New Models for Old Questions: Evolutionary Robotics and the 'A Not B' Error

Rachel Wood and Ezequiel Di Paolo

CCNR, University of Sussex, Brighton BN1 9QH, UK.
r.k.wood@sussex.ac.uk

Abstract. In psychology the 'A not B' error, whereby infants perseverate in reaching to the location where a toy was previously hidden after it has been moved to a new location, has been the subject of fifty years research since it was first identified by Piaget [1]. This paper describes a novel implementation of the 'A not B' error paradigm which is used to test the notion that minimal systems evolutionary robotics modelling can be used to explore developmental process and to generate new hypotheses for test in natural experimental populations. The model demonstrates that agents controlled by plastic continuous time recurrent neural networks can perform the 'A not B' task and that homeostatic mediation of plasticity can produce perseverative error patterns similar to those observed in human infants. In addition, the model shows a developmental trend for the production of perseverative errors to reduce during development.

1 Introduction

In the fifteen years since its inception Evolutionary Robotics (ER) has become a core methodology in Artificial Life. ER offers advantages of automated design, bias reduction and proof of concept capabilities which allow researchers to explore behaviour with models which although simplified, retain their explanatory power [2]. Central to ER is the use of evolutionary algorithms to design artificial 'brains' (and morphologies) for robots and simulated agents: such methods serve to reduce the degree of designer bias built into systems and support the discovery of novel mechanisms underlying behaviour. ER also affords the generation of existence proofs - model systems which demonstrate a particular cognitive capacity under specific conditions. Such systems can be used to devise a set of minimal conditions required for a particular behaviour. Minimal modelling is an integral aspect of ER which reflects a general concern for bias reduction and cognition as an interactional property rather than a computational one. ER has been applied to a very wide range of problems however little work has been done to explore the use of these methods to model *developmental process.* The work described in this paper uses ER to design developmental trajectories in simple simulated agents.

Piaget's delayed manual search task, (a.k.a. the 'A not B' error), is one of the most explored phenomena in developmental psychology [1]. The finding that

F. Almeida e Costa et al. (Eds.): ECAL 2007, LNAI 4648, pp. 1141–1150, 2007.

7-12 month old infants make perseverative errors in retrieving a hidden toy has sparked decades of debate. Subtle alterations in the way in which the task is administered have a important effect on the behaviour observed. The canonical 'A not B' error paradigm has the infant seated at a table fitted with two lidded wells. The experimenter attracts the infant's attention with a toy which is then hidden in one of the wells, after a short delay the child is allowed to retrieve it. This process is repeated a number of times before the toy is hidden in the second well. The delay is again imposed before the child is allowed retrieve the toy. Perseverative errors are observed as subjects reach back to the first hiding place despite having seen the toy being hidden in the second. A great deal of experimentation has been carried out as researchers attempt to establish exactly what skill or knowledge or the perseverating infant lacks. Explanations range from the Piagetian notion that errors arise from immature notions of object physics to Thelen and Smith's persuasive contention that conflict between motor memories and specific task inputs is responsible for perserverative reaching [4].

On this view, the delayed search paradigm is an excellent test case for a minimal, developmental ER approach: the required behaviour is interestingly complex and has been well studied in the natural domain. In order to perform delayed manual search the agent, must first attend to the relevant stimuli and then organise an appropriate motor response. The agent should discriminate the two cases (location 'A' versus location 'B') and produce a different motor response for each. Memory is also implicated: the agent must 'mark' the cued location to the end of a delay period and only then issue a motor response. Thus we are not primarily concerned with *engineering* a solution to the 'A not B' task, and we do not expect to 'explain' perseverative reaching with a simple evolutionary robotics simulation. Rather, our purpose is to *test* the notion that minimal simulation models can be a useful addition to the experimental toolkit for exploring complex developmental phenomena.

This paper describes two experiments designed to test the notion that we can use minimal ER methods to model complex developmental trajectories and asks if we can use this approach *generatively* i.e. can we use the results of such work to come up with new questions and hypotheses for empirical testing. Thus, our measure of success will be whether we can generate a novel hypothesis for theoretical and empirical research on this problem. In both experiments an agent is placed between two acoustic sources ('A' and 'B') one of which emits a signal. A delay is imposed before the agent is allowed to locate the signal source. Thus, the infant's manual search has become locomotion to a target and visual cues have become auditory signals but the crucial interaction between attention to a stimulus and the formulation of a motor response is preserved.

