Electronic version of an article published as Advances in Complex Systems,
Vol. 12, No. 3 (2009) 273-291, DOI: 10.1142/S0219525909002258

MODULATED EXPLORATORY DYNAMICS CAN SHAPE SELF-ORGANIZED BEHAVIOR

FRANK HESSE *
Max-Planck-Institute for Dynamics and Self-Organization,
Bernstein Center for Computational Neuroscience Göttingen and
Department for Nonlinear Dynamics, Georg-August-University Göttingen,
Bunsenstrasse 10, 37073 Göttingen, Germany
frank@nld.ds.mpg.de

RALF DER
Max-Planck-Institute for Mathematics in the Sciences,
Inselstrasse 22, 04103 Leipzig, Germany
ralfder@mis.mpg.de

J. MICHAEL HERRMANN†
University of Edinburgh, School of Informatics,
10 Crichton Street, Edinburgh, EH8 9AB, Scotland, UK
mherrman@inf.ed.ac.uk

July 22, 2009

Abstract

We study an adaptive controller that adjusts its internal parameters by self-organization of its interaction with the environment. We show that the parameter changes that occur in this low-level learning process can themselves provide a source of information to a higher-level context-sensitive learning mechanism. In this way the context is interpreted in terms of the concurrent low-level learning mechanism. The dual learning architecture is studied in realistic simulations of a foraging robot and of a humanoid hand that manipulated an object. Both systems are driven by the same low-level scheme, but use the second-order information in different ways. While the low-level adaptation continues to follow a set of rigid learning rules, the second-order learning modulates the elementary behaviors and affects the distribution of the sensory inputs via the environment.

1 Introduction

Homeokinesis [11, 5, 6] is based on the dynamical systems approach to robot control cf. e.g. [18, 19, 22] that may be understood as a dynamical counterpart of the principle of homeostasis [3, 1]. According to the latter, behavior is understood to result from the compensation

*corresponding author
†J.M.H. is also PI at the Bernstein Center for Computational Neuroscience Göttingen

Electronic version of an article published as Advances in Complex Systems,
Vol. 12, No. 3 (2009) 273-291, DOI: 10.1142/S0219525909002258

of perturbations of an internal homeostatic equilibrium. Although this approach proves successful in simple systems [12, 23], it remains elusive how it scales up to situations, where, e.g., internal nutrient levels in an agent are to give rise to the specific form of a resulting complex behavior.

Homeokinesis, in contrast, provides a mechanism for the self-organization of elementary movements that are not directly caused by a triggering stimulus, but are generated by amplification of intrinsic or extrinsic noise which is counterbalanced by the simultaneous maximization of the controllability of the resulting behavior. Homeokinesis provides a mechanism to produce behaviors that are not brought about by a reward signal or a prescribed target state, but that minimize an unspecific internal error functional [9, 7, 8, 10]. Homeostasis and homeokinesis are complementary principles rather than alternatives. While homeokinetic control serves to explore a behavioral manifold, homeostasis refers to the stationary state of a highly optimized system which, however, might be reached or re-approached by a coordinated exploration. It is also interesting to compare homeokinetic learning to reinforcement learning [20, 16]. While reinforcement learning tries to maximize the expected reward for a sequence of single actions, homeokinetics selects coherent sequences which may later be used as building blocks in a higher-order reinforcement learning algorithm.

The present paper explores the relation between homeostasis and homeokinesis in a specific case where certain exploratory parameter changes reoccur sufficiently often such that the result becomes predictable. In this way may the system be enabled to avoid the cause of this parameter change in favor of a more stable internal configuration. The obvious effect that the avoided situation deprives the system from informative signals is compensated by the persistence of the exploratory effects that eventually cause new encounters with the critical stimuli. We suggest an extension of the homeokinetic controller, where in addition to the minimization of temporally local errors also predictable future learning signals are taken into account. Predictability may be achieved by analyzing the time series of some sensory inputs that are processed in the sensorimotor loop while other sensors or other types of sensors could be available as *context* information. In addition to the low-level homeokinetic control a second layer will be adapted for feeding such context information in a suitable way into the first layer. We will illustrate this two-layer architecture by the example of a robot that changes its internal parameters in order to escape from stalling in front of an obstacle based on low level control. Subsequently it learns to avoid the collisions by advancing the wall-related parameter changes to the time before the collision with the help of the higher control layer.

