Adaptive Behavior

http://adb.sagepub.com

The Dynamics of Associative Learning in Evolved Model Circuits

Phattanard Phattanasri, Hillel J. Chiel and Randall D. Beer

Adaptive Behavior 2007; 15; 377
DOI: 10.1177/1059712307084688

The online version of this article can be found at:
http://adb.sagepub.com/cgi/content/abstract/15/4/377

Published by:
SAGE Publications
http://www.sagepublications.com

On behalf of:
ISAB
International Society of Adaptive Behavior

Additional services and information for *Adaptive Behavior* can be found at:

Email Alerts: http://adb.sagepub.com/cgi/alerts

Subscriptions: http://adb.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations (this article cites 30 articles hosted on the
SAGE Journals Online and HighWire Press platforms):
http://adb.sagepub.com/cgi/content/refs/15/4/377

Downloaded from http://adb.sagepub.com at PENNSYLVANIA STATE UNIV on April 16, 2008
© 2007 International Society of Adaptive Behavior. All rights reserved. Not for commercial use or unauthorized distribution.

Adaptive Behavior

The Dynamics of Associative Learning in Evolved Model Circuits

Phattanard Phattanasri[1], Hillel J. Chiel[2], Randall D. Beer[3]
[1]*Department of Electrical Engineering and Computer Science, Case Western Reserve University, Cleveland, OH 44106*
[2]*Departments of Biology, Neurosciences, and Biomedical Engineering Case Western Reserve University Cleveland, OH 44106*
[3]*Cognitive Science Program, Department of Computer Science, Department of Informatics, Indiana University, Bloomington, IN 47406*

In this article, we evolve and analyze continuous-time recurrent neural networks capable of associating the smells of different foods with edibility or inedibility in different environments. First, we present an in-depth analysis of this task, highlighting the evolutionary challenges it poses and how these challenges informed our experimental design. Next, we describe the evolution of nonplastic neural circuits that can solve this food edibility learning problem. We then show that the dynamics of the best evolved nonplastic circuits instantiate finite state machines that capture the combinatorial structure of this task. Finally, we demonstrate that successful circuits with Hebbian synaptic plasticity can also be evolved, but that such circuits do not utilize their synaptic plasticity in a traditional way.

Keywords evolution of learning · evolutionary robotics · dynamical systems theory · synaptic plasticity · neuroscience

1 Introduction

The ability to improve performance with experience is an essential and ubiquitous characteristic of biological organisms, from habituation to mechanical stimuli in protozoa (Wood, 1969), to drought avoidance learning in plants (Trewavas, 2003), to aversive olfactory learning in nematodes (Zhang, Lu, & Bargmann, 2005), and abstract concept learning in human beings. Yet many fundamental questions about learning remain unresolved. How many different kinds of learning are there and how can we distinguish between them? How does learning evolve? How are the mechanisms of learning integrated into the mechanisms of behavior? In this article, we explore some of these questions by evolving and analyzing continuous-time recurrent neural networks (CTRNNs) that can solve an associative learning task.

In any discussion of learning, there is a strong tendency to envision a hierarchical architecture in which the mechanisms responsible for behavior and the mechanisms responsible for learning are distinct, with the latter modifying the parameters of the former (Figure 1a). For example, control of behavior is generally associated with the electrical activity of nerve cells, while learning is generally associated with chemical

Correspondence to: Randall D. Beer, Cognitive Science Program, 1910 E. 10th St. – 840 Eigenmann, Indiana University, Bloomington, IN 47406.
E-mail: rdbeer@indiana.edu; *URL*: http://mypage.iu.edu/~rdbeer/
Tel.: (812) 856-0873; *Fax*: (812) 855-1086

Copyright © 2007 International Society for Adaptive Behavior
(2007), Vol 15(4): 377–396.
DOI: 10.1177/1059712307084688

Downloaded from http://adb.sagepub.com at PENNSYLVANIA STATE UNIV on April 16, 2008
© 2007 International Society of Adaptive Behavior. All rights reserved. Not for commercial use or unauthorized distribution.

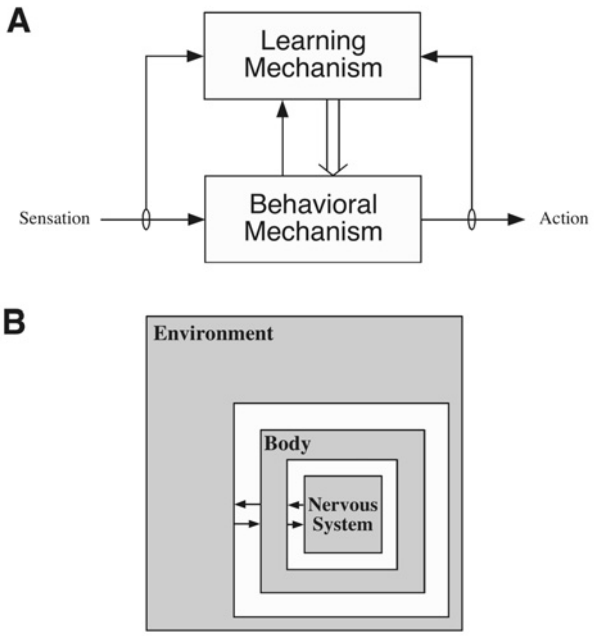


Figure 1 Perspectives on learning. (a) The traditional perspective on learning is hierarchical; a separate learning mechanism observes an agent's sensation, action and internal state (single arrows) and modifies the operation of the mechanisms responsible for its behavior (double arrow) so as to improve performance. (b) From a dynamical perspective, all behavior arises from the interaction between an agent's nervous system, its body and its environment. If the environment and the agent have dynamics on longer timescales, then some of that behavior may be interpretable as learning.

plasticity in the synapses which interconnect neurons. However, this identification of a behavioral distinction with a neuronal distinction is difficult to maintain in the face of growing evidence for considerable overlap between the time-scales of neural activity (Llinas, 1988; Toledo-Rodriguez, El Manira, Wallen, Svirskis, & Hounsgaard, 2005) and the timescales of synaptic plasticity (Bi & Rubin, 2005; Kandel & Siegelbaum, 2000; Sutton & Carew, 2002). For example, work on learning and memory in *Drosophila* has found that behavioral changes can occur over timescales ranging from seconds to days (Margulies, Tully, & Dubnau, 2005). In addition, work on multiple vertebrate systems has found that synaptic plasticity can be sensitive to differences in spike timing on the order of tens of milliseconds (Dan & Poo, 2004). Finally, forms of memory can be implemented using the dynamics of intrinsic membrane currents (Marder, Abbott, Turrigiano, Liu, & Golowasch, 1996) or through persistent activity in recurrent circuits (Major & Tank, 2004).

Modeling work on the evolution of learning has likewise tended to assume distinct mechanisms for behavior and learning. For example, Chalmers (1991) evolved the parameters of a general synaptic update equation on a supervised learning task and found that the well-known delta or Widrow-Hoff rule often evolved. As a second example, Miller and Todd explored the interaction of learning and evolution by evolving networks with genetically-selectable Hebbian plasticity on a food edibility task based on noisy sensory cues (Miller & Todd, 1991; Todd & Miller, 1991a, 1991b). Likewise, most work on the evolution of learning in the evolutionary robotics community has simply postulated the existence of various learning rules a priori and then evolved the parameters of such rules (Floreano & Mondada, 1996; Floreano & Urzelai, 2001). However, a traditional strength of evolutionary robotics has always been to minimize a priori assumptions, using evolution to explore the space of possible solutions to a problem without imposing our theoretical preconceptions.

Although many view synaptic plasticity as definitional of learning, strictly speaking, learning is a *behavioral* phenomenon, whose underlying mechanisms remain to be empirically investigated. These observations obviously do not undermine the central role that synaptic plasticity plays in biological learning. However, they do suggest the need for a more sophisticated perspective on the mechanisms of learning than that illustrated in Figure 1a. Thus, a much better strategy for exploring the evolution of learning would seem to be: (1) Set tasks that require behavioral plasticity for their solution; (2) Evolve agents that can accomplish these tasks *using a neural model that does not include an explicit learning mechanism*; (3) Examine how successful circuits actually implement the learning behavior. This approach fits in quite naturally with a dynamical perspective on behavior (Figure 1b), in which learning is interpreted as merely a particular kind of dynamics at a particular range of timescales arising from the interaction of a nervous system, body and environment (Beer, 1997).

In this article, we extend previous work on the evolution and analysis of neural circuits for learning simple sequential decision-making tasks such as those that arise in landmark-based navigation (Yamauchi &

Downloaded from http://adb.sagepub.com at PENNSYLVANIA STATE UNIV on April 16, 2008
© 2007 International Society of Adaptive Behavior. All rights reserved. Not for commercial use or unauthorized distribution.

Beer, 1994a, 1994b) to a more traditional associative learning task. The particular associative learning task, neural models, and evolutionary algorithm that we employ are described in Section 2. Section 3 presents an in-depth analysis of our learning task, highlighting the evolutionary challenges raised by this task and explaining how these challenges informed our experimental design. In Section 4, we demonstrate that CTRNNs lacking synaptic plasticity can be successfully evolved to exhibit associative learning. The dynamical operation of the best such circuit is then analyzed in detail in Section 5. For comparison, preliminary experiments with CTRNNs incorporating Hebbian synaptic plasticity are described in Section 6. Finally, Section 7 concludes with a discussion of the broader implications of our results and directions for future work.

2 Methods

2.1 Learning Task

The associative learning paradigm we study is abstracted from animal experiments on food edibility learning in *Aplysia* (Chiel & Susswein, 1993; Susswein, Schwarz, & Feldman, 1986) and is similar to the task explored by Todd and Miller (Miller & Todd, 1991; Todd & Miller, 1991a). We chose the simplest possible scenario that required associative learning for its solution. Two kinds of food are available to an agent. Two types of environments are distinguished by which food is edible and which is inedible. The agent receives two sensory inputs. A binary "smell" sensor S distinguishes the two types of food and a continuous reinforcement sensor R receives positive or negative reward for the agent's actions. The reinforcement can be loosely interpreted as coming from a gut sensor that signals the consequences of the agent's previous action after a delay. The only action that the agent can take is to open or close its mouth via a continuous effector output M.

A single trial is structured as follows (Figure 2). Normally, both the smell and reinforcement signals are 0. A trial begins with the presentation of a smell. After 10 time units, the smell is removed and the state of the mouth is evaluated for 10 additional time units. Next, a delay occurs whose duration is uniformly distributed over the range [8, 10] time units. This delay can be interpreted as the time it takes the agent to digest the food it consumes. Then the agent receives positive or negative reinforcement for 10 time units based on the correctness of its previous action for the given smell and environment type. A length-K trial sequence consists of a series of K such trials, sepa-

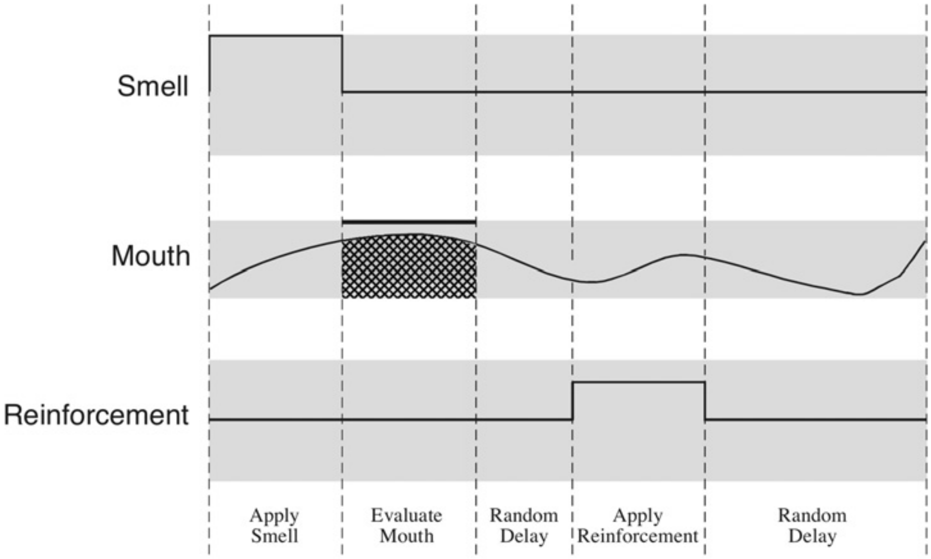


Figure 2 The structure of an individual trial. A trial is divided into five phases. First, a smell signal is applied. Second, the state of the mouth is evaluated relative to the correct action for the current environment (bold line). Third, there is a variable random delay. Fourth, a reinforcement signal proportional to the correctness (cross-hatched region) of the previous action is applied. Fifth, there is another variable random delay before the next trial begins.

Downloaded from http://adb.sagepub.com at PENNSYLVANIA STATE UNIV on April 16, 2008
© 2007 International Society of Adaptive Behavior. All rights reserved. Not for commercial use or unauthorized distribution.

rated by time intervals uniformly distributed over the range [16, 24] time units. This second delay can be interpreted as the time it takes an agent to encounter another patch of food. The type of environment can change between trials. In the course of a single experiment, the environment may change type multiple times.

The rationale for the experimental details that we are about to describe are based on a task analysis that will be given in Section 3. The error of an agent's action on the kth trial is given by

$$E_k = \int_{T_k+10}^{T_k+20} |A_k - M(t)| \psi(T_k + 10, t) \mathrm{d}t \tag{1}$$

where T_k is the time at which the kth trial begins, A_k is the correct motor output for the given food type and environment type on this trial, $M(t)$ is the agent's actual motor output during the evaluation period, and $\psi(t_0, t) \equiv \exp(-(-t - t_0 - 5)^2/5.12)/4.0034$ is a Gaussian weighting function normalized so that E_k runs between 0 and 1. Gaussian weighting assigns maximum importance to the error of the agent's action at the center of the evaluation period, with the importance smoothly falling off at earlier and later times. The normalized reinforcement is obtained from the error by $R_k = 1 - 2E_k$ and is applied for the entire reinforcement period.

2.2 Fitness Evaluation

The agent's task is to maximize its reinforcement by consuming as much of the edible food and as little of the inedible food as possible regardless of which environment it is in. A complete experiment consists of a set of P length-K trial sequences. The total fitness of an agent on a complete experiment is given by

$$F = 1 - \frac{1}{P} \sum_{p=1}^{P} \sum_{k=1}^{K} \alpha_{K,k} E_{p,k} \tag{2}$$

where $E_{p,k}$ is the error on the kth trial of the pth sequence and the $\alpha_{K,k}$ are trial weighting coefficients. For $K = 2$ we have $\alpha_2 = \{0, 1\}$, while for $K = 3$, $\alpha_3 = \{0, 0.33, 0.67\}$. For $K > 3$, there are K coefficients of the form

$$\alpha_K = \left\{0, \frac{0.5}{K-1.7}, \frac{0.8}{K-1.7}, \frac{1}{K-1.7}, \ldots, \frac{1}{K-1.7}\right\}$$

where the denominator serves to enforce the normalization condition $\sum_{k=1}^{K} \alpha_{K,k} = 1$. The weight of the first trial is always 0 so that learning has time to take place and so that the agent is not unduly punished for the unavoidable mistake it makes after each environment change.

A five-stage incremental shaping protocol was employed during evolution. In the first stage, agents were exposed to all possible combinations of two-trial sequences in both environments, for a total of eight trial sequences. Similarly, in the second stage, agents were exposed to all 16 possible three-trial sequences in both environments. The next three stages similarly involved all possible six, seven, and eight-trial sequences. However, in these later stages, each trial sequence began in a randomly-selected environment and then randomly switched to the opposite environment after the third, fourth or fifth trial. Note that the trial weighting scheme described above must be restarted after an environment switch. Transitions between stages were triggered whenever the fitness of the best agent in the population consistently exceeded 95%. Once again, this complex experimental protocol is based on a task analysis that will be described in Section 3.

2.3 Neural Model

The agent's behavior is controlled by a continuous-time recurrent neural network (CTRNN). We will call CTRNNs whose synaptic weights are fixed during the lifetime of the agent *nonplastic* CTRNNs, whereas CTRNNs whose synaptic weights can change during the agent's lifetime will be called *plastic* CTRNNs. Nonplastic CTRNNs are described by the following state equation:

$$\tau_i \dot{y}_i = -y_i + \sum_{j=1}^{N} w_{ji} \sigma(y_j + \theta_j) + s_i S(t) + r_i R(t)$$

$$i = 1, \ldots, N$$

where y_i is the state of the ith neuron, $\dot{y}_i$ denotes the time rate of change of this state, τ_i is the neuron's membrane time constant, w_{ji} is the weight of the connection from the jth to the ith neuron, θ_i is a bias term, and $\sigma(x) = 1/(1 + e^{-x})$ is the standard logistic output function. $S(t) \in [-1, 1]$ represents the weighted input from the binary smell sensor with weight s_i, and $R(t)$

Downloaded from http://adb.sagepub.com at PENNSYLVANIA STATE UNIV on April 16, 2008
© 2007 International Society of Adaptive Behavior. All rights reserved. Not for commercial use or unauthorized distribution.

$\in [-1, 1]$ represents the weighted input from the continuous reinforcement sensor with weight r_i. The output of neuron 1 is designated as the motor output $M(t) \equiv \sigma(y_1 + \theta_1)$. Neuron states were initialized to 0 at the beginning of each trial sequence and integrated with the forward Euler method using an integration step size of 0.1. With N time constants, N biases, N^2 connection weights and $2N$ sensor weights, a nonplastic N-neuron CTRNN is specified by $N^2 + 4N$ parameters.

Plastic CTRNNs obey a similar state equation, except that the weights of plastic synapses vary in time according to a Hebbian rule. There are many such rules to choose from (Baxter & Byrne, 1993). Here we utilize the covariance rule of Floreano and Mondada (1996):

$$\dot{w}_{ij} = \begin{cases} \eta_{ij}(w_{\max} - w_{ij})\lambda(o_i, o_j) & \text{if } \lambda(o_i, o_j) > 0 \\ \eta_{ij}w_{ij}\lambda(o_i, o_j) & \text{otherwise} \end{cases} \quad (3)$$

where w_{ij} is the synaptic weight from neuron i to neuron j, η_{ij} is the learning rate of w_{ij}, $o_i = \sigma(y_i + \theta_i)$ is the output of the ith neuron, $w_{\max}$ is the maximum weight of a synaptic connection, and $\lambda(o_i, o_j) \equiv \tanh(2 - 4|o_i - o_j|)$. This covariance rule strengthens a synapse when the difference between presynaptic and postsynaptic activity is less than 0.5, and weakens a synapse when this difference is greater than 0.5. The genetically-encoded weights w_{ij} are used as initial weights for plastic synapses. Note that the covariance rule cannot change the sign of a synapse, only its magnitude. Only synapses that interconnect two distinct neurons are plastic. Since each plastic synapse adds a parameter for learning rate η_{ij} to the genetic encoding, an N-neuron plastic CTRNN has $2N^2 + 3N$ parameters. A plastic synapse can become nonplastic if its η is set to 0.

2.4 Evolutionary Algorithm

A real-valued genetic algorithm was used to evolve CTRNN parameters. A population of 500 individuals was maintained, with each individual encoded as a vector of real numbers. Initially, a random population of vectors was generated by initializing each component of every individual to random values uniformly distributed over the range ±1 (they could move outside this range during evolution). Individuals were selected for reproduction using a linear rank-based method. A 5% elitist fraction of top individuals in the old population were simply copied to the new one. The remaining children were generated by either mutation or 2-point crossover with a crossover probability of 50%. A selected parent was mutated by adding to it a random displacement vector whose direction was uniformly distributed on the hypersphere and whose magnitude was a Gaussian random variable with 0 mean and variance of 0.5. A neuron's time constant, bias, sensor weights and input weights (and learning rates for plastic CTRNNs) were treated as a module during crossover. Unless otherwise specified, sensor weights, connection weights and biases were initialized to the range ±10, time constants were initialized to the range [1, 75], and the learning rates of plastic synapses were initialized to the range [0, 0.5].

3 Task Analysis

Despite the simplicity of our abstract food edibility learning task, our preliminary experiments demonstrated that associative learning is a surprisingly difficult behavior to evolve. In order to explain why the complex experimental design described above was necessary to reliably obtain highly fit agents, we analyzed the properties of this "simple" task in some detail.

Consider the combinatorial structure of the food edibility task. We denote the food type as either "↑" (for food producing an upward smell signal in S) or "↓" (for food producing a downward smell signal in S). The agent's action can be idealized as either mouth open (denoted "○") or mouth closed (denoted "●"). Reinforcement can be idealized as either positive (denoted "+") or negative (denoted "–"). Thus, an agent's idealized interaction history consists of a sequence of smell/action/reinforcement triples such as (↑●–)(↓●+)(↑○+)(↓●–)(↓○+). An interesting aspect of this example sequence is that the type of environment changes between the third and fourth trial from A (↑ edible, ↓ inedible) to B (↓ edible, ↑ inedible). Another interesting aspect of this sequence is that, for maximum fitness, the initial negative experience with the ↑ smell should be immediately transferred to the opposite ↓ smell in case a ↓ smell is encountered on the next trial. Note that such a sequence is cospecified by the agent and its environment. The environment determines the type of the next food item, the agent

Downloaded from http://adb.sagepub.com at PENNSYLVANIA STATE UNIV on April 16, 2008
© 2007 International Society of Adaptive Behavior. All rights reserved. Not for commercial use or unauthorized distribution.

determines an action, and then the food type, action and environment type collectively determine the reinforcement that the agent receives.

It is important to emphasize that the agent has no a priori knowledge of the structure or meaning of a sequence such as ↑●–↓●+↑○+↓●–↓○+. It merely receives a stream of sensory perturbations via S and R and is then selected for reproduction with a probability based on its overall performance on the food edibility learning task. In order to produce a successful agent the evolutionary algorithm must discover for itself that the sequence is divided into trials, that S carries information about food type and that R is a reinforcement signal, that the agent's action should be conditional on the smell, and that the relationship between smell and subsequent action should be conditional on reinforcement.

We can identify several strategy classes that might be utilized by an agent on this task, only one of which is true associative learning. Four classes in particular are worth considering. Each strategy can be described by an action map {↑→A_1,↓→A_2} that assigns to each smell the agent's corresponding action.

The first class consists of the two *fixed action* strategies

$$\begin{matrix}\{\rightarrow\bullet\} \\ \{\rightarrow\circ\}\end{matrix}$$

that always take the same action regardless of the food type. Because these strategies ignore both the smell and reinforcement signals, they only make sense for an agent in a fixed environment containing a single type of food. On fully random sequences, the best performance a fixed response strategy can obtain is 50%. However, for sequences containing more trials in which one of the actions is more appropriate than the other, the performance of these strategies can be higher.

The second class consists of the two *fixed response* strategies

$$\begin{matrix}\{\uparrow\rightarrow\bullet,\downarrow\rightarrow\circ\} \\ \{\uparrow\rightarrow\circ,\downarrow\rightarrow\bullet\}\end{matrix}$$

that produce the same response to a given food type regardless of the environment type. Because these strategies ignore the reinforcement signal, they only make sense in a fixed environment containing both types of food. On fully random sequences, the best performance a fixed response strategy can obtain is also 50%. However, for sequences containing more trials in one environment than another, the performance of one of these strategies can be higher.

Because their action maps are fixed, neither of the above strategy classes exhibits any learning. A learning agent changes its action map as necessary based on its experience: (*experience*) ⇒ {*action map*}. The simplest learning strategy is the *Law of Effect* strategy (Thorndike, 1898)

$$\begin{bmatrix}(\circ +) \Rightarrow \{\rightarrow\circ\} \\ (\bullet -) \Rightarrow \{\rightarrow\circ\} \\ (\circ -) \Rightarrow \{\rightarrow\bullet\} \\ (\bullet +) \Rightarrow \{\rightarrow\bullet\}\end{bmatrix}$$

in which the agent adopts one of the two fixed action strategies depending on how its last action was reinforced. This strategy ignores the smell signal. On fully random sequences, the best performance the Law of Effect strategy can obtain is also 50%. However, for sequences containing runs of trials in which the same action is correct, higher performances can be obtained.

Finally, we have the *associative learning* strategy

$$\begin{bmatrix}(\uparrow\circ +) \Rightarrow \{\uparrow\rightarrow\circ,\downarrow\rightarrow\bullet\} \\ (\uparrow\bullet -) \Rightarrow \{\uparrow\rightarrow\circ,\downarrow\rightarrow\bullet\} \\ (\downarrow\circ -) \Rightarrow \{\uparrow\rightarrow\circ,\downarrow\rightarrow\bullet\} \\ (\downarrow\bullet +) \Rightarrow \{\uparrow\rightarrow\circ,\downarrow\rightarrow\bullet\} \\ (\uparrow\circ -) \Rightarrow \{\uparrow\rightarrow\bullet,\downarrow\rightarrow\circ\} \\ (\uparrow\bullet +) \Rightarrow \{\uparrow\rightarrow\bullet,\downarrow\rightarrow\circ\} \\ (\downarrow\circ +) \Rightarrow \{\uparrow\rightarrow\bullet,\downarrow\rightarrow\circ\} \\ (\downarrow\bullet -) \Rightarrow \{\uparrow\rightarrow\bullet,\downarrow\rightarrow\circ\}\end{bmatrix}$$

in which the agent adopts one of two smell-sensitive action maps depending upon the pattern of reinforcement it receives. If that pattern of reinforcement changes (due to a change in environment type), then the action map will change accordingly. Only this strategy pays attention to the entire smell/action/reinforcement structure of a trial, and only this strategy can achieve 100% performance on the food edibility task in the presence of random food types and environment types.

This combinatorial analysis makes clear one obstacle to evolving associative learning: evolutionary searches can become trapped in either nonlearning or nonassociative learning strategies. Indeed, random

Downloaded from http://adb.sagepub.com at PENNSYLVANIA STATE UNIV on April 16, 2008
© 2007 International Society of Adaptive Behavior. All rights reserved. Not for commercial use or unauthorized distribution.

initial populations invariably contain one or more fixed-action and fixed-response strategies. Unless experiences are well-mixed, the fitness of these suboptimal strategies can exceed 50% and they can take over the population before associative learning has a chance to evolve. It was for this reason that we chose to expose the agent to all possible smell sequences of a given length and to randomize environment type during evolution.

A second obstacle to evolving associative learning stems from overspecialization on trial sequence length. In our preliminary experiments, we found that even when associative learning evolved, it would often fail to generalize to trial sequences longer than those on which it was evolved. In these cases, the evolved networks lose the ability to respond to changes in environment type after an initial critical period due to transients in the circuit dynamics. In order to achieve good generalization to sequences of arbitrary length, we found it necessary to expose the agent to long sequences of experiences during evolution. Note that, coupled with the need for exhaustive trial ordering described above, this can lead to long fitness evaluation times.

A final obstacle to the evolution of associative learning is the small fitness differences that often distinguish an associative learning strategy from a suboptimal one. For example, an agent that is correctly using its smell and reinforcement signals to guide its action, but is "sloppy" in the timing of its mouth actions may have lower fitness than a suboptimal strategy whose mouth actions are very accurate when they are correct. This is the reason that we use Gaussian weighting in evaluating trial error (Equation 1). In addition, the unavoidable mistakes that occur when an environment switch occurs can detract from the fitness of subsequent correct responses. This is the reason that we use trial weighting in our fitness calculation (Equation 2), so that accuracy on later trials in a sequence is weighted more heavily than on earlier ones in order to allow time for learning to actually take place without penalty. The problem of distinguishing small fitness differences is also greatly exacerbated by the long trial sequences required to obtain good generalization. It was for this reason that we employed a shaping protocol that first ensures correct responses for short trial sequences and then achieves generalization by incrementally lengthening the sequences of trials that must be correctly handled. Over 100 evolutionary experiments with different combinations of Gaussian weighting, trial weighting, and shaping demonstrated that all three were essential to the successful evolution of associative learning on this task (Phattanasri, 2002).

4 Learning Without Synaptic Plasticity

Our first set of experiments examined the ability of nonplastic CTRNNs to solve the food edibility learning task. Initially, 8 out of 12 evolutionary searches with six-neuron CTRNNs succeeded in achieving high fitness. An examination of the best evolved circuits revealed that one or two neurons were often saturated off or on, suggesting that fewer neurons were actually necessary. Thus, we next ran evolutionary searches with three or four neurons. We found that 8/10 evolutionary searches with three-neuron circuits and 7/10 searches with four-neuron circuits also achieved high fitness. In the remainder of this section, we describe the characteristics of the best evolved three-neuron circuit in some detail.

The progression of best fitness during the evolution of this circuit is shown in Figure 3. Recall that whenever the best fitness is consistently above 95%

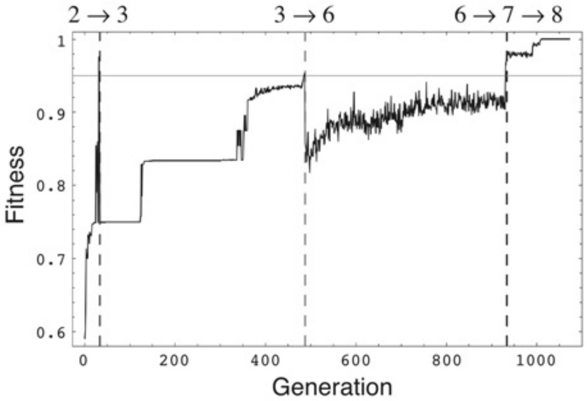


Figure 3 A plot of the best fitness versus generation for the best evolved three-neuron nonplastic circuit. Transitions between stages of our incremental shaping protocol are marked with dashed lines and labeled by the length of the trial sequences used in that stage. Transitions occur when the best fitness reliably exceeds 0.95 (horizontal gray line). Note that the fitness drops sharply after the 2→3 and 3→6 transitions before the circuit can generalize to sequences of arbitrary length. Indeed, the 6→7 and the 7→8 transitions occur so close together that they appear as a single line.

Downloaded from http://adb.sagepub.com at PENNSYLVANIA STATE UNIV on April 16, 2008
© 2007 International Society of Adaptive Behavior. All rights reserved. Not for commercial use or unauthorized distribution.

(gray horizontal line), our shaping protocol increases the length of trial sequences (dashed vertical lines). Note that following both of the first two transitions (from two-trial to three-trial sequences and from three-trial to six-trial sequences), there are sharp drops in fitness. This is because the best circuit has overspecialized on sequence length for two-trial and three-trial sequences. In contrast, there is no drop in fitness following the introduction of seven-trial sequences. This causes the immediate introduction of eight-trial sequences, after which the fitness also remains above 95%. Thus, at this point, the best circuit appears to have generalized to sequences of arbitrary length (this generalization typically occurred during the six-trial stage in all of our successful searches). There is some additional performance improvement before the end of the search.

One notable feature of the fitness graph (Figure 3) is the plateaus that are evident at several points during the search. These plateaus correspond to circuits that behave correctly on different trial subsets. For example, the first plateau during the three-trial stage occurs around a fitness of 0.75, which corresponds to correct behavior on 12 out of the 16 three-trial subsets. The second plateau around 0.84 falls between 13/16 and 14/16, and corresponds to a circuit that behaves correctly but somewhat inaccurately on 14 out of the 16 three-trial subsets. Finally, the plateau just before the three-trial to six-trial transition occurs at $0.94 \approx 15/16$. Similar plateaus were observed in all of our evolutionary searches, although the exact number and subsets of trials corresponding to each plateau varied from one search to the next. Note that the fitness plot is noisier after the transition to the six-trial stage due to the fact that the randomization of the trial at which environmental switches was introduced at this stage.

The best three-neuron circuit attained a fitness of 99.99% during evolution. In order to verify that this circuit had truly generalized, we tested it on 500 sets of all possible 10-trial sequences with a single environment switch randomly distributed between trials three and eight within each sequence. The circuit attained a fitness of 99.97% on this set of experiments. In addition, although exhaustive testing became increasingly difficult, successful tests on a subset of much longer trial sequences (up to 50 trials) indicated that this circuit did indeed represent a general solution to the food edibility learning task. We also probed the robustness of this circuit by varying the time delays between action and reinforcement and between one trial and the next outside the ranges that the circuit was exposed to during evolution. In all cases, we found this circuit to be quite robust to large variations in these delays.

The behavior of this circuit on a typical sequence of trials is shown in Figure 4. During this sequence, the environment type switches from A to B and then back to A again at the points indicated by dashed vertical lines. After each switch, the circuit initially produces an action that is incorrect for the new environment, receives negative reinforcement, and then modifies its action map to produce correct actions on subsequent trials. An interesting feature of the neural activity in this circuit is that the output of neuron 3 is nearly identical to the mouth action (neuron 1) output in environment B, but nearly a mirror image of it in environment A. This suggests that the "memory" of which environment the circuit is currently operating in is stored at least in part in the effective sign of the interaction between these two neurons. Of course, this memory cannot literally be stored in the connections between these neurons because the weights of these connections are fixed in nonplastic circuits. We will examine the dynamics of this circuit's operation in the next section.

5 Dynamical Analysis of Nonplastic Circuits

How do these nonplastic circuits work? In this section, we study how the best evolved three-neuron circuit actually accomplishes this feat. Strictly speaking, this circuit is a nonautonomous dynamical system (one that receives time-varying inputs) with two inputs S and R. Given the switch-like nature of these inputs, the best way to study the operation of such a system is to characterize its autonomous dynamics (its dynamics when inputs are fixed to particular values) for each possible combination of inputs and then to examine the transient dynamics induced by switching between these different phase portraits for typical trial sequences. If we idealize R as ± 1, then there are five possible phase portraits to consider: $\mathcal{P}_0$ (no input), $\mathcal{P}_\uparrow$ ($\uparrow$ smell), $\mathcal{P}_\downarrow$ ($\downarrow$ smell), $\mathcal{P}_+$ (positive reinforcement), and $\mathcal{P}_-$ (negative reinforcement). These five phase portraits are shown in Figure 5. These phase portraits plot the locations of all equilibrium points in the (y_1, y_2, y_3)

Downloaded from http://adb.sagepub.com at PENNSYLVANIA STATE UNIV on April 16, 2008
© 2007 International Society of Adaptive Behavior. All rights reserved. Not for commercial use or unauthorized distribution.

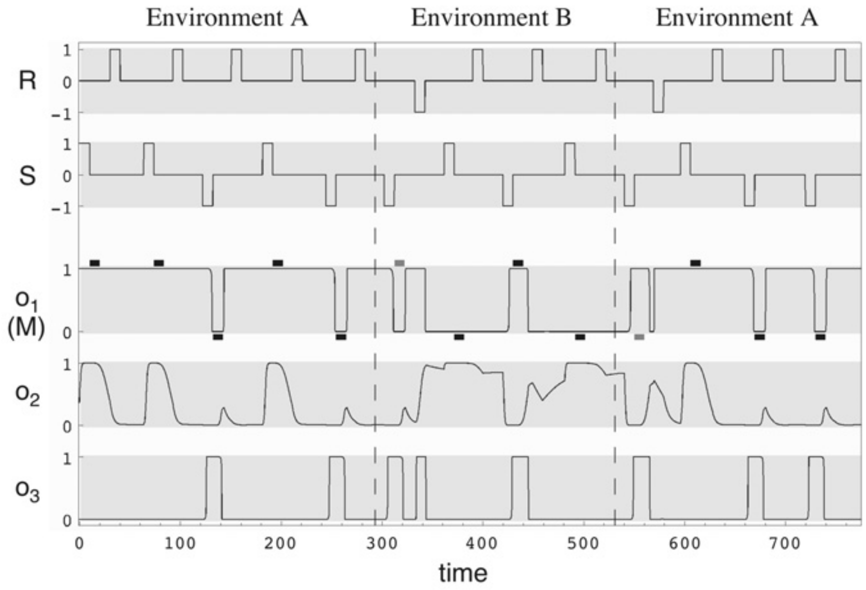


Figure 4 Activity of the best three-neuron nonplastic circuit on a typical trial sequence. From top to bottom the traces correspond to the reinforcement signal (R), the smell sensor (S), the mouth state (M, given by the output of neuron 1) and the outputs of the remaining neurons (o_i). Small rectangles mark the time during which the mouth state is evaluated and the state that the mouth should be in during this time, with correct actions denoted by black rectangles and incorrect actions denoted by gray rectangles. Transitions between environments are marked by dashed lines. Note that the circuit takes an incorrect action and receives negative reinforcement after each environment transition before modifying its action map to be appropriate to the new environment.

state space of the evolved three-neuron circuit. These equilibrium points can be stable attractors, unstable repellors, or saddle points (which have both stable and unstable directions). The three phase portraits $\mathcal{P}_0$, $\mathcal{P}_\uparrow$ and $\mathcal{P}_+$ are bistable; which attractor the state approaches will depend on the basin of attraction the circuit is in when the input changes. While the other two phase portraits $\mathcal{P}_\downarrow$ and $\mathcal{P}_-$ are unistable, the fact that they have a complicated saddle manifold structure implies that the path taken to the single attractor may vary greatly depending on where in the state space the circuit is when an input is presented.

How does the interaction of these phase portraits with the smell and reinforcement signals produce the observed behavior? Consider a three-trial sequence of ↑ smells in which the environment changes from type A to type B after the first trial. During the first trial in environment A (Top row of Figure 6), the circuit state is always attracted to the rightmost equilibrium points in $\mathcal{P}_\uparrow$, $\mathcal{P}_0$, $\mathcal{P}_+$, and $\mathcal{P}_0$ as first the ↑ smell is presented (1) and removed (2) and then the positive reinforcement is presented (3) and removed (4). When the mouth is evaluated during (2), the mouth motor neuron (neuron 1) is saturated on. This corresponds to an action of mouth open, which is the correct action for an ↑ smell in environment A. Note that the state is positioned in the basin of attraction of the *right* equilibrium point at the end of the trial (4).

The second trial begins in the same way as the first (Middle row of Figure 6), with a presentation (5) and removal (6) of an ↑ smell. However, the environment type has now changed to B and the mouth open action during (6) is no longer the correct response to an ↑ smell. The resulting negative reinforcement pulls the circuit state toward the upper left (7), leaving it in the basin of attraction of the *left* equilibrium point at the end of the trial rather than the right one (8). During the third trial (Bottom row of Figure 6), this difference in initial state leads to a different transient that produces a mouth closed action in response to an ↑ smell (10), which is then positively reinforced (11). Thus, the memory of environment type is held by which basin of attraction of $\mathcal{P}_0$ the circuit is operating in. Analogous dynamics underlie the circuit's response to

Downloaded from http://adb.sagepub.com at PENNSYLVANIA STATE UNIV on April 16, 2008
© 2007 International Society of Adaptive Behavior. All rights reserved. Not for commercial use or unauthorized distribution.

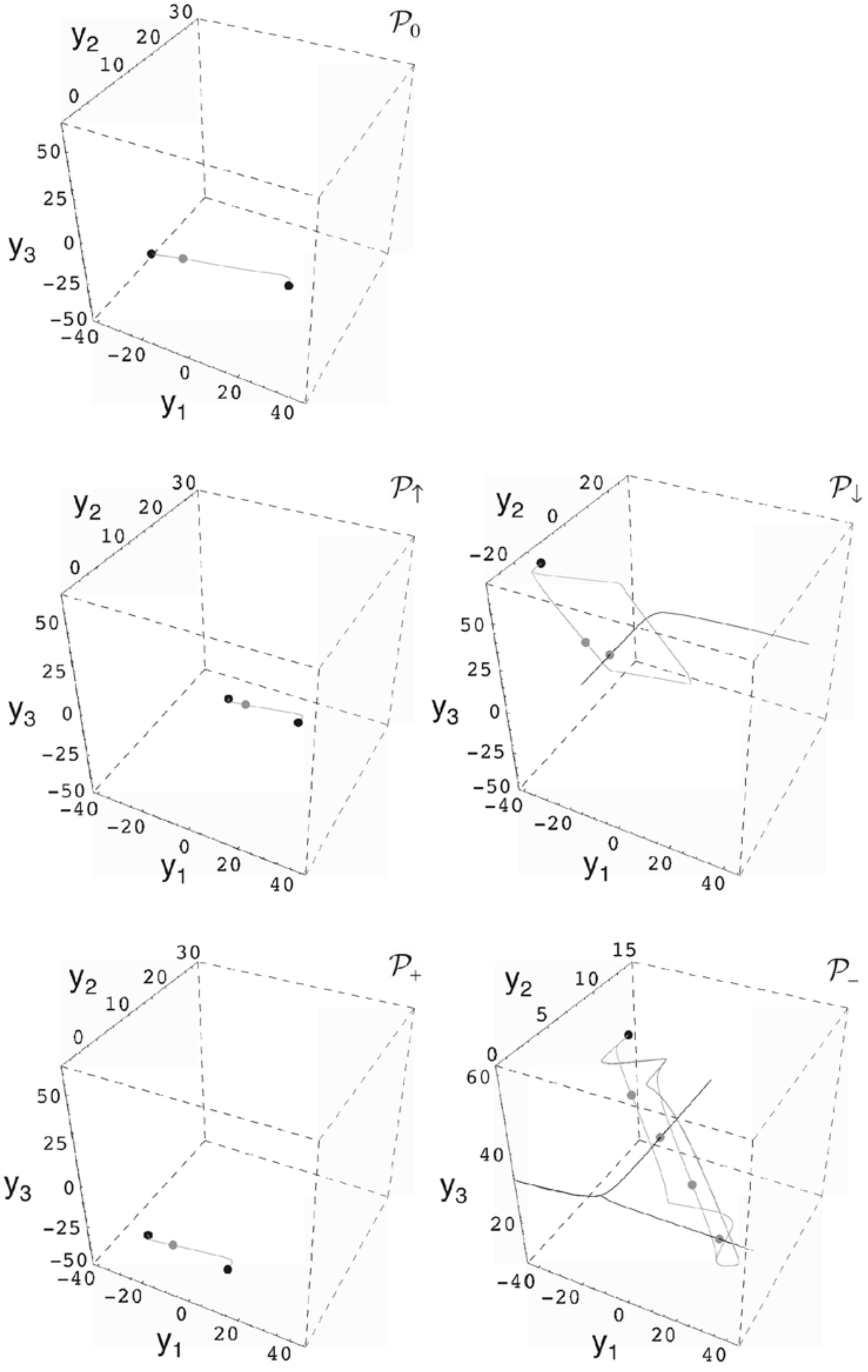


Figure 5 Phase portraits of the autonomous dynamics of the best three-neuron nonplastic circuit for no inputs ($\mathcal{P}_0$), an ↑ smell ($\mathcal{P}_\uparrow$), a ↓ smell ($\mathcal{P}_\downarrow$), positive reinforcement ($\mathcal{P}_+$) and negative reinforcement ($\mathcal{P}_-$). Stable and saddle equilibrium points are denoted by black and gray dots, respectively. The 1-dimensional stable and unstable manifolds of the saddle points are denoted by black and gray lines, respectively (two-dimensional saddle manifolds are not shown).

Downloaded from http://adb.sagepub.com at PENNSYLVANIA STATE UNIV on April 16, 2008
© 2007 International Society of Adaptive Behavior. All rights reserved. Not for commercial use or unauthorized distribution.

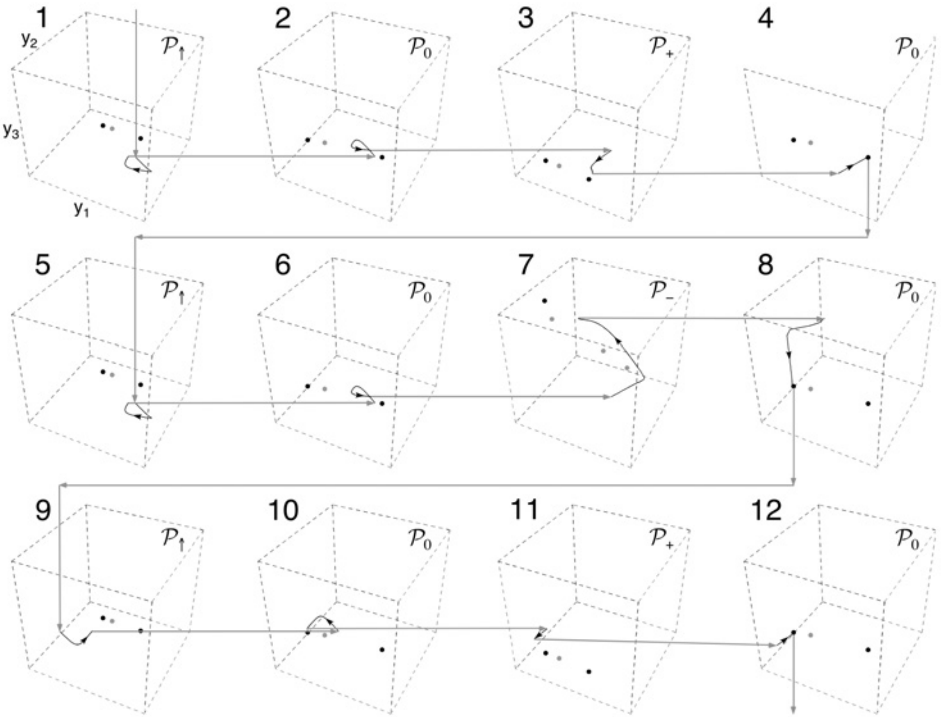


Figure 6 The nonautonomous dynamics of the best three-neuron nonplastic circuit during the trial sequence ↑○+↑○–↑●+, which switches from environment A to environment B after the first trial. As the input signals change throughout this sequence, the circuit's autonomous dynamics is switched between the different phase portraits shown in Figure 5 and its state is attracted to the stable equilibrium point in whose basin it finds itself at each point. The change in action map from environment A to environment B is accomplished by shifting the circuit's operating region from the neighborhood of the right stable equilibrium point of $\mathcal{P}_0$ to the neighborhood of the left stable equilibrium point of this phase portrait.

↓ smells, to mixed sequences of ↑ and ↓ smells, and to transitions from type B environments back to type A (Phattanasri, 2002).

How can we visualize the overall structure of this circuit's operation? Since the state often does not even reach an attractor before the input changes, transient responses clearly play a key role in this circuit. However, the details are complex and will of course vary with both the exact sequence of ↑ and ↓ smells and the random delays before reinforcement and between trials. One approach to visualization is to "strobe" the state of the system at selected times during a trial. In particular, observing the system state at the end of each smell and reinforcement signal over many trials reveals that these strobe states fall into relatively distinct clusters, which can be labeled by the signal they follow and the environment type (Figure 7a). An extended behavioral sequence such as the one shown in Figure 4 can then be understood as a set of trajectories between these strobe states (Figure 7b).

In fact, this strobed circuit dynamics can be interpreted as implementing a finite state machine (FSM) with input, with the strobe states corresponding to the FSM states and the trajectories between strobe states corresponding to input-driven transitions of the FSM (Casey, 1996). The FSM extracted from this circuit is shown in Figure 7c. It correctly classifies trial sequences into environment type and generates the correct motor actions for each food type in each environment type. Note that each FSM state actually encompasses several smaller subgroups of strobe states. These individual groups correspond to different paths to that state. For example, the left subgroup of A_4 in Figure 7a corresponds to receiving negative reinforcement following a ↓ smell (i.e., a transition from B_2), while the right subgroup corresponds to

Downloaded from http://adb.sagepub.com at PENNSYLVANIA STATE UNIV on April 16, 2008
© 2007 International Society of Adaptive Behavior. All rights reserved. Not for commercial use or unauthorized distribution.

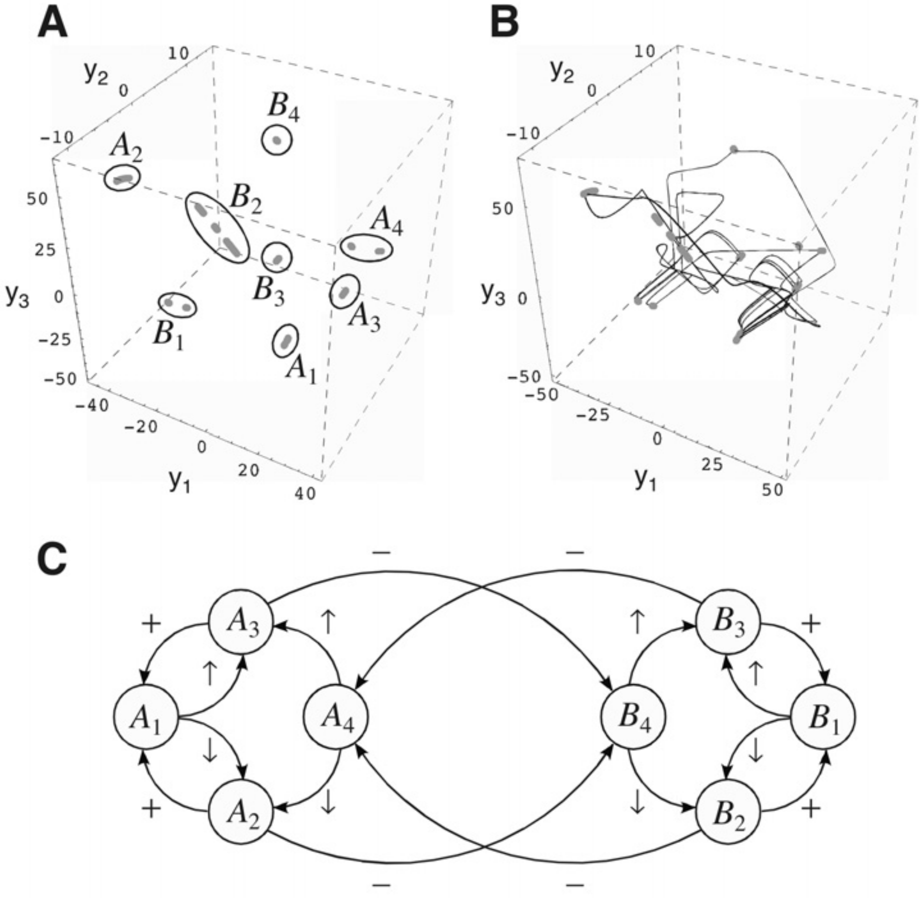


Figure 7 A finite state machine embedded in the best three-neuron nonplastic circuit. (a) The strobe states obtained by sampling the circuit state at the end of each smell and reinforcement signal. Note that these strobe states fall into distinct clusters. (b) Trajectories between the strobe states during the trial sequence shown in Figure 4. (c) The extracted finite state machine, where each state corresponds to a cluster of strobe states from part (a) and the state labels "A" and "B" refer to environment A and environment B, respectively. The state subscripts are defined as follows: 1 denotes the state of the FSM after receiving positive reinforcement, 2 denotes the state of the FSM after receiving an ↓ smell, 3 denotes the state of the FSM after receiving an ↑ smell, and 4 denotes the state of the FSM after receiving negative reinforcement.

receiving negative reinforcement following an ↑ smell (i.e., a transition from B_3). Thus, an FSM of finer granularity could also be extracted if desired. Note also that the extracted FSM is not minimal, since the two states A_4 and B_4 could in principle be collapsed onto A_1 and B_1, respectively. These extra states apparently arise from transients in the dynamics when the environment changes from one type to the other, which take one additional trial to settle out. We found that all successful nonplastic controllers we evolved implemented finite state machines in an analogous manner. Furthermore, although different evolutionary searches produced circuits that implemented different FSMs, they were all reducible to the minimal FSM for this task.

It is important to emphasize that the extracted FSMs merely summarize the normal operation of the circuit dynamics, and are not equivalent to this dynamics. For example, the strobe states do not necessarily represent attractors of the circuit dynamics; they describe only the typical range of states that the circuit is found in at key points in the interaction. Thus, a given circuit may only be able to "hold" some states for a limited amount of time. In addition, although the

Downloaded from http://adb.sagepub.com at PENNSYLVANIA STATE UNIV on April 16, 2008
© 2007 International Society of Adaptive Behavior. All rights reserved. Not for commercial use or unauthorized distribution.

overall operation of the circuit is robust to small perturbations to the strobe states, the extracted FSMs say nothing about how the circuit will respond to more general perturbations. Nevertheless, this analysis demonstrates that evolution has shaped the overall dynamics of the successful circuits to match the combinatorial structure of the food edibility learning task.

6 Learning with Synaptic Plasticity

In a second set of experiments, we examined the ability of plastic CTRNNs using the covariance learning rule (Equation 3) to solve the food edibility learning task. Surprisingly, all four-neuron and five-neuron experiments failed and only 2/15 evolutionary searches with six-neuron plastic circuits succeeded. The best of the other 13 searches never advanced beyond the six-trial stage, and many never even solved the three-trial stage. One might think that one solution would be to simply set all learning rates to 0 and use the same strategy as in the nonplastic CTRNN experiments. However, this never occurred.

The progression of best fitness during evolution of the best six-neuron circuit is shown in Figure 8. The pattern is very similar to that observed in the nonplastic searches (compare to Figure 3). The two-trial sequences are solved relatively quickly, but fitness drops sharply following the introduction of the three-trial sequences, indicating overspecialization. Plateaus are observed during evolution on three-trial sequences, and fitness once again drops after six-trial sequences are introduced. Once the circuit has successfully evolved to handle six-trial sequences, the transitions to seven-trial and eight-trial sequences occur rapidly, indicating successful generalization. Interestingly, this circuit took only about half the number of generations to evolve as did the best nonplastic circuit. However, the number of successful plastic searches is too small to determine whether or not this difference is statistically significant.

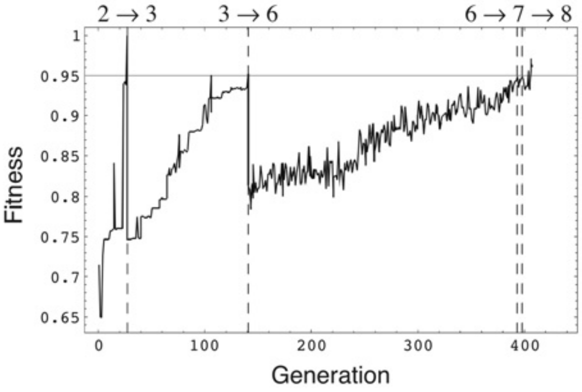


Figure 8 A plot of the best fitness versus generation for the best evolved six-neuron plastic circuit. Labeling conventions are the same as in Figure 3.

The best plastic six-neuron circuit achieved a fitness of 96% during evolution, which is somewhat lower than that obtained by the best nonplastic circuit. In a test of 500 sets of all possible 10-trial sequences this circuit attained a fitness of 95.4%, demonstrating that its performance generalizes to longer trial sequences than it was exposed to during evolution. We also found that this circuit was robust to variations in the delay between action and reinforcement and between one trial and the next outside the ranges that the circuit was exposed to during evolution.

The behavior of this circuit on a typical sequence of trials is shown in Figure 9. This is the same sequence that was earlier used to illustrate the behavior of the best nonplastic circuit (Figure 4), in which the environment type switches from A to B and then back to A again. As for the best nonplastic circuit, this circuit initially produces an incorrect action and receives negative reinforcement after each environment switch before modifying its action map to produce the correct responses. The outputs of the six neurons during the course of this trial are also shown in Figure 9. An interesting feature of this circuit is that the output of neuron 3 is nearly identical to the mouth action (neuron 1) in environment A, but is largely unrelated to the mouth action in environment B. Conversely, the output of neuron 4 is similar to the mouth action in environment B, but is unrelated in environment A.

The changes in magnitude of selected weights during the course of these trials are shown at the bottom of Figure 9. One striking feature of these plots is that the weights change very quickly. Indeed, the average value of the evolved learning rates in this circuit was 0.23 ± 0.12 (mean ± s.d.), with a minimum value of 0.06. These learning rates are too fast to properly integrate information across an entire trial, let alone multiple trials. Thus, despite the fact that Hebbian learning is available, this circuit solves the food edibility learning task using synaptic plasticity in a way that differs from the traditional view illustrated in

Downloaded from http://adb.sagepub.com at PENNSYLVANIA STATE UNIV on April 16, 2008
© 2007 International Society of Adaptive Behavior. All rights reserved. Not for commercial use or unauthorized distribution.

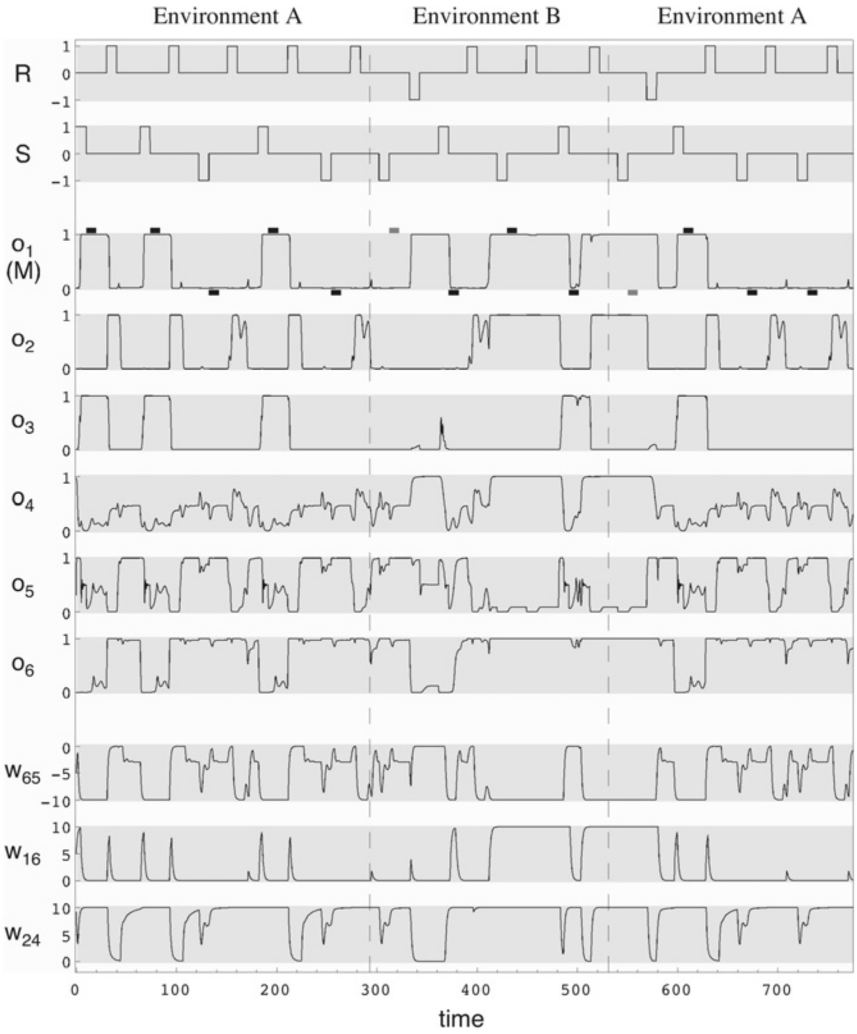


Figure 9 Activity of the best six-neuron plastic circuit on the same trial sequence shown in Figure 4. From top to bottom the traces correspond to the reinforcement signal (R), the smell sensor (S), the mouth state (M, given by the output of neuron 1), the outputs of the remaining neurons (o_i), and weight changes of three example connections (w_{ij}). Labeling conventions are the same as in Figure 4.

Figure 1a; the weight variables act more like additional *neuronal* degrees of freedom than they do traditional learning processes in the sense that the synaptic changes are occurring on the same timescale as the changes in neural activity. Similar results were reported by Floreano and Mondada (1996).

In an attempt to force a more traditional use of synaptic plasticity, we reduced the allowable range of learning rates from [0, 0.5] to [0, 0.03]. A number of experiments were run under these conditions, varying the number of neurons, the subset of connections that were plastic, and even the synaptic plasticity rule. In no case were we able to evolve completely successful circuits with these slower learning rates. Although the best agent achieved a fitness of 0.9 after 1000 generations, it was only able to handle environmental transitions in one direction. For example, it could perform correctly in environment A (resp. B) and then correctly make a transition to environment B (resp. A), but it was unable to subsequently transition back to environment A (resp. B). Thus, this circuit exhibited an extended critical period rather than general and flexible food edibility learning (cf. Miller & Todd, 1991; Yamauchi & Beer, 1994a).

Downloaded from http://adb.sagepub.com at PENNSYLVANIA STATE UNIV on April 16, 2008
© 2007 International Society of Adaptive Behavior. All rights reserved. Not for commercial use or unauthorized distribution.

7 Discussion

7.1 Summary

In this article, we have demonstrated that CTRNNs can be evolved to learn to associate the smell of food with its edibility based upon the circuit's experiences in an environment. Despite the fact that the basic CTRNN model lacks synaptic plasticity, an evolutionary algorithm was able to shape the network dynamics so that the agent could both generate bites in response to edible food while ignoring inedible food and learn which food was edible through the reinforcement it received from its interactions with its environment. In the best evolved circuits, this learning ability generalized to much longer trial sequences than they were evolved on. These circuits also maintained their ability to relearn throughout their operation, appropriately adjusting their responses to the two food types as they were switched between the two environment types. We also showed that these nonplastic circuits work by implementing finite state machines that capture the sensation-action-reinforcement structure of this task. Finally, we demonstrated that successful CTRNNs with Hebbian synaptic plasticity can also be evolved, but that such circuits do not use their synaptic plasticity in a traditional way.

7.2 The Neuronal Mechanisms of Learning

It may seem surprising at first that associative learning is even possible without explicit synaptic plasticity. However, learning is first and foremost a behavioral phenomenon. An animal is typically said to be learning when its behavior on some task improves over time as a result of its interactions with its environment. The actual mechanisms underlying this improvement in any particular instance must be elucidated by empirical investigation. The theoretical possibility of learning without synaptic plasticity follows immediately from the universal dynamics approximation capabilities of CTRNNs (Funahashi & Nakamura, 1993; Kimura & Nakano, 1998). Indeed, we have previously demonstrated such behavior (Yamauchi & Beer, 1994a, 1994b), and it has also been explored by several others (Blynel & Floreano, 2003; Cotter & Conwell, 1990; Feldkamp, Puskorius, & Moore, 1996; Izquierdo-Torres & Harvey, 2006; Tuci, Quinn, & Harvey, 2002; Younger, Conwell, & Cotter, 1999). Thus, although the traditional architecture sketched in Figure 1a is a possible organization of these mechanisms, it is by no means the only one.

Biologically, there is also reason to believe that at least some aspects of learning and memory are due not only to synaptic plasticity, but also neuronal dynamics. The intrinsic conductances of neurons endow them with the ability to show persistent changes even in the absence of changes in their synaptic connections (Marder *et al.* 1996). Similarly, long-term changes in firing levels have been observed in many regions of the brain and are often associated with memory processes (e.g., responses to delayed alternation tasks). Although the exact mechanisms of persistent activity are not yet clear, they almost certainly are due to a combination of intrinsic properties and recurrent connections among neurons (Major & Tank, 2004). These results suggest that, despite the fact that synaptic plasticity clearly plays an important role in biological learning, the strong tendency to equate every instance of learning behavior with synaptic plasticity must be resisted.

