**THE ROYAL
SOCIETY**

Evolving spike-timing-dependent plasticity for single-trial learning in robots

By Ezequiel A. Di Paolo

*School of Cognitive and Computing Sciences, University of Sussex,
Brighton BN1 9QH, UK* (ezequiel@cogs.susx.ac.uk)

Single-trial learning is studied in an evolved robot model of synaptic spike-timing-dependent plasticity (STDP). Robots must perform positive phototaxis but must learn to perform negative phototaxis in the presence of a short-lived aversive sound stimulus. STDP acts at the millisecond range and depends asymmetrically on the relative timing of pre- and post-synaptic spikes. Although it has been involved in learning models of input prediction, these models require the iterated presentation of the input pattern, and it is hard to see how this mechanism could sustain single-trial learning over a time-scale of tens of seconds. An incremental evolutionary approach is used to answer this question. The evolved robots succeed in learning the appropriate behaviour, but learning does not depend on achieving the right synaptic configuration but rather the right pattern of neural activity. Robot performance during positive phototaxis is quite robust to loss of spike-timing information, but in contrast, this loss is catastrophic for learning negative phototaxis where entrained firing is common. Tests show that the final weight configuration carries no information about whether a robot is performing one behaviour or the other. Fixing weights, however, has the effect of degrading performance, thus demonstrating that plasticity is used to sustain the neural activity corresponding both to the normal phototaxis condition and to the learned behaviour. The implications and limitations of this result are discussed.

Keywords: spiking neural networks; spike-timing-dependent plasticity;
activity-dependent synaptic scaling; single-trial learning; evolutionary robotics

1. Introduction

The aims of autonomous robotics, as conceived by W. Grey Walter (1953) and others, are scientific as well as practical. The design of autonomous robots has a lot to gain from drawing inspiration from the biological sciences, and, by constructing integrated adaptive agents in constant sensorimotor coupling with their environment, autonomous robotics offers a powerful modelling tool capable of informing biology. Grey Walter saw the potential significance of whole-agent modelling, even when using extremely simplified analogies to animal nervous systems, as a complement to detailed subsystem modelling, where typically abstractions occur at different levels. Computational neuroscience is a good example of this latter mode of research, where sensorimotor loops are opened by assuming known and independent input patterns to a system, but where the modelling granularity is richer.
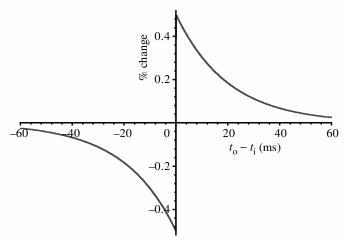
Figure 1. Time window for STDP. The percentage and direction of synaptic change are given by the time difference between pre-synaptic (t_i) and post-synaptic (t_o) spikes.

There is potentially much to be gained by using autonomous, and in particular evolutionary, robotics in an exploratory mode by looking for the consequences of including different mechanisms, analogous to neural counterparts, into a robot's controller and studying their effects on behaviour (Ruppin 2002). Examples of work of this kind include Webb's robotic model of cricket phonotaxis (Webb 1995, 2000), the study of effects of gaseous neuromodulators in artificially evolved neural controllers by Husbands *et al.* (1998), and the study of self-organizing neural network configuration by evolving synaptic plasticity by Floreano & Urzelai (2000). The current work seeks to similarly explore the effects of synaptic spike-timing-dependent plasticity (STDP) in a single-trial learning scenario using the evolutionary robotics methodology (Nolfi & Floreano 2000).

Grey Walter himself explored learning behaviour as a natural extension to the behavioural repertoire of his robotic tortoises. He reported a further development on *Machina speculatrix* which involved, besides phototaxis and obstacle avoidance, the capability of associating two classes of stimuli and so led to a form of conditional learning. He called this new robot *Machina docilis*. 'Memory' was stored in an oscillating circuit (Walter 1951, 1953). Oscillations would serve as a trace of the conjoint appearance of conditioned and unconditioned stimuli, so that after learning, the conditioned stimulus alone, together with the trace, would elicit the appropriate response. The majority of current models of learning tend to think of such memory traces in a different way. Influenced by connectionist models, learning is thought to be achievable thanks to synaptic plasticity and what is learned is reflected in long-term changes in the weight configuration of the whole network. The 'trace' was inherently dynamic for Walter, whereas connectionists more often think of a punctuated change followed by a static situation.

Work on evolutionary robotics has begun to challenge this stationary perspective on learning by providing examples of evolved learning behaviour in robots using neural networks that lack synaptic plasticity, and yet retain enough dynamic plasticity to respond with different patterns of activity in a history-dependent manner (Tuci *et al.* 2003; Yamauchi & Beer 1994). Other work has demonstrated that synaptic plasticity may form part of the complex dynamics of the network, never actually resting

on a stationary configuration, but remaining engaged in constant changes brought about by the closed sensorimotor loop (Di Paolo 2000, 2003; Floreano & Urzelai 2000). Synaptic plasticity plays a traditional role in none of these models; moreover, the models demonstrate that synaptic plasticity is, in the most general case, not necessary for learning to occur and that its presence can subserve functional purposes other than learning.

This paper is motivated by an exploration of this issue in the case of STDP in networks of spiking neurons. Recent work in evolutionary robotics has begun to explore the use of spike-based neural controllers (Floreano & Mattiussi 2001; French & Damper 2002). Spiking neural networks possess a number of attractive features. They have greater computational power than similar networks of threshold sigmoidal gates (Maass 1997). They can support a variety of functional specificity from rate-based codes to structured codes based on the timing of action potentials (Gerstner *et al.* 1997). They can perform novel kinds of computation, such as the recognition of temporal patterns using transient synchrony (Hopfield & Brody 2001) and real-time computation without stable states in high-dimensional 'liquids' of transient activity (Maass *et al.* 2002). Their complexity makes evolutionary robotics an appropriate tool of design and exploration.

STDP is a fast-acting mechanism which potentiates or depresses a synapse, depending on the relative timing between single pre- and post-synaptic spikes. The case most described in the literature is temporally asymmetric: potentiation occurs if the pre-synaptic spike arrives before the post-synaptic spike and depression occurs otherwise. The effective time-scale of STDP is of tens of milliseconds—no synaptic effect occurs between spikes separated by longer periods. Although its cumulative effect could lead to predictive learning in the presence of an iterated sequence of inputs (see below) it is not clear how other forms of learning, such as single-trial learning, could be achieved by STDP given the vast differences in time-scale between the mechanism and the task requirement (typically seconds at least). See also Mehta *et al.* (2002) on this issue. Grey Walter's oscillator circuit was, at least in principle, capable of such learning. Whether single-trial learning is possible at all using only STDP is the question we ask in this investigation. For this, we extend our recent results in evolving STDP controllers for simple phototaxis in order to evolve robots capable of reverting from positive to negative phototaxis in the presence of a sound stimulus.