In a sense the higher-order (predictive) learning receives its learning signals from the performance of the lower-level (homeokinetic) learning. The same effect may as well be chosen such that noticeable parameter changes increase the probability of remaining in a particular state as we demonstrate in a second example. Here a simulated human-like hand is shown to produce a gripping reflex by the information that is produced by the low-level learning.

Low-level reflexes that are produced in the current model by a homeokinetic controller can be interpreted by high-level structures in different ways. We posit that a main goal of the interference by the high-level control consists in the modulation of the distribution of sensory inputs to the low-level control system. If low-level errors are interpreted as risky, the high-level control should succeed in avoiding situations where these errors occur. In a different set-up, errors form a challenge to the insufficient internal model of the agent that may be adapted more effectively if the frequency of errors is increased. We will consider both schemes in some detail while other examples are mentioned in passing. Before this, we will present a brief summery on the principle of homeokinesis in the next section. A learning rule is derived

Electronic version of an article published as Advances in Complex Systems,
Vol. 12, No. 3 (2009) 273-291, DOI: 10.1142/S0219525909002258

from this principle in Section 3. Section 4 describes the second-order learning. Experimental results in the learning architecture in realistic simulations are presented in Sections 5 and 6.

2 Homeokinesis

Considering an adaptive control system, we will discuss requirements for an unspecific objective function that allows a robot to acquire autonomously reproducible relations between actions and environmental feedback.

The behavior of the robot is governed by a controller K that generates motor outputs

$$y_t = K\left(x_t; c\right) \tag{1}$$

as a function of the vector of sensory inputs $x_t = \{x_{t,1}, \ldots, x_{t,d}\}$. The effect of the controller depends on a parameter vector $c = \{c_1, \ldots, c_n\}$. For example, the c_i may represent weights of the input channels from which K calculates a squashed sum for each output. We further assume that the inputs are processed on a time scale that is short compared to the cycle time. Adaptivity of the controller is achieved based on an objective function E that contains possibly implicit information about desired behavior. In this sense it is sufficient to define E based on the probability of survival or, as we will proceed here, by a functional that is evaluated continuously by the robot itself.

Interaction of an agent with its environment includes sensitivity to sensory stimuli. The actions of the robot should appear to be caused by the inputs, although the particular form of the relation between inputs and outputs may change continuously due to the adaptation of the internal parameters. Whether or not a reaction of a robot is due to a particular stimulus cannot be decided by reinitialization in a similar perceptual situation if the robot is to learn autonomously. We therefore equip the robot with an internal model M mapping the motor command y_t to the sensor values x_{t+1}

$$x_{t+1} = M\left(y_t; a\right), \tag{2}$$

based on the parameters $a = \{a_i, \ldots, a_n\}$, which may represent weights of the input channels, from which M calculates a sum for each output. The model enables the robot to compare a situation with a similar one that was encountered earlier. If the robot's objective was solely the reproducibility of a reaction then the robot would tend to run into trivial behaviors. This could mean e.g. that the robot behaves such that the inputs remain as constant as possible, suppressing thus sensitive reactions. We focus therefore on the unavoidable differences between inputs and the corresponding predictions by the internal model. If these differences are small then the robot is obviously able to predict the consequences of its actions. If, in addition, the differences tend to increase due to new stimuli then the robot is also sensitive. Note that the differences can be decreased also by improving the internal model which induces in the robot a tendency to increase exploration beyond the growing region of predictable behavior.

The main idea of the approach derives from the fact that a destabilization in time is dynamically identical to a stabilization backward in time. Because predictability in a closed loop enforces stability, a virtual inversion of time is the road to a destabilization without the loss of predictability. This idea will become more clear in the following formal description. as a first step we introduce a virtual sensor value $\hat{x}$ by

$$\hat{x}_t = \arg\min_x \left\| x_{t+1} - \psi\left(x\right) \right\|, \tag{3}$$