We have also demonstrated that, even when synaptic plasticity is made available, it is not necessarily utilized in a way consistent with the traditional view illustrated in Figure 1a. Rather than integrate information across a trial or multiple trials, our successful plastic circuits appear to utilize plastic synapses simply as additional neuronal degrees of freedom. This aligns well with the growing realization that the short-term dynamics of synapses plays as important a role in nervous system function as the dynamics of neurons (Abbott & Regehr, 2004). Indeed, recent studies of spike timing dependent plasticity have demonstrated that timing differences as short as 20 milliseconds can profoundly alter the sign of plasticity from potentiation to depression (Dan & Poo, 2006). Interestingly, we found that, although successful plastic circuits evolved food edibility learning faster than successful nonplastic circuits, the success rate for plastic circuits was considerably lower than for nonplastic circuits. Furthermore, we were unable to evolve successful circuits if we forced synaptic plasticity to occur on timescales significantly longer than the neuronal timescale. Given the small number of experiments, it is difficult to know how to interpret these results. About the most that can probably be said is that further work is required to properly assess the relative tradeoffs between nonplastic and plastic circuits for this task.

Downloaded from http://adb.sagepub.com at PENNSYLVANIA STATE UNIV on April 16, 2008
© 2007 International Society of Adaptive Behavior. All rights reserved. Not for commercial use or unauthorized distribution.

Even if we accept the traditional view of synaptic plasticity, our results raise an even more fundamental question: How are the local synaptic changes underlying learning integrated into the mechanisms that coordinate behavior (Destexhe & Marder, 2004)? The coordination of the behavior of an entire animal is the responsibility of its nervous system as a whole. When an animal's behavior in nearly identical circumstances changes with experience, there is no question that something inside the animal changes as well. However, these changes can be highly distributed across the nervous system, making it difficult to assess the significance of any one change. For example, even in the habituation or sensitization of the gill withdrawal reflex in *Aplysia*, a paradigmatic illustration of learning due to changes in a single synapse, it was subsequently found that the activity of hundreds of neurons in the abdominal ganglion actually changed (Zečević, Wu, Cohen, London, Höpp, & Falk, 1989), suggesting that experience-dependent changes in the gill withdrawal reflex may involve much more than the plasticity of a single synapse. As a second example, Lockery and Sejnowski (1993) found that retraining a neural network model of the leech bending reflex to habituated or sensitized states produced very small changes in every synaptic weight in the network, a situation that would be extremely difficult to detect physiologically. Thus, understanding how local changes impact global dynamics is a very important open question in the neurobiology of learning (Münte, Altenmüller, & Jäncke, 2002).

Of course, despite their highly distributed organization, nervous systems are not architecturally uniform. Brain specializations related to learning are well-known. For example, lesion, brain imaging, and clinical correlation studies have conclusively shown that the hippocampus and cerebellum play crucial roles in mammalian learning (Kandel, Kupfermann, & Iversen, 2000). However, animals that lack such structures (e.g., octopi; Bullock & Horridge, 1965) are also capable of sophisticated forms of learning (Boal, Dunham, Williams, & Hanlon, 2000). Moreover, even in animals with these specializations, the relationship between activity within these regions and the rest of the brain during learning has turned out to be far more complex than first assumed. For example, recent brain imaging studies of activity in anterior cingulate, dorsolateral prefrontal cortex and hippocampus have shown that increases or decreases in activity in these regions are not as important as the neural context within which the activity occurs (McIntosh, 1999; Kelly, & Garavan, 2005). Thus, it is important to remember that behavioral plasticity is a systemic property of an entire animal, whose explanation is likely to involve not only functional specializations particular to learning, but also how those specializations are integrated into the rest of the nervous system.

7.3 The Difficulty of Evolving Learning

Behavioral plasticity is ubiquitous in the biological world. One can even argue that the evolution of learning is in some sense inevitable in a dynamical agent (Beer, 1997). Suppose that an agent has internal dynamics on a range of timescales and let us focus on the dynamics at timescales that are long relative to the timescale of the agent's actions. This long timescale dynamics can be deleterious, neutral, or advantageous. If, however, the agent is subjected to a selection process in which the long timescale dynamics is under evolutionary control, then any deleterious long timescale dynamics will be selected away. Furthermore, if there is any advantage to be gained from the longer timescale dynamics (e.g., if the environment itself is changing over longer timescales), then we would expect agents that exploit this advantage to proliferate over those that do not. In this sense, the improvement of behavior over time is an inevitable property of agents with dynamics on a range of timescales subjected to selection in a changing environment.

Yet we were able to evolve circuits capable of food edibility learning only by employing carefully crafted fitness weighting and shaping protocols. Why was learning so difficult to evolve? We believe that the difficulties we encountered stemmed not from the fact that we were trying to evolve learning per se, nor from the fact that only nonplastic synapses were available in some experiments. Rather, we believe that it was the discrete combinatorial nature of the food edibility learning task that made circuits so difficult to evolve. Nowhere is the combinatorial nature of this task more apparent than in the fact that the successful circuits implemented finite state machines. But inducing finite state machines from examples of a regular language is a very difficult problem in general, and an evolutionary approach is almost certainly not the most efficient procedure. More importantly, most animal learning does not have such a rigid combinatorial

Downloaded from http://adb.sagepub.com at PENNSYLVANIA STATE UNIV on April 16, 2008
© 2007 International Society of Adaptive Behavior. All rights reserved. Not for commercial use or unauthorized distribution.

character, except perhaps in rather artificial laboratory settings. This suggests that the evolution of learning should be explored in more natural ecological contexts. For example, one could imagine a version of the food edibility learning task with agents under energy constraints actually moving through an environment containing food of different types whose edibility varies from location to location.

7.4 Future Directions

There are a number of directions in which the work described in this article could be developed. First, we need both to better understand the operation of the plastic circuits we have evolved and to run a larger set of evolutionary experiments with systematic variations in learning rules, learning rates, and the distribution of plastic synapses within a circuit. Second, the food edibility task could be extended in various ways. For example, by allowing both food types to be edible or inedible, we increase the set of possible environments to four and require at least two experiences before appropriate behavior can be determined. In addition, we could provide a reinforcement signal only when a bite is actually taken (making it more like a gut sensor) and use the total energy intake as a fitness measure (with the consumption of inedible food counting negatively). This would be a first step toward a more ecological version of the food edibility learning task. Another interesting possibility would be to first evolve an agent on some fixed task and then subsequently introduce variability in the task that requires learning while continuing to evolve the agent. Finally, it would be interesting to systematically study the conditions under which various kinds of learning do or do not evolve in nonplastic CTRNNs as the spatiotemporal structure of the environment is changed. On a foraging task, for example, one would expect fixed behavior to evolve when the environment is completely predictable, reactive behavior to evolve when the environment has no long-term predictability, and various kinds of learning behavior to evolve when the structure of the environment varies over time in some systematic way.

7.5 What Is Learning?

There is one further issue that evolution and analysis of the sort of models we have described here might clarify. It might be objected that choosing between two possible action maps based upon a single experience in an environment is hardly deserving of the label "associative learning." There is no question that, by design, our version of the food edibility learning problem is highly simplified. But how many choices are enough to constitute learning? Many textbook examples of associative learning involve only a small number of choices. If there were three types of food, each of which could be either edible or inedible, the agent would require reinforcement from a sequence of three experiences in order to choose between the eight possible action maps. Are eight choices enough? One hundred?

More generally, which changes in behavior are deserving of the label "learning" and which are not? This is a very difficult question to answer. *The Oxford Companion to Animal Behavior* (McFarland, 1981) states

> Learning is a familiar enough phenomenon, but as is often the way, one not easily captured by the scientist's definition. … The variety is such that it may be difficult or impossible to formulate a definition of the conditions producing learning that is not either vacuous or so restrictive that it excludes certain cases which we should certainly want to regard as instances of learning. … It may be foolish, therefore, to waste too much time attempting to provide a precise, all-embracing definition of learning at this point. (pp. 336–337)

For example, recent work by Izquierdo-Torres and Harvey (2006) is aimed specifically at extending the approach we have described here to an imprinting task with a continuum of possibilities. However, imprinting is an irreversible change that occurs during a critical developmental period. Is this learning? Or consider CTRNNs that we previously evolved for generating walking in a legged agent (Beer & Gallagher, 1992). Depending on the initial orientation of the legs, it could take several steps before the dynamical transients of the coupled neuromechanical system settled out. During this transient, the locomotion performance steadily improved with each step as a direct result of the sensory feedback. Is this learning? These circuits were also able to adjust their motor patterns to the changing geometry of a growing body through entrainment between neural oscillations and the rhythmic sensory feedback from a stepping leg. Is this learning?

Downloaded from http://adb.sagepub.com at PENNSYLVANIA STATE UNIV on April 16, 2008
© 2007 International Society of Adaptive Behavior. All rights reserved. Not for commercial use or unauthorized distribution.

While it is relatively straightforward to recognize learning in textbook tasks with discrete sensation, action and reinforcement signals, it is surprisingly difficult to articulate a sharp behavioral definition of learning for freely-behaving animals.

Combined with our demonstration that multi-timescale dynamics can exhibit learning-like behavior regardless of whether or not explicit synaptic plasticity is available, the difficulty of defining learning behavior or isolating its neuronal properties suggests a rather radical possibility: Perhaps learning is not actually a natural kind at either the neuronal or the behavioral level. Rather, an agent whose internal dynamics has been shaped by evolution to survive in a dynamical environment may necessarily exhibit changes in its behavior on a continuum of timescales from milliseconds to its entire lifetime. If this is correct, then studies of learning may be better served by subsuming learning and its neuronal mechanisms into a more general study of the dynamics of behavior rather than trying to impose artificial boundaries between learning and nonlearning at either the behavioral or neuronal levels. The relatively unbiased nature of evolutionary approaches seems ideal for exploring the many possible ways in which the mechanisms of learning can be integrated into the mechanisms of behavior.

Acknowledgments

This research was supported in part by DMI-9720309 from NSF.

References

Abbott, L. F., & Regehr, W. G. (2004). Synaptic computation. *Nature 431*: 796–803.

Baxter, D. A., & Byrne, J. H. (1993). Learning rules from neurobiology. In D. Gardner (Ed.), *The neurobiology of neural networks* (pp. 71–105). Cambridge, MA: MIT Press.

Beer, R. D. (1997). The dynamics of adaptive behavior: A research program. *Robotics and Autonomous Systems 20*: 257–289.

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

Bi, G-Q., & Rubin, J. (2005). Timing in synaptic plasticity: From detection to integration. *Trends in Neurosciences 28*: 222–228.

Blynel, J., & Floreano, D. (2003). Exploring the T-maze: Evolving learning-like robot behaviors using CTRNNs. In C. Ryan, T. Soule, M. Keijzer, E. Tsang, R. Poli, & E. Costa (Eds.), *Applications of evolutionary computing* (Lecture Notes in Computer Science 2611) (pp. 593–604). Berlin: Springer.

Boal, J. G., Dunham, A. W., Williams, K. T., & Hanlon, R. T. (2000). Experimental evidence for spatial learning in Octopuses (*Octopus bimaculoides*). *J. Comp. Psych. 114*: 246–252.

Bullock, T. H., & Horridge, G. A. (1965). *Structure and function in the nervous systems of invertebrates*. W.H. Freeman.