Section 2 provides some background on recent work on STDP, §3 describes the methods, §4 describes the results, and the last section draws some conclusions.

2. STDP

Experimental neuroscientific evidence suggests that the degree and direction of change in the strength of a synapse subject to repeated pairings of pre- and post-synaptic action potentials depend on their relative timing (Bi & Poo 1998; Markram *et al.* 1997). (See Bi & Poo (2001) for a review.) Synaptic modification depends on whether the pre- and post-synaptic spikes are separated in time by less than a critical window of the order of a few tens of milliseconds. In most cases studied, if a pre-synaptic spike precedes the post-synaptic spike, the synapse is potentiated, whereas the opposite relation leads to depression of the synapse. This results in a temporally asymmetric plasticity rule (figure 1), which deserves the name 'Hebbian' because

of its tendency to strengthen causal correlations between spikes. There is empirical evidence, however, for non-Hebbian plasticity of this kind (Abbott & Nelson 2000; Bi & Poo 2001). Many theoretical studies have concerned themselves with this rule of plasticity and its desirable properties, such as a trend towards inherent stability in weight distribution and neural activity, unlike purely rate-based Hebbian rules which often require additional constraints (Kempter *et al.* 1999; Rubin *et al.* 2001; Song *et al.* 2000). One possible expression for this rule is

$$\Delta w = \begin{cases} w_{\max} A^+ \exp(-s/\tau^+) & \text{if } s \geqslant 0, \\ -w_{\max} A^- \exp(s/\tau^-) & \text{if } s < 0, \end{cases}$$

where $s = t_{\mathrm{o}} - t_{\mathrm{i}}$ is the time difference between a post-synaptic and pre-synaptic spike, A^+ and A^- are positive constants, and τ^+ and τ^- correspond to an exponential decay of the order of a few tens of milliseconds. Other filters may be used instead of the exponential decay, but this form is particularly suitable for implementation in an evolutionary robotics context, as will be shown in the next section.

One of the key concerns when studying rules for synaptic plasticity is their regulatory properties. Hebbian learning on its own leads to runaway processes of potentiation and cannot account for the stability of neural function. Additional elements, such as the directional damping of synaptic change (Rubin *et al.* 2001) or longer-term stabilizing regulation based on post-synaptic activity (Horn *et al.* 1998; Turrigiano 1999), may come to the rescue. These can lead to unsaturated distributions of synaptic strengths in the first case and to regulated neuronal firing in the second, and will also be investigated in this work.

Although STDP is a topic that has drawn much attention recently, most theoretical studies have concerned themselves with the properties of the temporally asymmetric plastic rule. There are, however, a few hypotheses about its functional role. For instance, Abbott & Blum (1996) show in a general model how firing patterns in a neural array (such as a receptive field), where neurons fire preferably at certain input values in a sequence of inputs, can, by means of temporal asymmetry in plasticity, lead to prediction of the inputs in a sequence through repeated presentation. This is because the synapses of neurons that fire in succession are strengthened from those that fire first to those that fire later (and are depressed in the opposite direction). Empirical evidence in the experience-dependent change in skewness in place fields in the rat hippocampus supports the findings of this model (Mehta *et al.* 1997, 2000). Related to this, Yao & Dan (2001) have found that repetitive pairing of visual stimuli at two different orientations induced a shift in orientation tuning in cat cortical neurons depending on the relative timing of presentation and is compatible with STDP.

Other related functional implications have also been suggested. Rao & Sejnowski (2001) suggest that STDP could be involved in implementing some form of temporal-difference learning (Sutton 1988) and show this in a model of input spike prediction; and Chechik (2003) has recently compared theoretical rules of plasticity derived from the principle of information maximization of relevant input with empirical rules to conclude that temporal asymmetry can increase input information to near optimal levels.

This kind of functionality is hard to compare with the results obtainable from the present work on a simple robotic task, as it is more likely to play a significant

role when sensory surfaces or arrays are included in the robot model as well as something equivalent to receptive fields. Because of the constraints put by the number of evolutionary evaluations, such elements are not included in this initial study but will be of central importance in the future.

3. Methods

(a) Robots and tasks

Since we are interested in exploring a novel mechanism for robot control, the chosen task is at this stage deliberately simple so as to facilitate comparisons with alternative approaches. Simulated robots are evolved to perform phototaxis on a series of light sources. The robots have circular bodies of radius $R_0 = 4$, with two motors and two light sensors. The angle between sensors is $120°$, but a small random displacement of between -5 and $5°$ is added at the start of each evaluation. Motors can drive the robot backwards and forwards in a two-dimensional unlimited arena.

The neural network consists of six nodes and is fully connected except for self-connections. Neurons can be either excitatory or inhibitory and this is set genetically. Trials with larger numbers of neurons have been carried out successfully, but have not been systematically studied.

The whole system is simulated using an Euler integration method with a time-step of 1 ms (25% of the minimum time-scale). Robots are run for two independent evaluations, each consisting of the sequential presentation of two distant light sources. Only one source is presented at a time for a relatively long period T_S chosen randomly for each source from the interval [7.5 s, 12.5 s] (each evaluation consists therefore of an average of 2×10^4 update cycles). The initial distance between robot and new source is randomly chosen from [60, 80], the angle from [0, 2π) and the source intensity from [3000, 5000]. The intensity decays in inverse proportion to the square of the distance to the source.

The simulated robots use photoreceptors that are activated by the light intensity corresponding to their current position if the light source is directly visible (i.e. an angle of acceptance of $180°$). This intensity is multiplied by the sensor gain equal for both sensors (genetically set from range [0.1, 20]) and clipped for values beyond a maximum of 20. A spike train is generated using a Poisson process with variable rate (maximum 100 Hz) by linearly transforming the sensor value into the instantaneous firing frequency. The Poisson spike trains coming from the left and right sensors are fed into neurons N2 and N3, respectively. Additionally, uniform noise is present in the sensors (and motors) with range 0.2 (before scaling by gains): this results in spikes that fire randomly with very low probability when the sensor is not stimulated. A 'sound' sensor is used for the aversive stimulus and is activated in half of the trials, when the robot must learn to perform negative phototaxis. In such trials, the sound sensor is activated when the robot has approached a source of light more than half the initial distance.

Two motors control the robot wheels. Each motor is controlled by two neurons, one that drives it forwards and the other one backwards, using a spike-based leaky integrator. The left motor is controlled by neurons N0 (forward) and N4 (backward) and the right motor by neurons N1 (forward) and N5 (backward).

A population of 30 robots was evolved using a generational GA with truncation selection. In the plastic scenarios described below initial weights are randomly chosen

at the start of each evaluation from the interval $[0, w_{max}]$, while the parameters for the plasticity windows and scaling constants are evolved. In the non-plastic scenario, synaptic strengths are encoded genetically. Other genetically set parameters include sensor and motor gains, motor decay constant and whether neurons are inhibitory or excitatory. All parameters are encoded in a real-valued genotype, each gene assuming a value within $[0, 1]$, threshold of 0.5 to encode whether a neuron is excitatory or inhibitory; each parameter is linearly scaled to the corresponding range of values, except for sensor and motor gains which are scaled exponentially. Only vector mutation (Beer 1996) is used with a standard deviation of vector displacement of 0.5 (maximum genotype length is 220), genetic boundaries are reflective. Fitness for normal phototaxis is calculated according to

$$F_{p} = \frac{(1 - M^2)}{T_S} \int f \, dt; \quad f = 1 - \frac{d}{D_i}$$

if the current distance to the source d is less than the initial distance D_i, otherwise $f = 0$. M measures the average difference in activity between the motors divided by the motor gain,

$$M = \frac{0.125}{T_S} \int \frac{(M_L - M_R)}{M_G} \, dt.$$

Near-optimal fitness will be obtained by robots approaching a source of light rapidly and with minimal integrated angular movement. This measure is applied to half the number of trials per individual evaluation. In the other half the robot must learn, on sensing the aversive stimulus, to avoid the light. This may be achieved in two different ways, either by making $F_{np} = 1 - F_p$ during the whole trial, or by making $F_{np} = 1 - F_p$ only if the aversive stimulus has been sensed at least once, otherwise $F_{np} = F_p$. The difference between the two is that in the first case, positive phototaxis must be pre-evolved on its own for a small number of generations, otherwise the initial random population already performs to 50% of the maximum fitness and will not evolve to regulate behaviour according to the presence of light. In the second case, this pre-evolution stage is not necessary, as the fitness rewards positive phototaxis as long as the sound is not heard. Only robots that perform a certain degree of phototaxis will find themselves in a condition where the fitness measure changes so they must additionally evolve light avoidance from that point onwards.

(b) Neural controller

An integrate-and-fire model with reversal is used for the neural controller. The time evolution of the membrane potential V of a neuron is given by

$$\tau_m \frac{dV}{dt} = V_{rest} - V + g_{ex}(t)(E_{ex} - V) + g_{in}(t)(E_{in} - V),$$

where τ_m is the membrane time constant (range $[10 \, ms, 40 \, ms]$), $v_{rest} = -70 \, mV$ is the rest potential, the excitatory and inhibitory reversal potentials are, respectively, $E_{ex} = 0 \, mV$ and $E_{in} = -80 \, mV$.

A noisy threshold value, V_{thres}, is given by a normal distribution with a genetically set mean value (range $[-60 \, mV, -50 \, mV]$) and a deviation of 1 mV. When the membrane potential reaches this threshold, an action potential is fired and V is reset

to V_{rest}. A random refractory time distributed uniformly in the interval $[2\,\text{ms}, 4\,\text{ms}]$ prevents the neuron from firing another spike within this period.

Every time a spike arrives to neuron j from an excitatory pre-synaptic neuron i the (non-dimensional) excitatory conductance of j is increased by the current value of the synaptic strength $(w_{ij}(t))$:

$$g_{\text{ex}}(t) \rightarrow g_{\text{ex}}(t) + w_{ij}(t).$$

The inhibitory conductance g_{in} is similarly affected by spikes coming from inhibitory neurons. The conductances otherwise decay exponentially,

$$\tau_{\text{ex}}\frac{\mathrm{d}g_{\text{ex}}}{\mathrm{d}t} = -g_{\text{ex}}, \qquad \tau_{\text{in}}\frac{\mathrm{d}g_{\text{in}}}{\mathrm{d}t} = -g_{\text{in}},$$

with τ_{ex} and τ_{in} genetically set for each neuron from the range $[4\,\text{ms}, 8\,\text{ms}]$.

The current motor value is stored in variables $M_{\text{L,R}}$, which are directly translated into the left and right velocities, respectively,

$$\tau_{\text{mot}}\frac{\mathrm{d}M_{\text{L,R}}}{\mathrm{d}t} = -M_{\text{L,R}} + M_{\text{G}}\left(\sum \delta(t - t_{\text{forw}}^{(\text{f})}) - \delta(t - t_{\text{back}}^{(\text{f})})\right),$$

with τ_{mot} genetically set from the range $[40\,\text{ms}, 100\,\text{ms}]$ and M_{G} from $[0.1, 50]$. Both motors have a same value for their gains and decay constants. For this low inertia model the output of the motor provide the instantaneous velocities on the right- and left-hand sides. This approach marks a difference with previous work on the evolution of spiking controllers which have used a neural rate estimation method for driving the motor (Floreano & Mattiussi 2001).

(i) *STDP*

The properties of plastic windows (figure 1) are evolved for each synapse in the neural network controller. Following Song *et al.* (2000), synaptic change is implemented using two recording functions per synapse $P^-(t)$ and $P^+(t)$. Every time a spike arrives at the synapse, the corresponding $P^+(t)$ is incremented by A^+ and, every time the post-synaptic neuron fires, the corresponding $P^-(t)$ is decremented by A^-. Otherwise, these functions decay exponentially with time constants τ^- and τ^+, respectively. $P^-(t)$ is used to decrease the synaptic strength every time the pre-synaptic neuron fires:

$$w_{ij}(t) \rightarrow w_{ij}(t) + w_{\text{max}}P^-(t).$$

Analogously, $P^+(t)$ is used to increase the synaptic strength every time the post-synaptic neuron fires:

$$w_{ij}(t) \rightarrow w_{ij}(t) + w_{\text{max}}P^+(t).$$

Accordingly, coincidental spikes produce both depression and potentiation in the synapse. The maximum synaptic strength is $w_{\text{max}} = 1$. This method facilitates the computational implementation of STDP by eliminating the need for keeping track of spike trains or calculating other response functions which could be more costly.

The values for A^+ and A^- and τ^+ and τ^- are genetically set per synapse from the ranges $[0.0001, 0.05]$ and $[10\,\text{ms}, 40\,\text{ms}]$, respectively. In all the experiments reported here the plastic windows are Hebbian, that is, spikes arriving before a post-synaptic

action potential always potentiate a synapse and those arriving after always depress it. Experiments relaxing this constraint, i.e. allowing anti-Hebbian or purely potentiating or depressing windows, have also been carried out successfully, but are not reported here.

(ii) *Activity-dependent scaling (ADS) of synapses*

Some of the mechanisms used by neurons to regulate their firing rate homeostatically are thought to affect all incoming synapses scaling them up or down independently of the pre-synaptic activity (Turrigiano 1999). If the post-synaptic activity is above a certain target, excitatory synapses are scaled down; otherwise, they are scaled up, thus preventing sustained levels of activity that are too high or too low. Following van Rossum *et al.* (2000), excitatory synapses are modified according to

$$\tau_{\text{ADS}}\frac{\mathrm{d}w_{ij}}{\mathrm{d}t} = w_{ij}(z_{\text{goal}} - z_j),$$

where $z_{\text{goal}} = 40\,\text{Hz}$ and τ_{ADS} is genetically set from the range $[1\,\text{s}, 10\,\text{s}]$. The firing rate z_j of a neuron is estimated by a leaky integration of the spike train,

$$\tau_z\frac{\mathrm{d}z_j}{\mathrm{d}t} = -z_j + \sum \delta(t - t^{(\text{f})}),$$

where $t^{(\text{f})}$ are the times when the neuron emits a spike (the sum runs over all previous spikes) and $\tau_z = 100\,\text{ms}$.

In real neurons, this is a mechanism that acts over long time-scales (over hundreds to thousands of seconds) (Turrigiano *et al.* 1998), but due to computational limitations (the very long evaluation runs that would be required) the chosen time-scale ($\sim \tau_{\text{ADS}}$) is faster than this but still significantly slower than the rest of the time-scales in the system. Even though the above mechanism acts on excitatory synapses, in the current context it has also been applied when the pre-synaptic neuron is inhibitory by multiplying the right-hand side above by -1. A similar homeostatic mechanism has been successfully implemented in robots capable of adapting to sensorimotor disruptions not previously experienced (Di Paolo 2000).

(iii) *Directional damping*

Synaptic weights are constrained within the range $[0, 1]$. This can be achieved simply by a stop condition at the boundaries or by means of damping factors that vanish as the weight value approaches a boundary. The choice can have important consequences. No damping leads to a bimodal distribution of weights under random stimulation (Song *et al.* 2000), where most weights adopt the minimum or maximum values in the range, but few values in between. The same happens with purely positional damping, i.e. factors that slow down weight change near the boundaries, but depend only on the current weight value. A biologically plausible alternative is directional damping, whereby if a weight value is near a boundary, changes that push this value towards the boundary are slowed down, but changes that push it away from the boundary are not. The equilibrium weight distribution in this case tends to be unimodal and centred around the point where potentiation and depression equilibrate (Rubin *et al.* 2001). Directional damping is supported empirically by the

observation that spike-driven potentiation is more pronounced than the expected linear variation at synapses of relatively low initial strength in cultured hippocampal cells (Bi & Poo 1998). It was also observed that the mean fractional negative change was constant over a wide range of initial weights, corresponding to a damping linear factor for absolute depression equal to the current weight value.

Linear directional, or multiplicative, damping is simply implemented by transforming a weight change (as resulting from STDP or ADS or both):

$$\left. \begin{array}{ll} \Delta w_{ij} \to (1 - w_{ij})\Delta w_{ij} & \text{if } \Delta w_{ij} \geqslant 0, \\ \Delta w_{ij} \to w_{ij}\Delta w_{ij} & \text{if } \Delta w_{ij} < 0, \end{array} \right\} \quad \text{for } w_{ij} \in [0, 1].$$

(iv) *Neural noise*

Different sources of neural noise have been modelled. At any given time, Gaussian noise with zero mean and 1 mV deviation is applied to the value of the firing threshold. This is the only source of neural noise in the first set of experiments. Additionally, for the second set, the refractory period is randomly set every timestep using a uniform distribution ([2 ms, 4 ms] for cases of short refractory period, [4 ms, 8 ms] for long refractory periods). Background noise is modelled as an incoming Poisson train to every neuron with a frequency of 10 Hz, and spontaneous firing has also been modelled using a baseline 10 Hz Poisson process for each neuron, but subject to refraction.

(v) *Synaptic decay*

In order to test the robustness of the evolved controllers to perturbations in their internal configuration, synaptic weights are allowed to decay exponentially to zero with a time constant that can be as fast as 100 ms. Synaptic decay is not affected by directional damping and is not applied during evolution but only during behavioural tests.

(vi) *Poisson filters and randomized delays*

In order to test the reliability of the evolved controllers on the precise timing of spikes, the simple expedient of filtering the output of a neuron with a Poisson process emitting random spikes at the same instantaneous rate has been used. Information about firing rate is conserved, but precise spike-timing is disrupted. Because the rate, z, is estimated using only previous spikes, it is only possible to approximate the instantaneous firing rate of the neuron in this manner. It is expected, however, that if controllers rely heavily on firing rates, the disruption in performance should not be too strong. A more sophisticated method consists of introducing artificial random delays in the firing time of single or multiple neurons. This is done by keeping a short subtrain corresponding to the last T ms of activity and swapping the current fire state of a neuron with a randomly selected state in the subtrain, thus conserving the number of spikes. This is similar in objective to tests *in vivo* on honeybee odour discrimination demonstrating the role of synchronized neural assemblies (Stopfer *et al.* 1997). These tests are applied only after evolution.
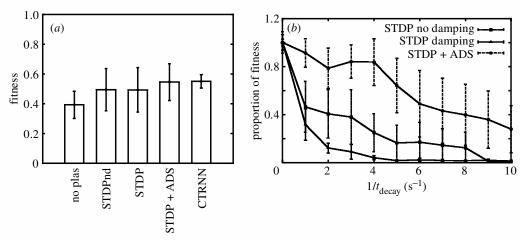
Figure 2. Fitness and robustness. (*a*) Average fitness of the best individual in the last generation (five independent runs for each condition, no plasticity, STDP with no directional damping, STDP with damping, STDP + ADS and CTRNN controllers for comparison). (*b*) Robustness against synaptic decay; t_{decay} indicates the speed with which weights decay exponentially to zero (10 evaluations per point, error bars indicate standard deviation).

(vii) *Continuous-time, recurrent neural networks (CTRNNs)*

Control runs using rate-based CTRNNs (Beer 1990) have been performed. These are defined by

$$\tau_i \frac{dV_i}{dt} = -V_i + \sum_j w_{ji} z_j + I_i, \quad z_j = \frac{1}{1 + \exp[-(V_j + b_j)]},$$

where V_i represents the membrane potential of neuron i, τ_i the decay constant (range [0.4 s, 4 s]), b_i the bias (range $[-3, 3]$), z_i the firing rate, w_{ij} the strength of synaptic connection from node i to node j (range $[-8, 8]$), and I_i the degree of sensory perturbation for sensory nodes. A plastic version of these controllers has also been used, and is described in more detail later.

4. Results

(*a*) *Normal phototaxis*

For completeness, results for normal phototaxis are summarized in this section. A more detailed study can be found in Di Paolo (2003). Robots are able to evolve normal phototaxis using STDP and STDP + ADS controllers, as well as non-plastic ones (figure 2*a*). It is found that selecting for plastic properties operational at the millisecond range, successfully produces controllers capable of adaptive behaviour at a time-scale of many seconds. The effect of synaptic decay has been studied and STDP + ADS controllers have been found to perform more robustly under such perturbations (figure 2*b*). This is not surprising given the compensatory nature of the ADS mechanism. Interestingly, it was often found that ADS plays a balancing role for neural input which often leads to synaptic oscillations and a bursting firing pattern and 'rhythmic' patterns of movement as shown in figure 3.
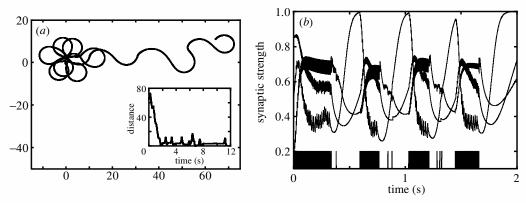
Figure 3. Example of evolved STDP + ADS controller. (*a*) Trajectory; the small circle indicates the position of the light source and the robot's distance from it is plotted in the inset. (*b*) Oscillatory weight dynamics for three synapses affecting node N0, together with its firing pattern. Other nodes and synapses show similar behaviour.
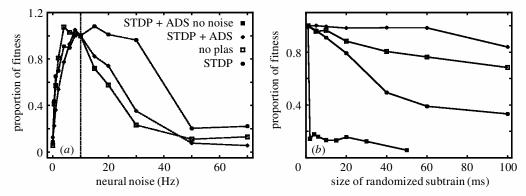

Figure 4. Noise and disruption of spike timing: (*a*) effect of increased and decreased spontaneous random neural firing on network performance on three independently evolved individuals for STDP + ADS controllers (similar dependence for the other cases). All networks were evolved with a background random firing of 10 Hz (vertical lines) and performance in normalized at this point. Each point is the average of 20 runs; (*b*) effect of spike-train randomization for all neurons, averaged over five independent runs, for the case of noisy neural controllers and STDP + ADS low neural noise controllers (squares).

Controllers including neural noise do not greatly change these results, with two important exceptions. Firstly, it is found that noise has functional value for the neural controllers as shown by the effects on performance of modifying the level of noise. Larger levels of neural noise lead to poorer performance, as expected, but so do smaller levels, thus indicating that the controllers have evolved to rely on an optimal presence of neural noise (figure 4*a*). Secondly, the response to spike randomization is also different for noisy and non-noisy controllers. The latter tend to fire regular and often synchronized spike trains, which, on disruption or application of Poisson filters, produce catastrophic reductions in performance (figure 4*b*). In contrast, noisy controllers are much more robust, even though spike trains also show some evidence of temporal coordination (not shown). The unintuitive conclusion is that, despite its

reliance on precise spike-timing, STDP can manage to configure a successful neural controller even when spike-timing information is lost.

The asymmetric STDP rule retains certain interesting properties in this case when compared with a rate-based synaptic controller. It is found that STDP controllers are able to reach a stable state more rapidly and more reliably than the rate-based counterparts under the same conditions, thus achieving higher fitness by being able to perform the task earlier in their lifetimes. This is shown by comparing the time variance across many runs for STDP and rate-based controllers. Unlike the latter, the former decreases to very low values, thus indicating that the final distribution of weight values does not depend strongly on spike-timing information (Di Paolo 2003).

(*b*) *Single-trial learning*

A simple learning task is implemented using the above scenario. The robot must approach sources of light as usual, but if an aversive stimulus ('sound') is presented once during the approach, then it must avoid them from then on. Robots are then evaluated on two different conditions: (i) normal phototaxis in the absence of aversive stimulus, and (ii) negative phototaxis in the presence of aversive stimulus. For balance, each evaluation consists of an equal number of trials under each condition. Because the two conditions are described by opposite fitness criteria (see § 3), it is actually extremely improbable that an initially random population will be able to evolve the desired behaviour, as from the start it will achieve an average fitness of 50% by not doing anything (implicitly, not approaching the light under condition (i)). Of the two solutions to this problem proposed above, here the first one is used: the population is pre-evolved to perform normal phototaxis for a few generations. Preliminary tests with the other option showed it was harder to evolve.

Additionally, an incremental method is used for the robots to evolve appropriate responses to the aversive stimulus. Each evaluation consists of the presentation of five sources of light in sequence. If the stimulus is presented only once, on approaching the first source, the opportunity presented during evolution for the controller to explore reasonable responses is available only during a short time window and absent during the rest of the trial. A solution to this problem is to present the aversive stimulus on approaching each of the five sources of light and evolve a reactive withdrawal during a fixed number of generations. After that, the stimulus is presented on approaching the first four sources of light, but not the last one, again for a fixed number of generations, and so on, until, finally, the stimulus is presented on approaching only the first source of light. In this way, what is initially a withdrawal response requiring the constant presence of the stimulus evolves into single trial learning requiring its presence only in one instance. Other variants of this process are possible, including making the aversive stimulus less and less reliable over generations for all except the first source of light (e.g. by introducing noise, or reducing its intensity).

(i) *Evolution and behaviour*

Figure 5 shows one run at the end of which single-trial learning evolved using only STDP with eight fully connected noisy neurons (5 Hz background firing). Each evaluation consisted of one trial under each condition. Normal phototaxis is pre-evolved during the first 10 generations (zone I in the figure). It is not necessary (and perhaps inconvenient) to achieve a high level of fitness during this stage. In zone II
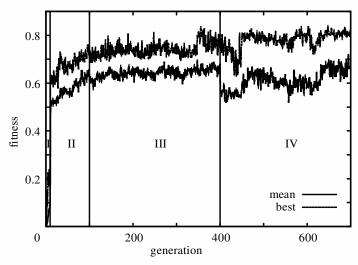
Figure 5. Average and maximum fitness using an incremental approach to single-trial learning. Stage I corresponds to normal phototaxis in all trials and stage II to withdrawal reaction in half of the trials with constant presentation of aversive stimulus; in stage III the aversive stimulus is presented in the first four of five light sources, and in stage IV the stimulus is presented in the first two of six sources.
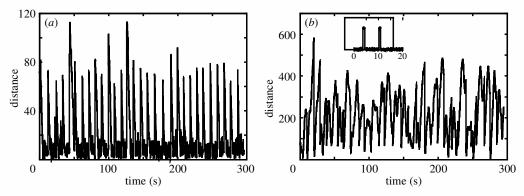

Figure 6. Distance between robot and light source during the sequential presentation of 30 sources: (*a*) normal phototaxis; (*b*) negative phototaxis after presentation of aversive stimulus on first light source. The inset shows the time when stimulus is presented.

the aversive stimulus is presented on approaching each of the five light sources, in zone III on approaching four of the five, and in zone IV the number of light sources is extended to six and the stimulus is presented on approaching the first two. A further stage was found to be unnecessary; by this point the robots were able to perform single-trial learning. In this case, the controller is made of five excitatory and three inhibitory neurons (both backward motors and the interneuron). In other runs a similar balance has been obtained with either three excitatory and five inhibitory neurons or equal numbers of inhibitory and excitatory neurons. Sensor neurons have always tended to be excitatory.