2 Methods

The agent is modelled as a circular body, radius $R_0 = 4$, with two diametrically opposed motors and a two acoustic sensors positioned symmetrically at a 45 degree angle to the motors. The agent can rotate its sensors part way around its body and thus has the possibility to actively 'manage' attention by controlling sensor orientation. Motors can drive forward and in reverse and agent bodies are modelled as rigid, small and having very little mass. Thus motor output translates directly into tangential velocity as the point where the motor is located. The agent's behaviour is controlled by a continuous time recurrent neural network (CTRNN) controller with nine nodes: two sensor nodes, a fully connected, four node intranet, two motor nodes, and an additional motor node to control sensor orientation. Each node in the intranet receives an incoming synapse from each sensory node and each motor node receives an incoming synapse from each node in the intranet; nodes in the intranet also have self connections. Spatial discrimination is provided by the relative activity of the two sensors which, being in different positions on the agent's body, receive differing intensities of external perturbation depending on their positions relative to the source. Neural activity is calculated using:

$$\tau_i \dot{y_i} = -y_i + \sum_j w_{ji} z_j; \quad z_j = \frac{1}{1 + exp[-(y_j + b_j)]} \tag{1}$$

where y_i represents the cell potential, τ the decay constant, b_i the bias, z_i the firing frequency, w_{ij} the strength of the synaptic connection from node i to node j. Uniformly distributed noise, scaled by a gain factor is added to all sensor and motor transduction steps.

Sound is modelled as an instantaneous, additive field with volume attenuated in inverse proportion to distance squared. The intensity of incoming signal is further attenuated by a self-shadowing factor derived from the distance travelled by the signal through the body of the robot. This mechanism devised by Di Paolo [3], provides a simple but biologically plausible means for the agent to distinguish the direction of a sound source. The degree of attenuation is determined by movement and angular orientation of the robot. This body shadowing factor D_{sh} is modelled as linear attenuation proportional to the distance travelled through the body (ibid.).

$$Dsh = Dsen(1 - A), 0 \leq A < 1, A = \frac{D^2 - R_0^2}{D_{sen}^2} \tag{2}$$

where D_{sen} is the distance between the source and the sensor, and D is the distance between the source and the centre of the body. If $A \leq 1$, there is a direct line between source and sensor and $D_{sh} = 0$, maximum attenuation occurs when the sensor is on the opposite side of the body to the sound source. The intensity of incoming signal is calculated by taking the volume at sensor (attenuated by source to sensor distance) and then multiplying it by a linear scale derived from the degree of shadowing (1.0 when $D_{sh} = 0$, 0.1 when $D_{sh} = 2R_0$).

A tournament style GA with rank based selection and strong elitism is used. Populations (N = 60) are evolved for 1000 generations; vector mutation is the sole genetic operator employed. Twelve trials are given with fitness averaged across trials. Fitness is based on final position relative to the target ($F_A = 1 - D_{final}/D_{initial}$) and efficiency of approach trajectory (i.e. the proportion of approach time spent within 4 body radii of the target (F_D). Scores are modulated by an additional factor based on the agent's energetic output during the approach phase. The individual fitness F for a single trial is obtained using

$$F = (a_d F_D + a_a F_A) exp(-a_e \int_{t_i}^{t_f} \sqrt{V_R^2 + V_L^2 dt}) \tag{3}$$

where a_d= 0.25 and a_a=0.75 are respectively weighting factors for approaching and maintaining proximity to the target, $a_e = 0.005$ scales the exponential and V_R and V_L represent the translation speeds of the right and left motors respectively. Network parameters are encoded as real values in a vector of fixed dimensions. Mutation is carried out by perturbing the genome **G** with the probability $\mu = 0.005$ in a random direction by adding a normalized random vector **p** multiplied by a distance m (in the range [0,1]), thus $\mathbf{G} \longrightarrow \mathbf{G} + \mathbf{mp}$. Unless otherwise indicated parameters in the range [0,1] are mapped to [-3,3] for biases, [0.4,2] time constants and [0.05, 10] sensor gains, [-5, 5] motor gains. Parameter ranges were selected having been used successfully in previous ER work.

