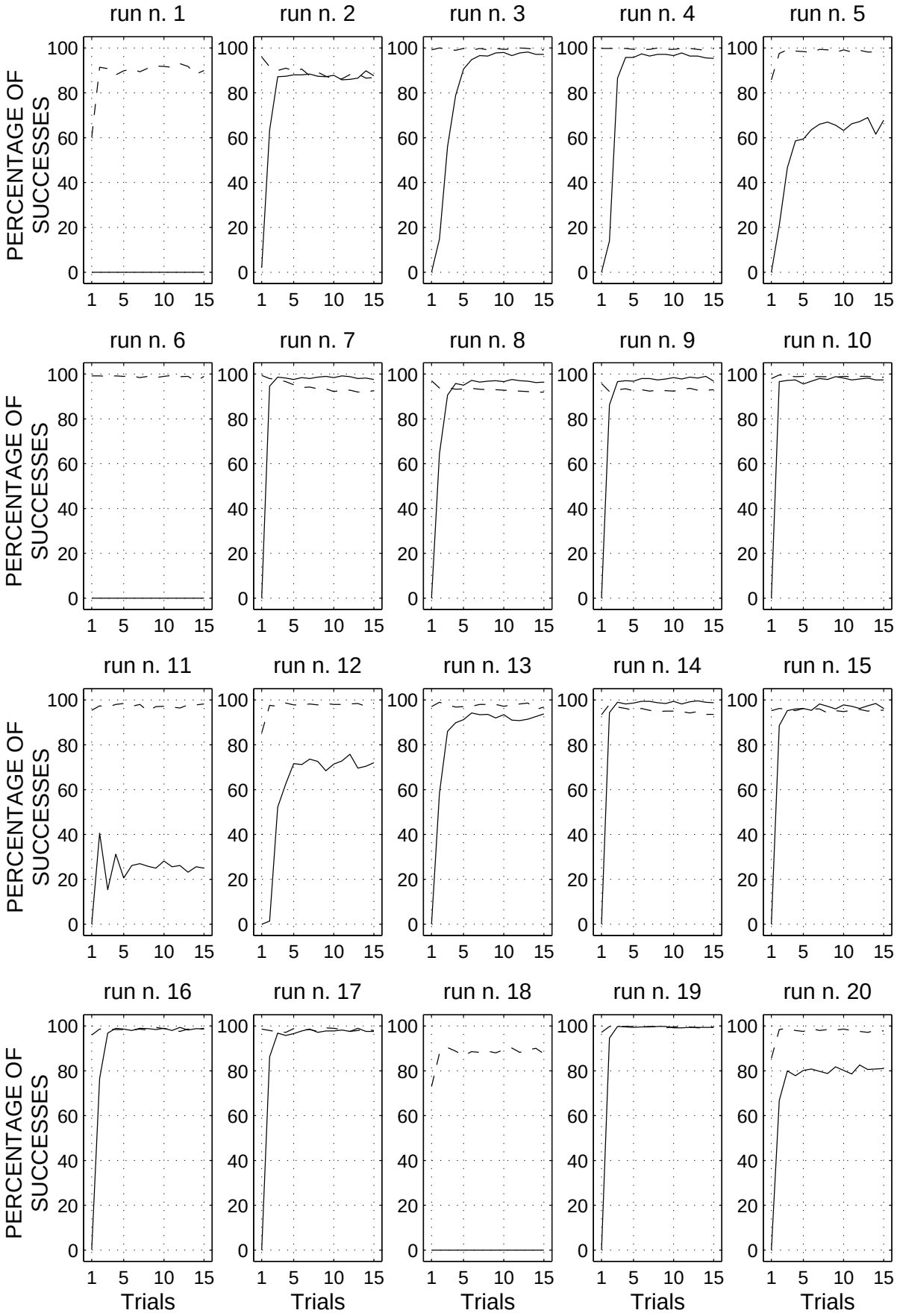
run n. 1
run n. 2
run n. 3
run n. 4
run n. 5
run n. 6
run n. 7
run n. 8
run n. 9
run n. 10
run n. 11
run n. 12
run n. 13
run n. 14
run n. 15
run n. 16
run n. 17
run n. 18
run n. 19
run n. 20
PERCENTAGE OF SUCCESSES
Trials
0
20
40
60
80
100
1
5
10
15

Figure 5:

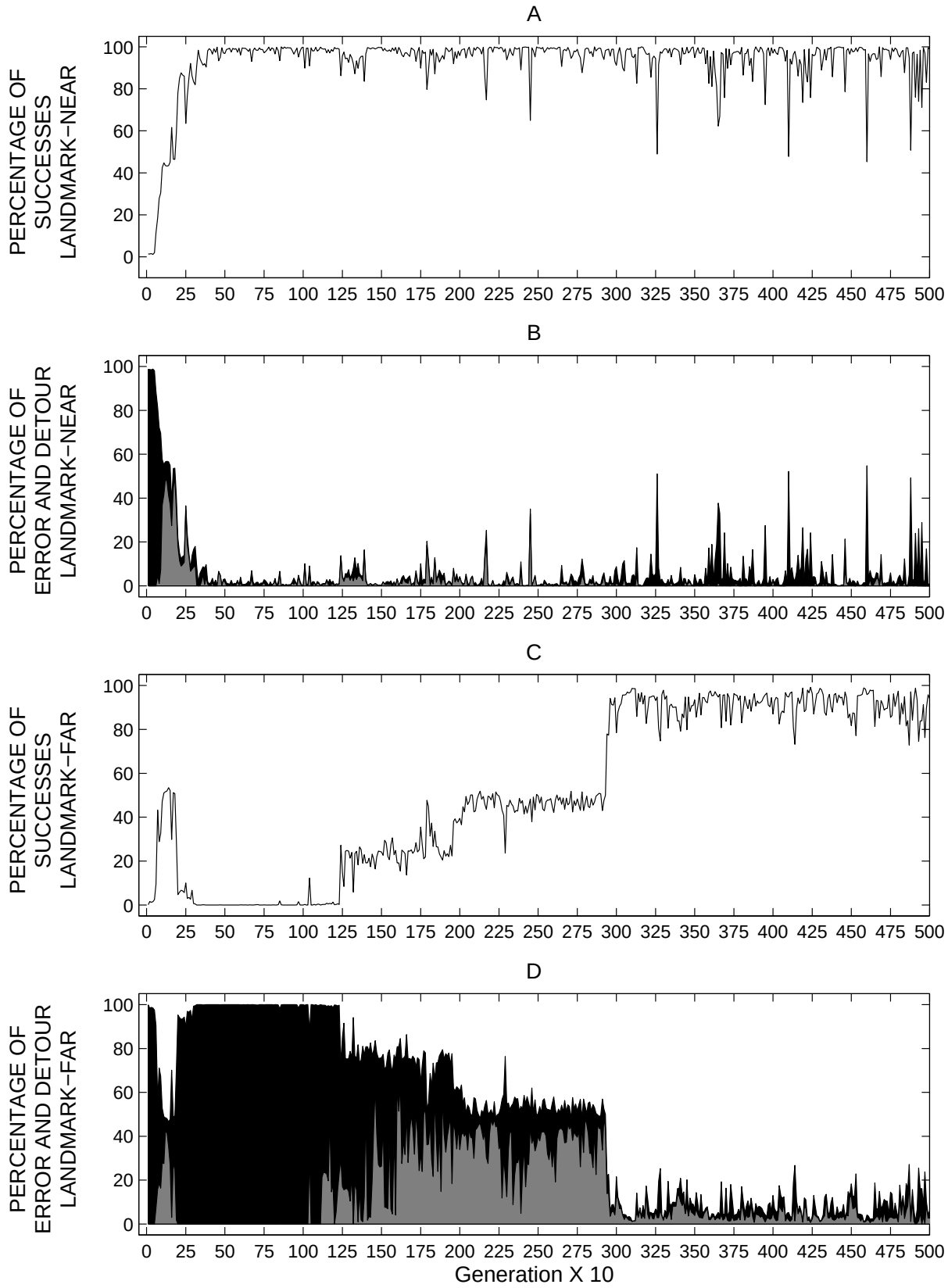
A
PERCENTAGE OF SUCCESSES LANDMARK−NEAR
B
PERCENTAGE OF ERROR AND DETOUR LANDMARK−NEAR
C
PERCENTAGE OF SUCCESSES LANDMARK−FAR
D
PERCENTAGE OF ERROR AND DETOUR LANDMARK−FAR
0 20 40 60 80 100
0 25 50 75 100 125 150 175 200 225 250 275 300 325 350 375 400 425 450 475 500
Generation X 10

Figure 6: