Fast heterosynaptic learning in a robot food retrieval task inspired by the limbic system

Bernd Porr [a] & Florentin Wörgötter [b,c]

[a] *Department of Electronics & Electrical Engineering, University of Glasgow, Glasgow, GT12 8LT, UK*
[b] *Bernstein Center of Computational Neuroscience, University Göttingen, Germany*
[c] *Computational Neuroscience, Psychology, University of Stirling, FK9 4LR Stirling, UK*

Abstract

Hebbian learning is the most prominent paradigm in correlation based learning: If pre- and postsynaptic activity coincides the weight of the synapse is strengthened. Hebbian learning however, is not stable because of an autocorrelation term which causes the weights to grow exponentially. The standard solution would be to compensate the autocorrelation term. However, in this work we present a heterosynaptic learning rule which does not have an autocorrelation term and therefore does not show the instability of Hebbian learning. Consequently our heterosynaptic learning is much more stable than the classical Hebbian learning. The performance of our learning rule is demonstrated in a model which is inspired by the limbic system where an agent has to retrieve food.

Key words: Heterosynaptic learning, Hebbian learning, Limbic system, Robotics, Closed loop

1. Introduction

Hebbian type plasticity at a synapse correlates presynaptic with postsynaptic activity. Such learning is inherently unstable because of its *autocorrelation* term in the learning rule: weight growth will cause a higher postsynaptic potential and therefore even more weight growth. Additional measures have to be taken to prevent unlimited weight growth (Oja, 1982; Bienenstock et al., 1982; Verschure and Coolen, 1991). In this study we present a novel heterosynaptic learning rule which has been derived from our differential Hebbian learning rule isotropic sequence order (ISO) learning (Porr and Wörgötter, 2003). Our new learning rule remains stable without any additional measures. In this article we will show that our new heterosynaptic learning rule is much faster and more stable than the corresponding Hebbian learning rule. In addition, we will demonstrate the applicability of the rule first with a simulated and then in a real robot. The robot has to find food disks from the distance. Initially the robot has only a pre-wired reflex which enables it to react to food disks at close range only. During learning this reflex reaction is correlated with distant stimuli which enable the robot to target food disks from the distance. Finally, we will show similarities to the limbic system and argue that motor learning in this brain area might be driven by heterosynaptic learning.

Email addresses: b.porr@elec.gla.ac.uk (Bernd Porr), worgott@chaos.gwdg.de (Florentin Wörgötter).

2. Open loop: Heterosynaptic learning

2.1. *The neural circuit*

The dotted box in Fig. 1 shows the basic components of the neural circuit. The learner consists of two inputs x_0 and x_1 which are filtered with $u_0 = x_0 * h_0$ and $u_1 = x_1 * h_1$ where $*$ denotes the con-

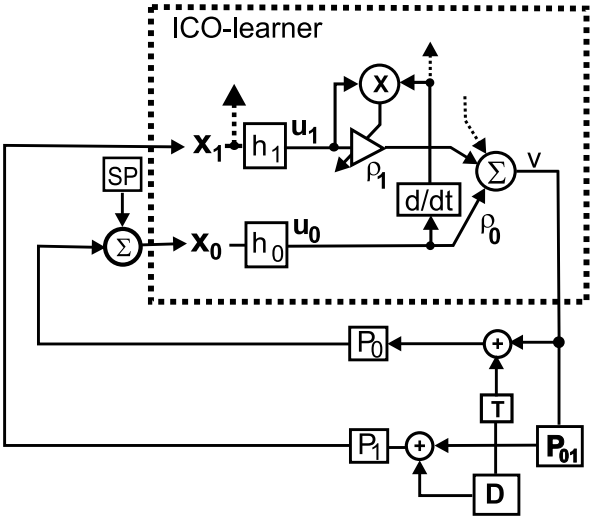

Fig. 1. Input correlation (ICO) learning in the closed loop: x_0, x_1 are sensor inputs, h_k are resonators and ρ_k are weights. The symbol d/dt denotes the derivative. $\otimes$ is a correlator and $\sum$ is a summation node. SP is the setpoint. D is a disturbance. P_0, P_{01}, P_1 are environmental transfer functions. T is a delay.