3 Experiment One

The agent moves on an unbounded, 2-D plane, its task, to locate and approach the source of an auditory signal emitted by one of two randomly placed beacons ('A' and 'B') located symmetrically with respect to the agent's initial position (initial distance in the range [35, 70] units). Agents begin each presentation at a fixed orientation. A 400 timestep trial comprises three phases: signal phase when a high volume, irregular signal is emitted by one beacon, delay period and approach phase when the agent is permitted to move toward the target (see fig. 1). Although the agent cannot move during the signal and delay phases sensor rotation is permitted. Both beacons emit a low volume variable signal during the delay and approach phases.

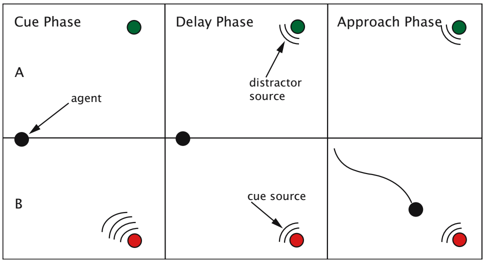

Fig. 1. Schematic of the delayed search task environment

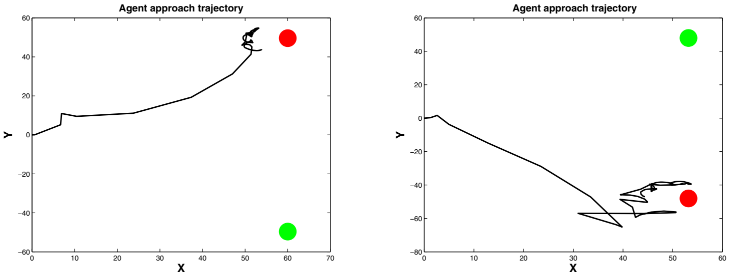

Fig. 2. Experiment one: approach trajectories for agent with increased source salience, figure shows two correct approaches one to the ‘A’ location (l) and one to the ‘B’ location (r).

3.1 Experiment One: Plasticity

The weightings on synaptic connections are set by a mechanism based on Hebbian learning. Four constants (α, β, γ and ϵ) and a learning rate, all in the range [-1,1] and under evolutionary control, determine the mapping between synaptic activity and weight change. The learning rate provides a degree of localised plasticity by modulating the rate at which individual synapses adapt. Weights are initialised with small random values and constrained to the range [-8.0, 8.0]. At each update step modifications to the weights on each synapse are obtained via

$$\frac{dw_{ij}}{dt} = \eta(\alpha y_i y_j + \beta y_j + \gamma y_i + \epsilon) \tag{4}$$

where η represents the learning rate for the synapse, and α, β, γ and ϵ are constants in the range [-1,1].

3.2 Experiment One: Results

A first set of evolutionary runs produced agents that could achieve the task but with difficulty and stereotyped behaviours. It was found that the task was easier to evolve if the beacons both emitted sound at a high volume during the approach phase (this does not break the conditions of the A not B paradigm since manipulations increasing saliency are also used in infant experiments). Five evolutionary runs have been carried out and four produced agents with fitness of 0.70 or higher. The best agent has a score of 0.74, its approach strategy differs from that used by the agents described above. Here, the agent does not approach all targets from the same direction and its motion trajectories are less stereotyped. Performance of the task is more efficient both in terms of the number of targets correctly located and the agent’s approach trajectories see fig 2. The agent finds its target in 10 trials (out of twelve) and approaches both ‘A’ and ‘B’ targets efficiently indicating that it discriminates between signals (rather than responding to the presence/absence of a cue from one side). However, there is little evidence of a perseverative pattern in its behaviour.

4 Experiment Two: The Role of Homeostatic Mechanism

In the first experiment described, the agent makes approach errors but they are not systematically perseverative: the agent does not regularly go to A when the signal has swapped to B. Iizuka and Di Paolo describe a minimal form of behavioural preference in simulated agents: in their model a homeostatic mechanism is used to implement selective behaviour which is temporally sustained without being invariant i.e. the preference develops over time and may also be altered [7]. A plastic neurocontroller is evolved with two high dimensional 'boxes' or regions defined in the space of neural dynamics. An association between the regions and the behavioural options available to the agent is created through evolution. Their model shows durability of preference and preference transitions through a process of mutual constraint between internal and external dynamics which provides the link between environmental change and contingent neural reorganisation. Their model has some interesting features in common with the A not B error paradigm; the delayed manual search task can be conceived in terms of choice, (i.e. the infant has two options for retrieving the toy), and, on this view, perseverative errors take the form of a preference or habit enacted at some level of internal dynamics. In the dynamic fields model of perseverative reaching, the error is produced through the action of a 'pre-shape' in motor planning space which biases the system toward repeated actions rather than the generation of new ones [9]. This 'motor memory' based model of perseveration is highly persuasive but requires a relatively complex modelling approach. So, it is interesting to ask if a very minimal model of homeostatic behaviour control can be used to invoke perseverative responding in simple simulated agents.

4.1 Experiment Two: Methods

To explore this question further experiment one has been extended to include homeostatic mediation of synaptic plasticity. Two homeostatic regions are specified as a function of the firing-rate of the post-synaptic node and these determine the rate of change in the pre-synaptic weights. Plasticity is zero inside the regions and elsewhere is determined by a function based on post-synaptic firing. The regions correspond to firing rates of [0.15, 0.4] and [0.6, 0.85], (their locations are arbitrary and here we take the ranges used in [7]). At the beginning of each evolutionary run correspondence between the regions and the 'A' and 'B' targets is randomly specified for each node. Thus one neuron's 'A' box maybe the high firing region and another's the low firing region. Two region control of plasticity is not applied to sensory or motor neurons to avoid over biasing their activity. For these neurons a single region is specified in the range [0.15, 0.85]. Weight change is calculated using a Hebbian rule with linear dependence on the firing rate of the pre-synaptic node, an evolved learning rate is also applied (as experiment one.) Weights are updated according to:

$$\delta w_{ji} = \eta_{ji} z_i p(z_j) \tag{5}$$

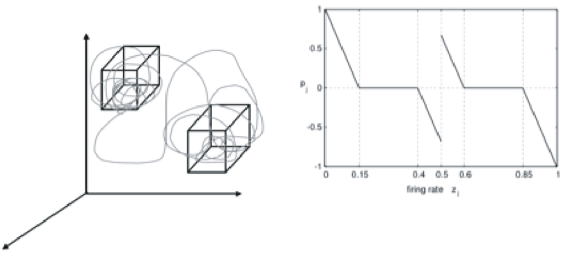

Fig. 3. Schematic of plastic facilitation for intranet nodes, two high-dimensional ‘boxes’ or regions are specified in the space of neural firing. Right: how plasticity corresponds to the homeostatic regions, the plot shows the plasticity function (P_j) as a function of neural firing (z_j). When post-synaptic firing is in one of two flat regions $P_z = 0$ and plasticity is inhibited (schematic adapted from [7]).

where δw_{ji} is the change in w_{ji} per unit of time, $p(x)$ is the plasticity function and η_{ji} the learning rate. Parameters are restricted to positive numbers to ensure that the function brings firing rates back toward the specified firing regions. Fitness is calculated as previously with the addition of a term designed to reward homeostatic agents. For every time step that a specified neuron behaves homeostatically a counter is incremented by 1.0; thus, on an ‘A’ trial the fitness is awarded for homeostasis if the firing rate of the specified neuron remains within the ‘A’ box. If it fires at a rate which is inside the ‘B’ box no fitness is awarded and if the firing rate is outside both an increment of 0.5 is given. This score is averaged for all nodes and the term obtained F_H is used to scale fitness awarded for approach and proximity to target. In all other respects the model is as described in experiment one:

$$F = (F_D + F_P)F_H \tag{6}$$

4.2 Experiment Two: Results

Five evolutionary runs have been completed, two produced agents with fitness of 0.6 or above and the best agent scored 0.62, here we examine a single agent. The approach paths taken by this agent are direct and having located the source it remains there for the remainder of the trial. The agent is equally successful on both ‘A’ and ‘B’ trials. The reduction in overall fitness reflects scaling of scores for approach and proximity by F_H, the score for homeostasis. The error pattern produced by the homeostatic agent is unlike that observed in experiment one. Non-homeostatic agents make more frequent errors and mistakes do not follow a particular pattern. In the homeostatic system errors occur less often and tend to be clustered around ‘swaps’ (i.e. the presentation immediately following transition from signal delivered at ‘A’ to signal delivered at ‘B’ and vice