Electronic version of an article published as Advances in Complex Systems,
Vol. 12, No. 3 (2009) 273-291, DOI: 10.1142/S0219525909002258

which is optimal with respect to the prediction of the following input x_{t+1}, although generally different from the real input x_t. The predictor

$$\psi = M(K(x); a), \tag{4}$$

cf. Eq. 2 is realized by a parametric function, e.g. an artificial neural network that receives the sensor vector x_t as an input and generates an estimate of the subsequent input x_{t+1}. In principle, the representation (4) of ψ comprises both the controller K and the internal model M, because the prediction of the future output necessarily depends on the current action. In this sense the function ψ depends on both parameter vectors c and a (1, 2), but assuming (cf. below) that the adaptation of a occurs on a slower time scale than that of c, we may drop the argument a in most cases. When we want to emphasize the dynamical systems aspect rather than the learning dynamics, we shall focus on the dynamical variables and omit both arguments c and a.

We can interpret the calculation of $\hat{x}_t$ as a mechanism for editing an earlier input which is invoked once x_{t+1} is available. In order to minimize the effect of the editing, one should require that

$$\|x_t - \hat{x}_t\| \to \min, \tag{5}$$

which actually turns out to be the central criterion of the approach. Eq. 3 is nothing but a regularized solution of the possibly ill-posed equation

$$\hat{x}_t = \psi^{-1}\left(x_{t+1}\right), \tag{6}$$

which reveals that essentially the inverse of the predictor ψ (4) is used in order to produce $\hat{x}_t$, see also Fig. 1. In this sense we consider the sensory dynamics in inverted time while avoiding a conflict with causality.

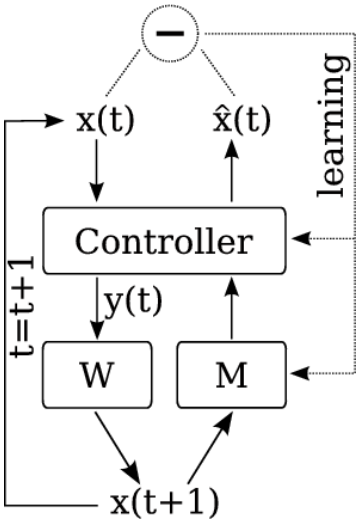

Figure 1: Scheme of the homeokinetic control scheme based on a sensorimotor loop. Sensor values $x(t)$ are used by the controller to generate motor commands $y(t)$ which are executed in the environment (W). Subsequently, new sensor values $x(t+1)$ become available. A world model (denoted by M) that realizes a function $\psi(x(t)) \approx x(t+1)$ (4) is simultaneously adapted. The goal of the parameter adaptation is to minimize the difference between the virtual and the true sensor value $x(t)$.

Electronic version of an article published as Advances in Complex Systems,
Vol. 12, No. 3 (2009) 273-291, DOI: 10.1142/S0219525909002258

Often instead of (5) the problem

$$\|x_{t+1} - \psi(x_t)\| \to \min \tag{7}$$

is considered which measures the forward prediction error $\|\xi\|$, where $\xi = x_{t+1} - \psi(x_t)$.

Eqs. 5 and 7 have in common that they optimize the predictability of sensory inputs by the predictor ψ (4). The dynamical effects are, however, quite different. (7) causes the controller to decrease the distance of any noisy trajectory from the predicted value. This leads generically to a convergence of nearby trajectories, i.e. tends to stabilize the current behavior. Condition (5), in contrast, causes nearby trajectories to diverge, which is, however, counterbalanced by the simultaneous optimization of predictability. In (5) we shall use the abbreviation $v_t = x_t - \hat{x}_t$ in order to denote the predictor-based sensory shift. It is used in the definition of an energy function

$$E = \|v\|^2 . \tag{8}$$

Because minimizing E minimizes the sensitivity of ψ^{-1}, cf. Eq. 6, the sensitivity of the function ψ with respect to variations of its arguments is maximized. The shift v is small if both $\xi = x_{t+1} - \psi(x_t)$ is small and the derivative of ψ is large. Hence, the two goals of sensitivity and predictability are implicit in (8). This becomes more obvious when studying the parameter dynamics in the approximation of small v. The sensor value x_{t+1} can be expressed in either way

$$\psi(x_t) + \xi = x_{t+1} = \psi(x_t + v). \tag{9}$$

The shift v can be used to generate a Taylor expansion of the environmental effects implicit in ψ