volution operator and the filters h_0 and h_1 basically smooth out the input signals. The circuit can easily be extended to a bank of filters with different resonators $h_j, j > 0$ and individual weights $\rho_j, j > 0$ to generate complex shaped responses (Grossberg, 1995). The individual filters thereby generate traces with different duration so that their superposition is able to generate coordinated behaviour in response to a stimulus.

To make our new heterosynaptic learning rule comparable to our older ISO-learning we will employ mostly resonators but at a later stage in the work we will also use other filter-functions. These resonators generate a damped oscillation when triggered by a delta pulse. In discrete time the resonator responses are given by: $h(n) = \frac{1}{b} e^{an} \sin(bn)$. The index for the time steps is n. The parameters are defined as $a = \mathrm{Re}(p) = -\pi f/Q$ and $b = \mathrm{Im}(p) = \sqrt{(2\pi f)^2 - a^2}$ respectively where Q is the quality of the filter.

2.2. *The learning rule*

The learning rule for the weight change $\frac{d}{dt}\rho_j$ is:

$$\rho_j' = \mu u_j u_0' \qquad j > 0 \tag{1}$$

where only *input signals are correlated with each other*. For that reason we will call our rule input correlation (ICO) learning. A sequence of events $x_1 \rightarrow x_0$ leads to a weight increase at ρ_1, whereas the reverse sequence $x_0 \rightarrow x_1$ leads to a decrease. The weights stabilise if the input x_0 is constant on average (or if x_1 is zero).

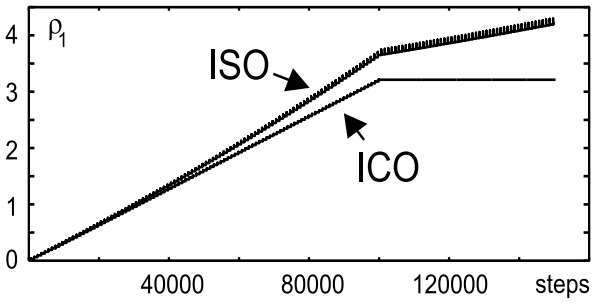

Fig. 2. Behaviour of the only ($N = 1$) weight ρ_1 for ICO-learning as compared to ISO-learning while stimulating first x_1 and then x_0 with delta pulses which are $T = 25$ apart. The pulse-sequence was repeated every 2000 time steps until step 100, 000. After step 100, 000 only input x_1 receives delta pulses. The learning rate was $\mu = 0.001$ and the parameters of the resonators were $f_0 = f_1 = 0.01, Q_0 = Q_1 = 0.6$.

Fig. 2 shows the behaviour of ICO-learning compared to ISO-learning for a relatively high learning rate for two filters ($N = 1$). Clearly one sees that differential Hebbian learning (ISO-learning) contains an exponential instability, which leads to an upward bend. For a more detailed discussion of this instability we refer the reader to Porr and Wörgötter (2003). This is different for heterosynaptic learning (ICO-learning) which does not contain this instability because it does not use the output to change its weight ρ_1.

3. Closed loop: Input control

3.1. *ICO-learning embedded in the environment*

ICO-learning only makes sense in a closed loop system where the output of the learner v feeds back to its inputs x_j after being modified by the environment (Fig. 1). The resulting structure, similar to that described in Porr et al. (2003), is that of an subsumption architecture where we start with an inner feedback loop which is superseded by an outer loop (Brooks, 1989). For a more detailed discussion of such nested structure we refer to Porr et al. (2003).

Initially only a stable reflex or *feedback loop* exists which is established by the two transfer functions H_0, P_0 and the weight $\rho_0 \neq 0$. The feedback loop has the task to maintain the setpoint SP as precisely as possible. A disturbance D causes deviations from

the setpoint but stability requires that the feedback loop must be able to bring the system back to its setpoint. In a behavioural context a disturbance is, for example, finding food by random walks.

From the point of view of the feedback it is desirable to *predict* the disturbance D to preclude the unwanted triggering of the feedback loop (Palm, 2000). Fig. 1 accommodates this in the most general way by a formal "delay" parameter T, which assures that the input x_1 receives the disturbance D earlier than input x_0. This establishes a second *predictive loop*, which is inactive at the start of learning ($\rho_1 = 0$). The learning goal is to find values for $\rho_j, j > 0$ so that the learner can use the earlier signal at x_1 to generate an anticipatory reaction which prevents x_0 from deviating from the setpoint SP. In the case of the food an anticipatory reaction can be generated if the food can be seen from the distance and then a reaction towards the food is generated.

3.2. *Benchmark: The simulated robot*

This section presents a benchmark application which compares Hebbian (ISO) learning with our new heterosynaptic (ICO) learning rule. Fig. 3A shows the circuit diagram of a simulated robot which has the task to learn to retrieve white "food disks" in a black arena. The reflex (via x_0) is established by two proximal sensors (LD) which draw the robot into the centre of the white disks. Learning has the task to use the distal sensors (SD) which feed into x_1 to generate an anticipatory reaction towards the "food disk". This simple benchmark application has already been used in Porr and Wörgötter (2003) and Verschure et al. (2003), however, with Hebbian learning only.

We quantify successful and unsuccessful learning for increasing learning rates μ. Learning was considered successful when we received a sequence of four contacts with the disk at a sub-threshold value of $|x_0| < 0.2$. We recorded the actual number of contacts until this criterion was reached. The log-log plots of the number of contacts in Fig. 3D,E show that both rules follow a power law. The similarity of the curves for small learning rates reflects the mathematical equivalence of both rules for $\mu \to 0$. For low learning rates weight changes are so slow that they do not cause a substantial change in the output v *during* learning so that u_0' can be replaced by the derivative by the output v' which means that we turn ICO-learning (Eq. 1) back into ISO-learning ($\rho_j' = \mu u_j v', j > 0$).

The dependence of failures on the learning rate is quite different for ISO- as compared to ICO-learning. For differential Hebbian (ISO) learning (Fig. 3G), errors increase roughly exponentially up to a learning rate of $\mu = 3 \cdot 10^{-4}$ while saturating at even higher learning rates. This behaviour reflects errors caused by the autocorrelation terms. For ICO-learning (Fig. 3F) failures remain essentially zero up to $\mu = 1 \cdot 10^{-4}$; the learned behaviour diverges only above that value. In contrast to the ISO-rule, this effect is here due to "over-learning" where the learning gain of the predictive pathway is higher than the gain of the feedback loop. Thus, the predictive pathway becomes unstable during the first learning experience.

In addition we have plotted the weight development for ICO- and ISO-learning, respectively (Fig. 3B,C). Looking at ISO learning (Fig. 3C) it is apparent that there is always a significant weight drift due to self amplification of the weights. ICO-learning, on the other hand is stable.

In summary the simulations demonstrate that ICO-learning is much more stable than the Hebbian ISO-learning. Consequently, ICO-learning can operate at more than ten times higher learning rates than ISO-learning.

3.3. *The real robot*

In this section we will demonstrate that ICO-learning is also able to master the "food-disk" targeting in an physically embodied agent (Ziemke, 2001). ISO-learning fails here completely (data not shown) because of its destabilising autocorrelation terms (see Fig. 2) which drive the weights either very quickly to infinity or, alternatively, one has to run the robot for hours to see anticipatory behaviour which is impractical.

In addition, we will show that it is possible to use other filters than resonators in the predictive pathway.

As before, the task of the robot is to target a white disk from a distance. As in the simulation the robot has a reflex reaction which pulls the robot into the white disk just at the moment the robot drives over the disk (Fig. 4A1). This reflex reaction is achieved by analysing the bottom scanline of a camera with a fisheye lens mounted on the robot. The predictive pathway is created in a similar way: A scanline which views the arena at a greater distance

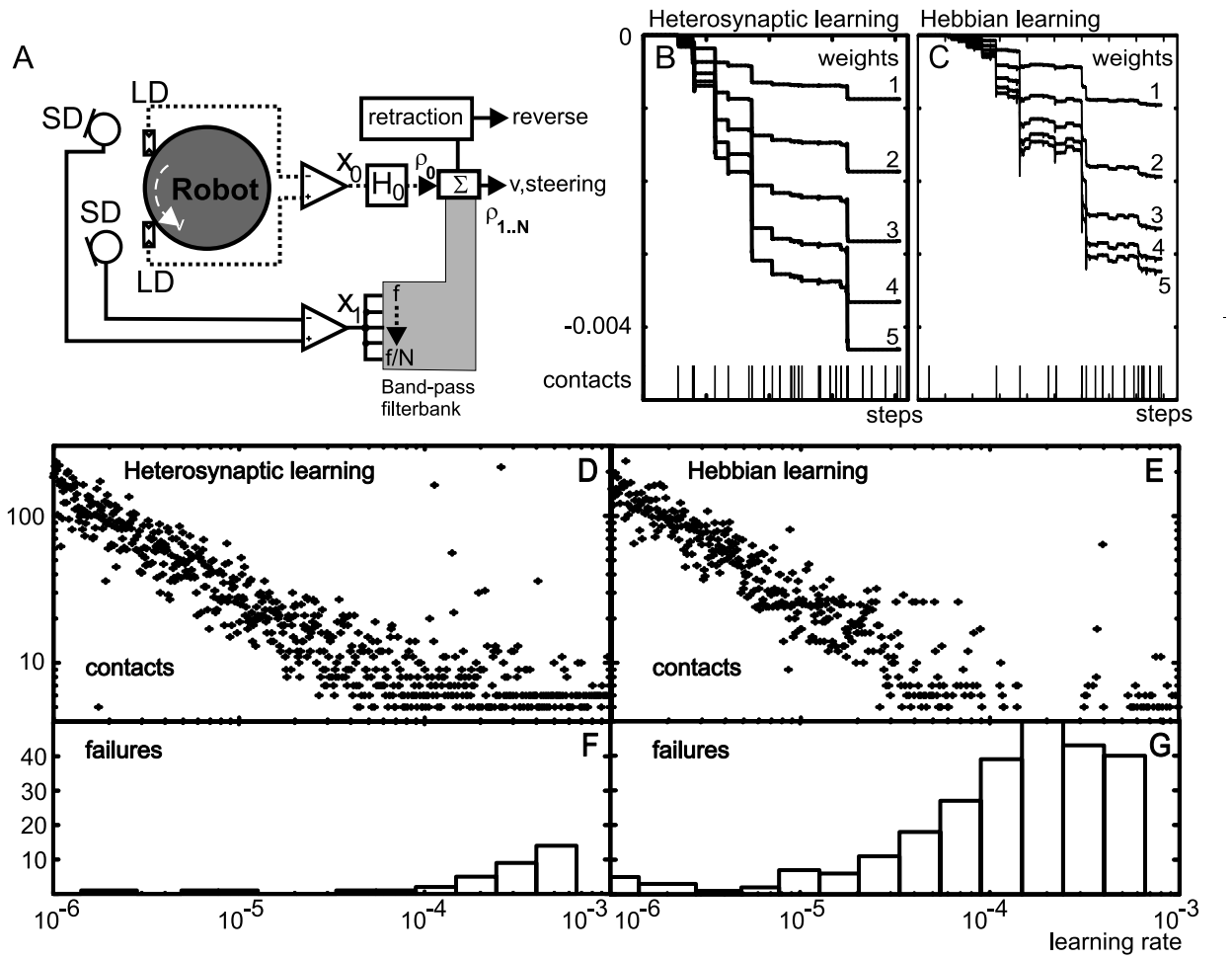

Fig. 3. The robot simulation. A) The robot has two light detectors (LD) which establish with the filter H_0 and the fixed weight ρ_0 the reflex reaction. The sound detectors (SD) establish with a filter bank the predictive loop. The weights $\rho_1 \dots \rho_N$ are variable and are changed either by ISO- or ICO-learning. The robot has also a simple retraction mechanism when it collides with a wall ("retraction") which is not used for learning. The output v is the steering angle of the robot. Filters are set to $f_0 = 0.01$ for the reflex, $f_j = 0.1/j, j = 1 \dots 5$ for the filter bank where $Q = 0.51$. Reflex gain was $\rho_0 = 0.005$. B) and C) show weight development for ICO- and ISO-learning at a learning rate of $\mu = 10^{-5}$. D) and E) plot the number of contacts for both learning rules needed for successful learning against the learning rate. F) and G) document the number of failures against the learning rate.

from the robot (hence "in its future") is fed into a bank of of five filters. This enables the robot to learn to drive *towards* the white disk (Fig. 4A2). In contrast to the simulation these filters are set up as *FIR filters*. Learning is successful and, except of fluctuations, the weights stay stable.