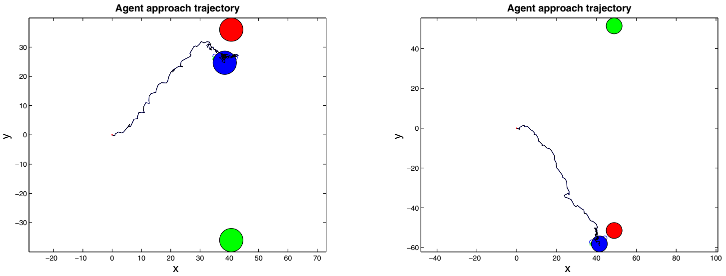

Fig. 4. Experiment two: approach trajectories for agent with homeostatic mediation of plasticity, figure shows one ‘A’ (l) and one ‘B’ (r) trial.

versa). The rate of perseverative errors in homestatic versus non-homeostatic agents is shown in figs 5 and 5. Here, trials comprising six presentations to one side followed by 6 presentations to the other are administered. Start order is randomised so that agents receive an equal number of trials beginning on each side. The first presentation in each trial has been discarded as errors here cannot be considered perseverative as the agent has had no previous exposure to the test environment.

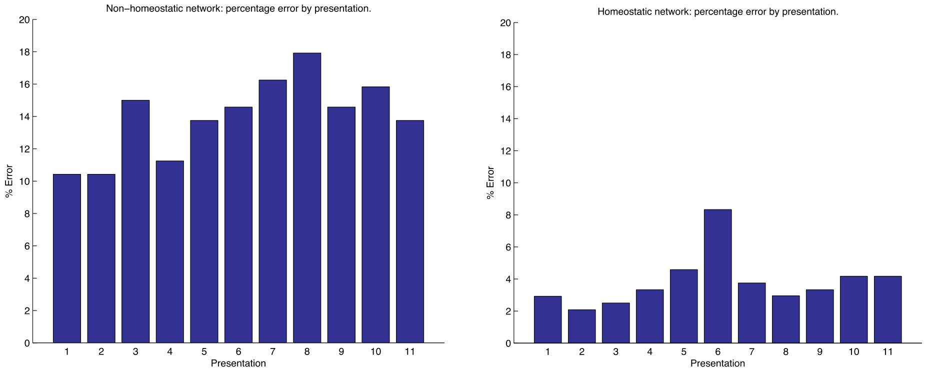

Fig. 5. Percentage error by presentation (l) non homeostatic network, (r) Homeostatic network, 240 independent trials administered.

Figure 5 (left) shows the percentage error made on each presentation by a non-homeostatic agent evolved under the conditions described in experiment one, 240 trials were given. The error rate varies between 10 and 20 per cent and errors are made equally often on all presentations. There is a slight increase in the number of errors following the signal swap but there is no evidence of a peak in errors on the presentation immediately following it. Figure 5 (right) shows

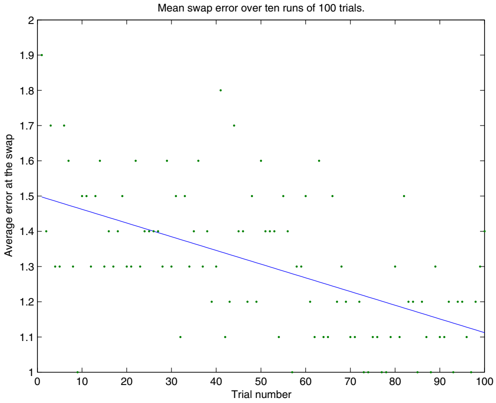

Fig. 6. Average of ten independent runs of 100 trials comprising 6 ‘A’ and 6 ‘B’ presentations, line = linear regression on the data, negative slope indicates that the swap error reduces over time.