$$\psi(x_t + v) = \psi(x_t) + Lv, \tag{10}$$

where L is the Jacobian matrix of the system defined as $L_{ij} = \frac{\partial}{\partial x_j}\psi_i(x)$. Using Eq. 10 in (9) we find $v = L^{-1}\xi$. The energy function is then

$$E = \left(L^{-1}\xi\right)^T \left(L^{-1}\xi\right). \tag{11}$$

For this formulation of E is is immediately clear that that by gradient descent on E the modeling error ξ is decreased whereas the sensitivity of the system is increased by increasing the Jacobian L. However, an increase in sensitivity will tend to lower predictability and vice versa, such that the actual behavior can be expected to oscillate between periods of exploration and stabilization in a way which reflects the quality of the predictor and the complexity of the environment.

3 Learning rules for control

Because the cost function (11) depends via the behavior of the robot also on the controller parameters (1), adaptive parameter changes can be achieved by a gradient flow on $E(x_t, c_t)$, where x_t denotes the trajectory of the robot in the sensory state space and c_t the current values of the controller parameters. The combined dynamics

$$x_{t+1} = \psi(x_t; c_t) + \xi_t \tag{12}$$

$$c_{t+1} = c_t - \varepsilon_c \frac{\partial}{\partial c} E(x_t, c_t) \tag{13}$$

Electronic version of an article published as Advances in Complex Systems,
Vol. 12, No. 3 (2009) 273-291, DOI: 10.1142/S0219525909002258

describes both the effects of the environment and the controller on the sensory state of the agent as well as the adaptation of the controller parameters. The resulting state (12) and parameter dynamics (13) run concomitantly and form a dynamical system in the product space formed by x and c. Learning, in this sense, means to identify both a set of parameters as well as a region in x-space which are optimal with respect to E. It is possible that the learning process results in a limit cycle involving both parameters and states, or it may be even open-ended by allowing a robot to gradually explore a virtually unbounded environment.

The parameters a of the internal model M (2) predicting how sensor values x are influenced by controller outputs y and the environment, are adapted by supervised learning. Since training data are given by the pairs (y_t, x_{t+1}) the prediction error $\|x_{t+1} - \psi(x_t)\|$ is suitable to improve the model with learning rate ε_a. Note that the model adaptation has in principle an effect on the loop function ψ (4), but we should consider how the time scales of the adaptation processes are related. The ratio of ε_c in (13) and ε_a is crucial for the learning process. If $\varepsilon_a \approx \varepsilon_c$ the model is in principle able to track the changes in the behavior of the agent, while for $\varepsilon_a \ll \varepsilon_c$ the model is rather accumulating information about the environment. Note that ε_c must correspond to a time scale comparable to the maximal rate of change of the sensory inputs in order to allow for an immediate reaction.

Ignoring the effects of the nonlinearity in Eqs. 12 and 13 we find that the state in (12) is dominated by the eigenvector of L with largest eigenvalue. Therefore, a learning rule based on Eq. 7 reduces the maximal eigenvalue of L. A learning rule based on (8), (11) will instead tend to increase the minimal eigenvalue of L by minimizing the maximal eigenvalue of L^{-1}. In this way more and more modes become activated and, if the forward prediction error ξ does not increase, the behavior becomes sensitive with respect to many different stimuli.

In the following we will use a pseudo-linear controller K with $y = tanh(z)$, where $z = cx + h$ and the loop function is $\psi(x_t) = a \tanh(cx_t + h)$. For this controller Eq. 13 becomes

$$\Delta c = \mu a - 2\mu x (z - h) \tag{14}$$

$$\Delta h = -2\mu (z - h) \tag{15}$$

where a is the linear response of the environment to the action y obtained by the internal model $M(y_t) = a(y_t)$ and $\mu > 0$ is a modified learning rate including the energy function E (11). The update of the bias h has the opposite sign of z and hence of the motor command y. This leads to a hysteresis effect where the sign of Δh changes as soon as the motor command changes sign. The interesting point is once $|h| > |cx|$ (and $h\ y < 0$, that is h and y have opposite sign) the bias will initiate a change of the motor command. Since the learning rate μ of the bias is modulated by the energy function E, with varying E a robot will execute an irregular searching behavior.