4. Limbic system

The way the limbic system operates seems to be closely related to ICO-learning. The limbic system is known for its role in the mediation of motivation and reward (Cardinal et al., 2002; Berthoud, 2005). We will focus on the nucleus accumbens core (NAcc) as the central integrator of sensor and motor information. This structure is considered important in the co-ordination of simple goal directed or instrumental behaviours (Killcross and Coutureau, 2003), like finding food. Plasticity in the NAcc is modulated by dopamine which originates from the ventral tegmental area (VTA) which in turn is innervated by the lateral hypothalamus. The downstream outputs of the lateral hypothalamus control directly motor actions of eating. Beninger and Gerdjikov (2004) and Kelley (2004) have both proposed similar heterosynaptic learning rules which might drive learn-

ing in the limbic system. They have proposed that glutamatergic input mainly from the cortex has to coincide with dopaminergic transients coming from the VTA which in turn is driven by primary eating reflexes originating from the lateral hypothalamus (LH). More specifically, the activation of the NMDA channels by glutamate has to coincide with a transient dopaminergic activity. Such learning could be formalised in the following way:

$$\Delta\rho = \mu \cdot NMDA \cdot DA' \tag{2}$$

whereas ρ is the synaptic weight of a glutamatergic input, μ is the learning rate, $NMDA$ is the activity of the NMDA channel and DA is the activity