The behaviour of the best robot of the last generation can be seen in figure 6, where the distance to the light source is plotted over time. If the sound stimulus is not
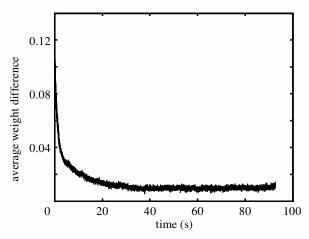
Figure 7. Mean squared difference between weight values over time for the positive and negative phototaxis conditions for a same controller averaged for 10 comparisons and for all the weights.

presented during the trial the robot performs normal phototaxis in a stable manner (figure 6*a* shows the presentation of 30 light sources, whereas the robot has been evolved for at most six light sources). However, on presentation of the sound stimulus after approaching the first source of light, the robot learns to avoid other sources from that point onwards (figure 6*b*, note the vertical scale). It is interesting that there is no extinction of the learned behaviour. Under the conditions of evaluation, there is no adaptive gain to forgetting what has been learned, even though it is possible to conceive similar scenarios where this would be advantageous.

(ii) *Synaptic and firing patterns*

On analysing the differences in weight distributions for the robot in the presence and absence of aversive stimulus, it was found that although some weights show some degree of difference between the two situations, the majority tend to be distributed in the same regions of the range. Figure 7 shows the squared difference of corresponding weight values over time between runs for normal and negative phototaxis (averaged for all weights and for 10 pairs of runs). Although the difference does not vanish, a good part of the remaining variance can be attributed to the firing pattern when the robot performs negative phototaxis, as detailed below. Some weights show a more pronounced difference in their distribution over time; for instance, figure 8 compares two synapses from the neuron receiving the sound stimulus to the backward left and right motor neurons. These two cases were found to show qualitatively the most significant differences in distribution.

A more pronounced difference is found in the firing pattern of the neural controller. Figure 9 shows the activity of the neural network for a robot at the moment of switching between positive and negative phototaxis on sensing the stimulus (top of the figure). The firing pattern previous to the stimulus being sensed is characteristic of positive phototaxis; it shows a response to light sensors (not shown) typical of robots performing only phototaxis (Di Paolo 2003). On sensing the stimulus the receptor neuron (N6) as well as others increase their firing frequency for a period of rapid firing during as long as the duration of the stimulus. This rapid firing activates
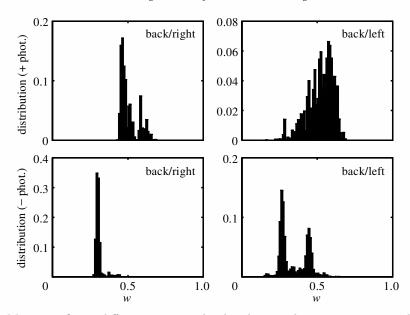
Figure 8. Most significant differences in weight distributions between positive and negative phototaxis conditions (top and bottom, respectively) for the same controller. The two synapses correspond to connections from the sound-receptor neuron to the backward left and right motor neurons.
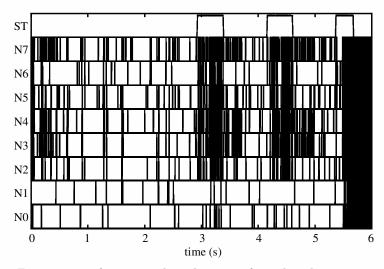

Figure 9. Firing pattern for a network at the point of switching between positive and negative phototaxis. Sound stimulus is shown at the top.

the two backward motors (N4 and N5) provoking a strong withdrawal reaction. When the stimulus is sensed for the third time, however, a stable rapid firing pattern emerges throughout the network. It would seem that this pattern effectively produces a sensor shutdown as sensor information apparently can do little to influence the firing of the neurons. The robot simply ignores the presence of light and because of its fast speed it never stays near it for long enough. However, a comparative test
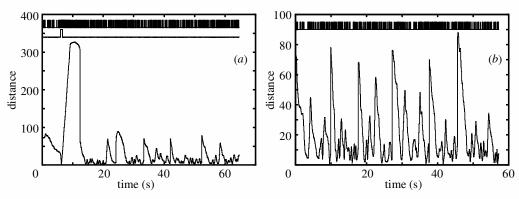
Figure 10. Behaviour with perturbed timing information (distance to source in sequential presentation of six light sources). (*a*) The negative phototaxis condition with a small randomization of spike train (subtrain of 5 ms); the time of sound activation is also shown at the top together with the sound-receptor spike train. Fast firing becomes active after sound is sensed and remains active during fast withdrawal from the source. However, the pattern is unstable and the robot reverts to normal phototaxis. (*b*) The same controller under a positive phototaxis condition with a significant degree of spike-train randomization (subtrain of 50 ms); the sound-receptor firing pattern is included for comparison.

between a normal robot and a robot with artificially removed sensors shows that the behaviour is different. The normal robot seems to be using light sensor activation to sustain the rapid firing dynamics, whereas the blind robot completely ignores the light even when the rapid firing is lost, and so moves more slowly.

If the difference between the two modes is to be found mainly in the firing pattern, then it may be possible to induce the behavioural change at different points in the robots lifetime. In effect, trials in which sound stimulus is presented arbitrarily result in a switch from positive to negative phototaxis. Thus, the stimulus is acting not so much as a conditioned stimulus, i.e. as associated with an unconditioned light stimulus, but rather as a release mechanism capable of inducing a change in behaviour at any point in the robot's lifetime. With hindsight, this is not so surprising, as the fitness function only asks for a change in behaviour to occur, not for an association between two classes of stimuli.

(iii) *Spike timing and noise*

An interesting asymmetry is found when the controller is subject to spike-train randomization. Whereas the controller doing normal phototaxis is quite robust to loss of timing information due to the endogenous level of neural noise (Di Paolo 2003), the learning of light avoidance is, on the contrary, quite brittle. The robot may withdraw from the light in the presence of sound, and rapid firing may be established, but very soon, if sound is no longer present, the network reverts to normal activity, and the robot performs positive phototaxis. This is shown in figure 10*a*, where a small spike-train randomization has been applied (5 ms subtrain). In contrast, in the normal phototaxis condition the robot performs well even with significant loss of timing information (figure 10*b* shows the distance to the light for 50 ms randomized subtrain). This indicates that, in the negative phototaxis condition, the maintenance
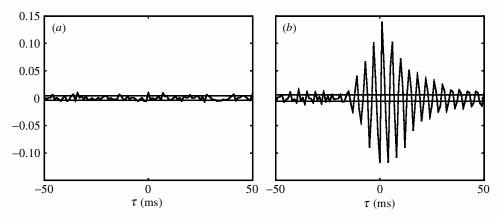
Figure 11. Covariograms between left and right forward motor neurons: (*a*) positive phototaxis, (*b*) negative phototaxis. Bands show the estimated interval of significance.
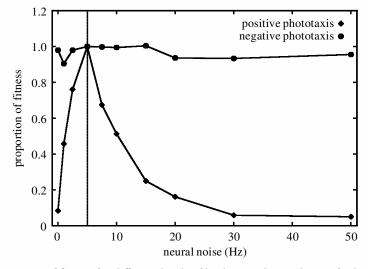

Figure 12. Proportion of fitness for different levels of background neural noise for both conditions. The vertical line indicates the level of noise used during evolution; fitness has been normalized at this level.

of rapid firing can only be a stable solution by heavy reliance on entrained spikes, as is clear in figure 11, which compares the covariograms between the left and right forward motor neurons for both positive and negative phototaxis (horizontal lines show interval of expectation for random firing corrected by frequency (Di Paolo 2003)).

Responses to variations in background noise are also different between the two conditions. Figure 12 shows the average performance over 10 trials for different levels of spontaneous random firing. Fitness is normalized to the evolved noise level (5 Hz). While performing positive phototaxis, the robot's response to noise is similar to that described in the previous section, i.e. a decay in performance for both increased and reduced levels of noise. Some degree of decay can also be observed for the negative phototaxis condition, but the effect is much less pronounced. This result shows that
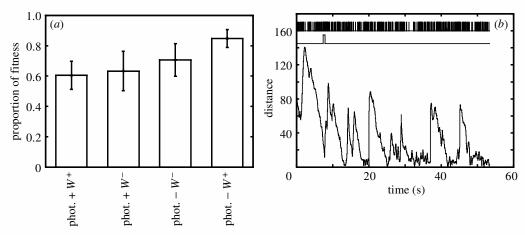
Figure 13. Performance for fixed weight networks: (*a*) proportion of fitness for each condition using fixed weights taken from a long run of the same and opposite conditions; (*b*) negative phototaxis with fixed weights (W^-) together with sound stimulus and spike train corresponding to sound sensor neuron.

the addition of background noise is a very different perturbation from the randomization of spike trains, and with different effects. Background Poisson noise does little to degrade the timing information of spike trains in the negative phototaxis regime because of its characteristic high-frequency activity.

(iv) *Removal of plasticity*

Finally, test runs were performed by fixing the weights of the network under each condition to the value acquired at the end of long run (30 light sources) with plasticity. Figure 13*a* shows the reduction in fitness for positive and negative phototaxis ('phot.+' and 'phot.−', respectively) when the fixed weights are taken from a long run of the same and opposite conditions (W^+ and W^- in the figure). It is clear that, for each condition, using fixed weights produces a reduction in fitness which does not depend strongly on which weight values are used, and whether those correspond to negative or positive phototaxis. It is possible to conclude that the final weight configuration carries little information about whether the robot should be approaching or avoiding lights (performance for negative phototaxis is even better when using the fixed set of weights W^+).

Even though the reduction in fitness for negative phototaxis does not seem too drastic, in fact, on observing the performance (using W^- weights), it is clear that the robot is not avoiding light sources, it is just taking longer to reach them (figure 13*b*). This means that although weight distribution seems less important a factor than the firing pattern in explaining the change of behaviour, synaptic plasticity does play a role in achieving and sustaining the necessary firing pattern. This is true also for the non-learning condition, where the reduction in fitness is even more significant. Synaptic plasticity is used less to 'store' a configuration of synaptic values than to generate the right neural dynamics.

5. Conclusions

The behaviour of plastic networks of spiking neurons as robot controllers is sufficiently rich to justify further exploration using synthetic design techniques such as evolutionary robotics. The search algorithm is able to find successful controllers by evolving only the rules of plastic change and the time properties of each neuron, all of which are active at the tens-of-milliseconds range. However, robots are able to behave appropriately over larger time-scales and to learn in a single trial over time-scales of tens of seconds. The traditionally connectionist answer to this dilemma is that relevant information about the robot's history is somehow retained in the network configuration in the form of appropriate weight distributions. If not, then some other mechanism is necessary to serve as a memory trace, such as sustained oscillations in Walter's learning robot. The result of this investigation shows that learning may not rely on either, but that different neural firing patterns may in themselves provide the substrate for learning to occur. This is different from other results in evolutionary robotics showing learning in CTRNNs in the absence of synaptic plasticity (Tuci *et al.* 2003; Yamauchi & Beer 1994). In such cases, neuron decay constants are allowed to be long enough to play a role similar to Walter's oscillating circuit. It is unclear how the present result would extend to more complex scenarios or what the implications to natural agents are. It is provided here as a working proof-of-concept, interesting enough to deserve further exploration regarding its generality.

Although robots achieve single-trial learning, the current result needs to be improved upon. It is clear that fitness function design is one of the central issues to be investigated further as, in the current set-up, the aversive stimulus cannot be rightly called a conditioned stimulus even if it is presented during evolution only very near the light source. The reason for this is that the fitness function rewards *not approaching* light sources on hearing the sound, and so a valid solution is to ignore the light sources once the stimulus is sensed. In that sense, sound acts as a release mechanism that switches between different stable behavioural modes. A more detailed fitness evaluation would differentiate between reaction to the aversive stimulus depending on the presence or absence of the unconditioned stimulus. The incremental method for fitness evaluations would also need to be improved to obtain more flexible controllers capable of 'forgetting' the conditioned behaviour, possibly by changing the class of unconditioned stimuli over evolutionary time. Further classes of stimuli would also be needed to make sure that learning occurs in association with the appropriate class and not just any class. This cannot be investigated in the current scenario, but it may easily be done with a more complex visual system, for instance in tasks of navigation or orientation towards an object.

Even though the bi-stable firing properties of the evolved controllers would seem to explain the origins of the bi-stable behaviour of the robot, STDP nevertheless plays a role besides the overall configuration of the neural network from a random initial state. No non-plastic controller has successfully evolved the learning task (the attempt is not *a priori* futile, since fixed synaptic weights do not impede the existence of multiple stable firing patterns that could supplement the lack of synaptic plasticity and subserve learning) and fixed-weight networks show decay in performance. Comparing performance with crossed sets of weight values shows that plasticity does not work by storing information in synapses. In contrast, the role played by STDP seems to be that of sustaining and possibly initiating the switch in firing patterns.

However, this is still far from clear. A possible clue is provided by the different effects of increasing neural noise and randomizing spike trains on learning. This is an issue that deserves further investigation and experiments are currently under way to determine how synaptic plasticity is involved.

As described in § 2, STDP seems to explain more naturally other forms of learning such as prediction of input in a repeated sequence, for instance, in visually guided navigation. Such forms of learning, requiring robots with a visual system and possibly an architecture including receptive fields of neurons with similar or equal properties, are currently being investigated using similar techniques to those described in this paper.

References

Abbott, L. F. & Blum, K. I. 1996 Functional significance of long-term potentiation for sequence learning and prediction. *Cereb. Cortex* **6**, 406–416.