the percentage of errors made by a homeostatic network under the same trial regime, here we see a fewer errors overall and a peak in the number of errors on the trial following the signal swap. Examination of the behaviour of agents on trial six, (following the swap), indicates that the errors made are perseverative i.e. agents continue to approach the source at which the signal was delivered in the previous six presentations. Figure 6 shows the mean swap error over ten independent runs of 100 trials (with the 6/6 presentation regime described above). Linear regression on the data produces a trend line with a negative slope (fig 6 solid line) indicating that swap errors decrease over time. Data points are widely scattered thus the R-square value indicates a weak fit. However, analysis of the mean swap error over time indicates that there is a developmental trend in the production of perseverative errors in the model. Further statistical analysis is required to properly explore this result.

5 Discussion

This paper presents results from two experiments: in the first agents must rely on the signal emitted by one source at the beginning of each trial as the only cue indicating the location to approach. In the second experiment the task of locating the selected source is simplified because sources emit sound throughout the trial, however the agent is still required to ‘remember’ where the signal came from. Plastic changes and processes of neural self-organisation play a significant role the behaviour of agents in this task. The homeostatic mechanism is introduced as a means to test if we can obtain perseverative responding; this strategy reflects a view of homeostasis as a mechanism for conservation in adaptation. In experiment two plasticity is mediated by an evolutionary requirement that the activity of internal neurons in the controller should remain within one of two regions in the space of neural activity corresponding with the two types of task environment (i.e. 'A' and 'B'). Thus, a constrained developmental process limits the means by which performance may be implemented. This aspect of the work is interesting for a number of reasons: very little work has been done with plastic CTRNN controllers using homeostatic mechanisms of adaptation, and this experiment combines exploration of homeostatic dynamics with further investigation of the application of the minimal modelling approach. The homeostatic mechanism reproduces a pattern of perseverative errors observed in human infants and also a developmental trend for such errors to disappear. Neither of these characteristics has been explicitly selected for and therefore we can hypothesise that these results are a consequence of having included homeostatic mechanisms in the model. These results allow the generation a hypothesis about mechanisms underlying the 'A not B' error in infants such that there is some process of regulation of plasticity which plays a role similar to the homeostatic mechanism in the model. This process supports the conservation of patterns of neural activation with the effect of adapting to repeated sensorimotor requirements. An original aim of this work was to test the notion that minimal ER simulation can be used to model complex developmental trajectories, the results discussed here indicate that such methods can be valuable tools for exploring empirical problems.

References

1. Piaget, J.: The Origins of Intelligence in the Child. Routledge & Kegan Paul, London (1952)
2. Harvey, I., Di Paolo, E., Wood, R., Quinn, M., Tuci, E.: Evolutionary Robotics: A New Scientific Tool for Studying Cognition. Artificial Life 11(1-2) (2005)
3. Di Paolo, E.A.: Behavioural Coordination, Structural Congruence and Entrainment in a Simulation of Acoustically Coupled Agents. Adaptive Behaviour 8(1), 27–48 (2000)
4. Thelen, E., Smith, L.B.: A Dynamic Systems Approach to the Development of Cognition and Action. MIT Press, Cambridge (1994)
5. Beer, R.: Toward the Evolution of Dynamical Neural Networks for Minimally Cognitive Behaviour. In: Maes, P., Mataric, M., Meyer, J-A., Pollack, J., Wilson, S. (eds.) From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behaviour, MIT Press, Cambridge (1996)
6. Ashby, W.: Design for a Brain. John Wiley and Sons Inc., Chichester (1960)
7. Iizuka, H., Di Paolo, E.A.: Toward Spinozist Robotics: Exploring the Minimal Dynamics of Behavioural Preference. Adaptive Behaviour (Accepted)
8. Tuci, E., Quinn, M., Harvey, I.: Using a Net to Catch a Mate: Evolving CTRNNs for the Dowry Problem. In: From Animals to Animats 7: Proceedings of the Seventh International Conference on Simulation of Adaptive Behaviour, MIT Press, Cambridge (2002)
9. Thelen, E., Schoener, G., Scheier, C., Smith, L.: The Dynamics of Embodiment: A Dynamic Field Theory of Infant Perseverative Reaching. Behavioural and Brain Sciences 24, 1–86 (2001)