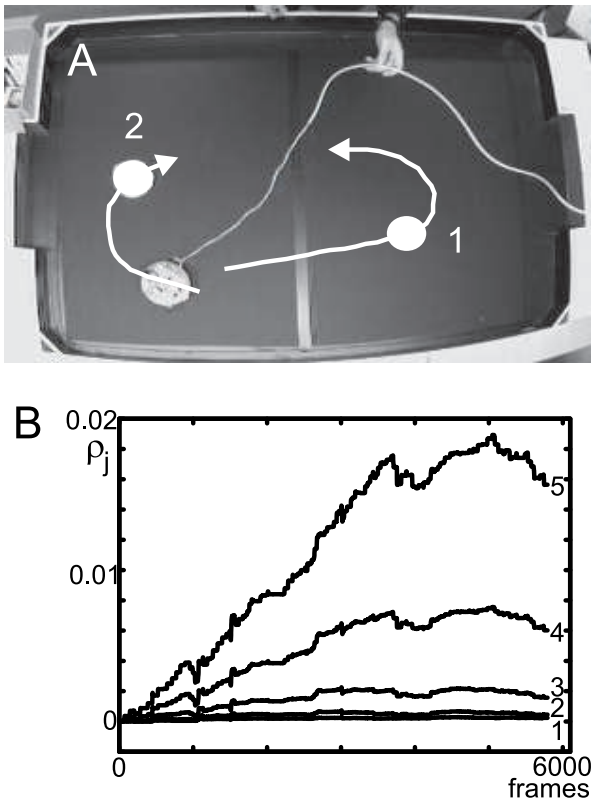

Fig. 4. Experiment with a real robot. A1: start of the run, A2: after 15 mins and 52 secs. The arrows at A1 and A2 show the trace of the robot while driving into food blobs (white circles). The weight development ($\rho_j, j = 1 \ldots 5$) is shown in B. Parameters: frame rate was 25 frames/ses. The video image $f(x = [0 \ldots 95], y = [0 \ldots 64])$ was evaluated at $y = 53$ for the reflex x_0 and at $y = 24$ for the predictive signal x_1. Reflex and predictive signal were calculated as a thresholded (> 240) weighted sum: $x_{0,1} = \sum_{x=0}^{95}(x - 96/2)^2\Theta(f(x, y))$. The reflex pathway was set to: $f_0 = 0.01, Q = 0.51$ with a reflex gain of $\rho_0 = 30$. The predictive finite impulse response (FIR) filters had $100, 50, 33, 25, 20$ taps where all coefficients are set to one. The learning rate was $\mu = 0.000005$.

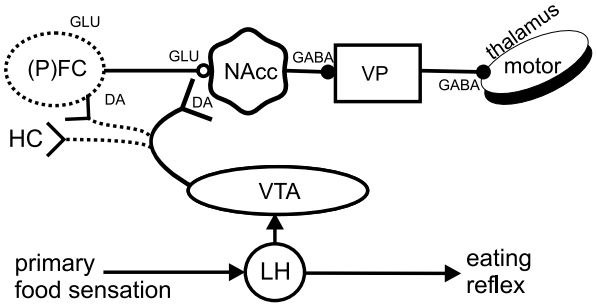

Fig. 5. Simplified diagram of the limbic system. NAcc=Nucleus Accumbens core, PFC=prefrontal cortex, VP=ventral pallidum, VTA=ventral tegmental area, LH=lateral hypothalamus.

of the dopaminergic neurons originating from the VTA. Our formalisation of Beninger's heterosynaptic learning (Eq. 2) is identical to our ICO-learning (Eq. 1) which demonstrates the similarities between ICO-learning and plasticity in the limbic system.

5. Discussion

To our knowledge ICO-is the only learning rule that operates strictly heterosynaptically. The neuronal literature on heterosynaptic plasticity normally emphasises that it is essentially a modulatory process which modifies (conventional) homosynaptic learning (Bliss and Lomo, 1973; Markram et al., 1997), but cannot lead to plasticity on its own (Bailey et al., 2000; Jay, 2003). As a consequence, heterosynaptic learning rules have so far mostly been used to emulate modulatory processes, for example, by the implementation of three-factor learning rules, trying to capture dopaminergic influence in the Striatum and the Cortex (Schultz and Suri, 2001).

ICO-learning bears some similarities to spike timing dependent plasticity (STDP). In Saudargiene et al. (2004) we have shown that STDP can be modelled by correlating the activity of an NMDA channel with the derivative of the postsynaptic potential. Here we correlate the activity of an NMDA channel with the derivative of the dopaminergic activity. Consequently, we could call our ICO-learning here "heterosynaptic spike timing dependent plasticity".

In the closed loop case the difference between ICO-learning and the classical Hebbian learning rules becomes even clearer: While classical Hebbian learning and STDP can be used in both open loop (Oja, 1982) and closed loop scenarios (Verschure and Coolen, 1991; Porr et al., 2003), ICO-learning

needs feedback from the environment. Otherwise ICO-learning is not able to reach its learning goal, namely to minimise the error at its input x_0. In this respect ICO-learning is different from other learning rules which calculate an error by comparing the output signal v with a desired response, like the delta rule (Widrow and Hoff, 1960) or temporal difference (TD) learning (Sutton, 1988).