Abbott, L. F. & Nelson, S. B. 2000 Synaptic plasticity: taming the beast. *Nature Neurosci.* **3**, 1178–1183.

Beer, R. D. 1990 *Intelligence as adaptive behavior: an experiment in computational neuroscience.* Academic.

Beer, R. D. 1996 Toward the evolution of dynamical neural networks for minimally cognitive behavior. In *From Animals to Animats 4. Proc. 4th Int. Conf. on Simulation of Adaptive Behavior* (ed. P. Maes, M. J. Mataric, J.-A. Meyer, J. B. Pollack & S. W. Wilson), pp. 421–429. Cambridge, MA: MIT Press.

Bi, G. Q. & Poo, M. M. 1998 Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and post-synaptic cell type. *J. Neurosci.* **18**, 10 464–10 472.

Bi, G. Q. & Poo, M. M. 2001 Synaptic modifications by correlated activity: Hebb's postulated revisited. *A. Rev. Neurosci.* **24**, 139–166.

Chechik, G. 2003 Spike-timing dependent plasticity and relevant mutual information maximization. *Neural Computation* **15**, 1481–1510.

Di Paolo, E. A. 2000 Homeostatic adaptation to inversion of the visual field and other sensorimotor disruptions. In *From Animals to Animats 6. Proc. 6th Int. Conf. on Simulation of Adaptive Behavior, Paris, France* (ed. J.-A. Meyer, A. Berthoz, D. Floreano, H. Roitblat & S. Wilson). Cambridge, MA: MIT Press.

Di Paolo, E. A. 2003 Spike-timing dependent plasticity for evolved robots. *Adapt. Behav.* **10**, 243–263.

Floreano, D. & Mattiussi, C. 2001 Evolution of spiking neural controllers for autonomous vision-based robots. In *Evolutionary Robotics IV* (ed T. Gomi). Springer.

Floreano, D. & Urzelai, J. 2000 Evolutionary robots with on-line self-organization and behavioral fitness. *Neural Netw.* **13**, 431–443.

French, R. L. B. & Damper, R. I. 2002 Evolving a circuit of spiking neurons for phototaxis in a Braitenberg vehicle. In *From Animals to Animats 7. Proc. 7th Int. Conf. on the Simulation of Adaptive Behavior, Edinburgh, UK* (ed. J. Hallam, D. Floreano, B. Hallam, G. Hayes, J.-A. Meyer & S. Wilson), pp. 335–344. Cambridge, MA: MIT Press.

Gerstner, W., Kreiter, A. K., Markram, H. & Herz, A. V. M. 1997 Neural codes: firing rates and beyond. *Proc. Natl Acad. Sci. USA* **94**, 12740–12741.

Hopfield, J. J. & Brody, C. D. 2001 What is a moment? Transient synchrony as a collective mechanism for spatiotemporal integration. *Proc. Natl Acad. Sci. USA* **98**, 1282–1287.

Horn, D., Levy, N. & Ruppin, E. 1998 Memory maintenance via neuronal regulation. *Neural Comput.* **10**, 1–18.

Husbands, P., Smith, T., Jakobi, N. & O'Shea, M. 1998 Better living through chemistry: evolving GasNets for robot control. *Connection Sci.* **10**, 185–210.

Kempter, R., Gerstner, W. & van Hemmen, J. L. 1999 Hebbian learning and spiking neurons. *Phys. Rev.* E **59**, 4498–4514.

Maass, W. 1997 Networks of spiking neurons: the third generation of neural network models. *Neural Netw.* **10**, 1656–1671.

Maass, W., Natschläger, T. & Markram, H. 2002 Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Comput.* **14**, 2531–2560.

Markram, H., Lubke, J., Frotscher, M. & Sakmann, B. 1997 Regulation of synaptic efficacy by coincidence of post-synaptic APs and EPSPs. *Science* **275**, 213–215.

Mehta, M. R., Barnes, C. A. & McNaughton, B. L. 1997 Experience-dependent asymmetric expansion of hippocampal place fields. *Proc. Natl Acad. Sci. USA* **94**, 8918–8921.

Mehta, M. R., Quirk, M. C. & Wilson, M. A. 2000 Experience-dependent asymmetric shape of hippocampal receptive fields. *Neuron* **25**, 707–715.

Mehta, M., Lee, A. K. & Wilson, M. A. 2002 Role of experience and oscillations in transforming a rate code into a temporal code. *Nature* **417**, 741–746.

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

Rao, R. P. N. & Sejnowski, T. J. A. 2001 Spike-timing-dependent Hebbian plasticity as temporal difference learning. *Neural Comput.* **13**, 2221–2237.

Rubin, J., Lee, D. D. & Sompolinsky, H. 2001 Equilibrium properties of temporally asymmetric Hebbian plasticity. *Phys. Rev. Lett.* **86**, 364–367.

Ruppin, E. 2002 Evolutionary autonomous agents: a neuroscience perspective. *Nat. Rev. Neurosci.* **3**, 132–141.

Song, S., Miller, K. D. & Abbott, L. F. 2000 Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nature Neurosci.* **3**, 919–926.

Stopfer, M., Bhagavan, S., Smith, B. H. & Laurent, G. 1997 Impaired odour discrimination on desynchronization of odour-encoding neural assemblies. *Nature* **390**, 70–74.

Sutton, R. S. 1988 Learning to predict by the method of temporal differences. *Mach. Learn.* **3**, 220–224.

Tuci, E., Quinn, M. & Harvey, I. 2003 An evolutionary ecological approach to the study of learning behaviour using a robot based model. *Adapt. Behav.* **10**, 201–221.

Turrigiano, G. G. 1999 Homeostatic plasticity in neuronal networks: the more things change, the more they stay the same. *Trends Neurosci.* **22**, 221–227.

Turrigiano, G. G., Leslie, K. R., Desai, N. S., Rutherford, L. C. & Nelson, S. B. 1998 Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature* **391**, 892–896.

van Rossum, M. C. W., Bi, G. Q. & Turrigiano, G. G. 2000 Stable Hebbian learning from spike-timing dependent plasticity. *J. Neurosci.* **20**, 8812–8821.

Walter, W. G. 1951 A machine that learns. *Scient. Am.* **185**, 60–63.

Walter, W. G. 1953 *The living brain*. London: Duckworth.

Webb, B. 1995 Using robots to model animals: a cricket test. *Robot. Auton. Syst.* **16**, 117–134.

Webb, B. 2000 What does robotics offer animal behaviour? *Anim. Behav.* **60**, 117–134.

Yamauchi, B. & Beer, R. D. 1994 Sequential behavior and learning in evolved dynamical neural networks. *Adapt. Behav.* **2**, 219–246.

Yao, H. & Dan, Y. 2001 Stimulus timing-dependent plasticity in cortical processing of orientation. *Neuron* **32**, 315–323.