Casey, M. (1996). The dynamics of discrete-time computation, with application to recurrent neural networks and finite state machine extraction. *Neural Computation 8*: 1135–1178.

Chalmers, D. J. (1991). The evolution of learning: An experiment in genetic connectionism. In D. S. Touretzky, J. L. Elman, T. J. Sejnowski & G. E. Hinton (Eds.), *Connectionist models: Proceedings of the 1990 Summer School* (pp. 81–90). San Mateo: Morgan Kaufmann.

Chiel, H. J., & Susswein, A. J. (1993). Learning that food is inedible in freely-behaving *Aplysia californica*, *Behavioral Neuroscience 107*: 327–338.

Cotter, N. E., & Conwell, P. R. (1990). Fixed-weight networks can learn. In *Proceedings of the International Joint. Conference on Neural Networks* (pp. II553–II559). IEEE Press.

Dan, Y., & Poo, M-M. (2004). Spike timing-dependent plasticity of neural circuits. *Neuron 44*: 23–30.

Dan, Y., & Poo, M-M. (2006). Spike timing-dependent plasticity: From synapse to perception. *Physiological Review 86*: 1033–1048.

Destexhe, A., & Marder, E. (2004). Plasticity in single neuron and circuit computations. *Nature 431*: 789–795.

Feldkamp, L.A., Puskorius, G.V., & Moore, P.C. (1996). Adaptation from fixed weight dynamic networks. In *Proceedings of the IEEE International Conference on Neural Networks* (pp. 155–160). IEEE Press.

Floreano, D., & Mondada, F. (1996). Evolution of plastic neurocontrollers for situated agents. In P. Maes, M. Mataric, J. Meyer, J. Pollack, & S. Wilson (Eds.), *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior* (pp. 402–410). Cambridge, MA: MIT Press.

Floreano, D., & Urzelai, J. (2001). Evolution of plastic control networks. *Autonomous Robotics 11*: 311–317.

Funahashi, K. I., & Nakamura, Y. (1993). Approximation of dynamical systems by continuous time recurrent neural networks. *Neural Networks 6*: 801–806.

Izquierdo-Torres, E., & Harvey, I. (2006). Learning on a continuum in evolved dynamical node networks. In L. Rocha et al. (Eds.), *Artificial Life X: Proceedings of the Tenth International Conference on the Simulation and Synthesis*

Downloaded from http://adb.sagepub.com at PENNSYLVANIA STATE UNIV on April 16, 2008
© 2007 International Society of Adaptive Behavior. All rights reserved. Not for commercial use or unauthorized distribution.

of Living Systems (pp. 507–512). Cambridge, MA: MIT Press.

Kandel, E. R., Kupfermann, I., & Iverson, S. (2000). Learning and memory. In E. R. Kandel, J. H. Schwartz, & T. M. Jessell (Eds.), *Principles of neural science*, 4th ed. (pp. 1227–1246). McGraw-Hill.

Kandel, E. R., & Siegelbaum, S.A. (2000). Transmitter release. In E. R. Kandel, J. H. Schwartz, & T. M. Jessell (Eds.), *Principles of neural science*, 4th ed. (pp. 253–279). McGraw-Hill.

Kelly, A. M. C., & Garavan, H. (2005). Human functional neuroimaging of brain changes associated with practice. *Cerebral Cortex 15*: 1089–1102.

Kimura, M., & Nakano, R. (1998). Learning dynamical systems by recurrent neural networks from orbits. *Neural Networks 11*: 1589–1599.

Llinas, R. R. (1988). The intrinsic electrophysiological properties of mammalian neurons: Insights into central nervous system function*. Science*. *242*: 1654–1664.

Lockery, S. R., & Sejnowski, T. J. (1993). Voyages through weight space: Network models of an escape reflex in the leech. In R. D. Beer, R. E. Ritzmann & T. McKenna (Eds.), *Biological neural networks in invertebrate neuroethology and robotics* (pp. 251–266). San Diego: Academic Press.

Major, G., & Tank, D. (2004). Persistent neural activity: prevalence and mechanisms. *Current Opinion in Neurobiology 14*: 675–684.

Marder, E., Abbott, L. F., Turrigiano, G. G., Liu, Z., & Golowasch, J. (1996). Memory from the dynamics of intrinsic membrane currents. *Proc. Nat. Acad. Sci. USA 93*: 13481–13486.

Margulies, C., Tully, T., & Dubnau, J. (2005). Deconstructing memory in *Drosophila*. *Current Biology 15*: R700–R713.

McFarland, D. (Ed.) (1981). *The Oxford companion to animal behavior*. Oxford University Press.

McIntosh, A. R. (1999). Mapping cognition to the brain through neural interactions. *Memory 7*: 523–548.

Miller, G. F., & Todd, P. M. (1991). Exploring adaptive agency: I. Theory and methods for simulating the evolution of learning. In D. S. Touretzky, J. L. Elman, T. J. Sejnowski & G. E. Hinton (Eds.), *Connectionist models: Proceedings of the 1990 summer school* (pp. 65–80). San Mateo: Morgan Kaufmann.

Münte, T. F., Altenmüller, E., & Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nature Reviews Neuroscience 3*: 473–478.

Phattanasri, P. (2002). *Associative learning in evolved dynamical neural networks*. Ph.D. Dissertation, Department of Electrical Engineering and Computer Science, Case Western Reserve University.

Susswein, A. J., Schwarz, M., & Feldman, E. (1986). Learned changes of feeding behavior in *Aplysia* in response to edible and inedible foods. *Journal of Neuroscience 6*: 1513–1527

Sutton, M. A., & Carew, T. J. (2002). Behavioral, cellular and molecular analysis of memory in *Aplysia* I: Intermediate-term memory. *Integrative and Comparative Biology 42*: 725–735.

Thorndike, E.L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *The Psychological Review, Series of Monographs Supplements* (Supplement 2) *4*: 1–109.

Todd, P.M., & Miller, G.F. (1991a). Exploring adaptive agency II: Simulating the evolution of associative learning. In J.A. Meyer & S.A. Wilson (Eds.), *From Animals to Animats 1: Proceedings of the First International Conference on Simulation of Adaptive Behavior* (pp. 306–315). Cambridge, MA: MIT Press.

Todd, P. M., & Miller, G. F. (1991b). Exploring adaptive agency III: Simulating the evolution of habituation and sensitization. In H. P. Schwefel & R. Manner (Eds.), *Proceedings of the First International Conference on Parallel Problem Solving From Nature* (pp. 307–313). Berlin: Springer.

Toledo-Rodriguez, M., El Manira, A., Wallen, P., Svirskis, G., & Hounsgaard, J. (2005). Cellular signaling properties in microcircuits. *Trends in Neurosciences 28*: 534–540.

Trewavas, A. (2003). Aspects of plant intelligence. *Annals of Botany 92*: 1–20.

Tuci, E., Quinn, M., & Harvey, I. (2002). An evolutionary ecological approach to the study of learning behavior using a robot-based model. *Adaptive Behavior 10*: 201–221.

Wood, D. C. (1969). Parametric studies of the response decrement produced by mechanical stimuli in the protozoan *Stentor coeruleus*. *J. Neurobiology 1*: 345–360.

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

Yamauchi, B., & Beer, R. D. (1994b). Integrating reactive, sequential and learning behavior using dynamical neural networks. In D. Cliff, P. Husbands, J. Meyer & S. Wilson (Eds.), *From Animals to Animats 3: Proceedings of the Third International Conference on Simulation of Adaptive Behavior* (pp. 382–391). Cambridge, MA: MIT Press.

Younger, A. S., Conwell, P. R., & Cotter, N. E. (1999). Fixed-weight on-line learning. *IEEE Transactions in Neural Networks 10*: 272–283.

Zečević, D. Wu, J-Y., Cohen, L. B., London, J. A., Höpp, H-P., & Falk, C. X. (1989). Hundreds of neurons in the *Aplysia* abdominal ganglion are active during the gill-withdrawal reflex. *J. Neuroscience 9*: 3681–3689.

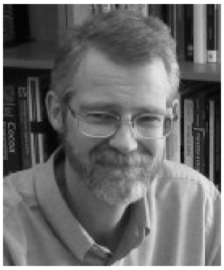
Zhang, Y., Lu, H., & Bargmann, C. I. (2005). Pathogenic bacteria induce averse olfactory learning in *Caenorhabditis elegans*. *Nature 438*: 179–184.

Downloaded from http://adb.sagepub.com at PENNSYLVANIA STATE UNIV on April 16, 2008
© 2007 International Society of Adaptive Behavior. All rights reserved. Not for commercial use or unauthorized distribution.

About the Authors

Phattanard Phattanasri received his M.S. and Ph.D. in system and control engineering from Case Western Reserve University in 1998 and 2002, respectively. Recently, he was a research associate in the Department of Electrical Engineering and Computer Science at Case. His research interests are computational modeling of learning, nonlinear adaptive control system and optimization of discrete event systems. He currently works at Fabrinet, Co. Ltd, Thailand in the field of optoelectronics manufacturing. *Address:* 54 Setsiri Rd, Phayathai, Bangkok, Thailand 10400. *E-mail*: phattanardp@fabrinet.co.th

Hillel J. Chiel received his B.A. from Yale University and his Ph.D. from M.I.T. He did postdoctoral work in the Center for Neurobiology and Behavior at Columbia University and the Molecular Biophysics Research Department at AT&T Bell Laboratories. He is currently Professor of Biology, Neurosciences and Biomedical Engineering at Case Western Reserve University. His research focuses on the neuromechanics of adaptive behavior in the marine mollusk *Aplysia californica*. He has more than eighty publications and serves as an editor of the Journal of Neural Engineering. In 2004, he was elected a Fellow of the Institute of Physics, London, England. *Address*: Department of Biology, DeGrace 304, 2080 Adelbert Road, Case Western Reserve University, Cleveland, OH 44106-7080. *E-mail*: hjc@case.edu



Randall D. Beer received his Ph.D. in computer science in 1989. From 1989 to 2006, he was a professor of electrical engineering and computer science, biology, and cognitive science at Case Western Reserve University. He spent the 1995-1996 academic year as a visiting scientist at the Santa Fe Institute, where he also served as an external faculty member for the next 6 years. In 2006, he joined the cognitive science program at Indiana University, where he is currently a professor of computer science and of informatics, as well as a member of the Center for the Integrative Study of Animal Behavior. Prof. Beer's research is broadly concerned with understanding how coordinated behavior arises from the dynamical interaction of an animal's nervous system, its body, and its environment.

Downloaded from http://adb.sagepub.com at PENNSYLVANIA STATE UNIV on April 16, 2008
© 2007 International Society of Adaptive Behavior. All rights reserved. Not for commercial use or unauthorized distribution.