Acknowledgements

We thank the referees for their very helpful feedback.

References

Bailey, C. H., Giustetto, M., Huang, Y. Y., Hawkins, R. D., Kandel, E. R., Oct 2000. Is heterosynaptic modulation essential for stabilizing Hebbian plasticity and memory? Nat Rev Neurosci 1 (1), 11–20.

Beninger, R., Gerdjikov, T., 2004. The role of signaling molecules in reward-related incentive learning. Neurotoxicity Research 6 (1), 91–104.

Berthoud, H., May 2005. Brain, appetite and obesity. Physiol Behav 85 (1), 1–2.

Bienenstock, E., Cooper, L., Munro, P., 1982. Theory for the development of neuron selectivity, orientation specifity and binocular interpretation in visual cortex. J. Neurosci 2, 32–48.

Bliss, T., Lomo, T., 1973. Long-lasting potentiation of synaptic transmission in the dentrate area of the anaesthetized rabbit following stimulation of the perforant path. J Physiol 232 (2), 331–356.

Brooks, R. A., 1989. How to build complete creatures rather than isolated cognitive simulators. In: VanLehn, K. (Ed.), Architectures for Intelligence. Erlbaum, Hillsdale, NJ, pp. 225–239.

Cardinal, R., Parkinson, J., Hall, J., Everitt, B., May 2002. Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. Neurosci Biobehav Rev 26 (3), 321–352.

Grossberg, S., 1995. A spectral network model of pitch perception. J Acoust Soc Am 98 (2), 862–879.

Jay, T., Apr 2003. Dopamine: a potential substrate for synaptic plasticity and memory mechanisms. Prog Neurobiol 69 (6), 375–390.

Kelley, A. E., 2004. Ventral striatal control of appetitive motivation: role in ingestive behaviour and reward-related learning. Neuroscience and Biobehavioural Reviews 27, 765–776.

Killcross, S., Coutureau, E., Apr 2003. Coordination of actions and habits in the medial prefrontal cortex of rats. Cereb Cortex 13 (4), 400–408.

Markram, H., Lübke, J., Frotscher, M., Sakman, B., 1997. Regulation of synaptic efficacy by coincidence of postsynaptic aps and epsps. Science 275, 213–215.

Oja, E., 1982. A simplified neuron model as a principal component analyzer. J Math Biol 15 (3), 267–273.

Palm, W. J., 2000. Modeling, Analysis and Control of Dynamic Systems. Wiley, New York.

Porr, B., von Ferber, C., Wörgötter, F., 2003. Iso-learning approximates a solution to the inverse-controller problem in an unsupervised behavioural paradigm. Neural Comp. 15, 865–884.

Porr, B., Wörgötter, F., 2003. Isotropic Sequence Order learning. Neural Comp. 15, 831–864.

Saudargiene, A., Porr, B., Wörgötter, F., 2004. How the shape of pre- and postsynaptic signals can influence stdp: A biophysical model. Neural Comp. 16, 595–626.

Schultz, W., Suri, R. E., 2001. Temporal difference model reproduces anticipatory neural activity. Neural Comp. 13 (4), 841–862.

Sutton, R., 1988. Learning to predict by method of temporal differences. Machine Learning 3 (1), 9–44.

Verschure, P., Coolen, A., 1991. Adaptive fields: Distributed representations of classically conditioned associations. Network 2, 189–206.

Verschure, P. F. M. J., Voegtlin, T., Douglas, R. J., 2003. Environmentally mediated synergy between perception and behaviour in mobile robots. Nature 425, 620–624.

Widrow, G., Hoff, M., 1960. Adaptive switching circuits. IRE WESCON Convention Record 4, 96–104.

Ziemke, T., 2001. Are robots embodied? In: First international workshop on epigenetic robotics Modeling Cognitive Development in Robotic Systems. Vol. 85. Lund.