4 Hebbian second-order learning

The homeokinetic controller (1, 14-15) generates simple reactive behaviors that are interesting because of their flexibility. We are now going to modify the controller such that in addition prospective information can be exploited in order to generate preventive action, thereby relying on other information which may be available from more complex sensors and predictors. We propose to interpret such information in terms of the low-level control which may be advantageous if no background information can be referred to for the interpretation of the high-level information.

We extend the homeokinetic controller by an additional learning mechanism in order to avoid situations that cause low-level errors. The homeokinetic controller will eventually lead

Electronic version of an article published as Advances in Complex Systems,
Vol. 12, No. 3 (2009) 273-291, DOI: 10.1142/S0219525909002258

the agent out of these situations but only by accumulating the error-related cost during the time the error is above the baseline. In reoccurring situations the error is certainly predictable, which suggests to add a predictive component to the homeokinetic controller. Rather than replacing the controller by an eventually fixed forward system we aim at retaining flexibility and merely advance the low-level error by estimating the future error of the low-level predictor ψ in a higher control layer.

The error function (11) that is minimized by the homeokinetic control layer is thus extended by an additive contribution from context sensors which can be represented by

$$E = \left(L^{-1}(\xi + \zeta)\right)^T \left(L^{-1}(\xi + \zeta)\right). \tag{16}$$

Here we introduced the prediction ζ of the error $\xi = x_{t+1} - \psi(x_t)$ which is calculated based on the context input. In place of the pseudo-linear regression that will be used here for simplicity, more complex predictors are certainly possible, but the more interesting option is to substitute ζ by an arbitrary reward signal that then would become effective in shaping the behavior. In order to interpret such a reward signal only its order of magnitude has to be known in advance while its size relative to previous time steps or relative to the level of ξ can be evaluated by the algorithm.

In just the same way does the sign of ζ affect the learning of the prediction of the forward error ξ. The ambivalence of the error suggests different choices of the way the predicted error enters the behavioral learning. A natural way is to use the additional sensory information once it is available in order to avoid regions which tend to generate large errors. If, to the contrary, the large errors indicate a learning task than the prediction of these errors can be used for guidance towards challenging regions in the environment. The idea to use the predicted decrease in the forward error, i.e. expected success of learning, has been put forward in Ref. [13].

The channel by which the context information arrives will be denoted by x^H, where the index already points to the present Hebbian formulation [14, 4] of the functional of the the influence of the context information on the low-level learning rule (16). The additional contribution to the cost function (16) will be provided by a supervening adaptable layer that naturally uses the Hebbian rule, cf. Eq. 18 and is trained on-line to predict ζ, cf. Fig. 2. The robot is now controlled such that in addition to the state estimation error ξ also the

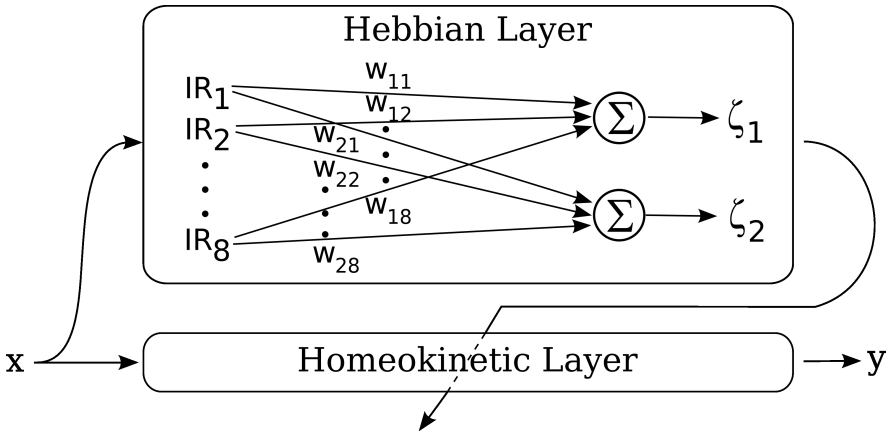

Figure 2: Scheme of the extended control structure. A Hebbian layer affects the homeokinetic controller by the predicted modeling error ζ which complements the low-level time-loop error.

Electronic version of an article published as Advances in Complex Systems,
Vol. 12, No. 3 (2009) 273-291, DOI: 10.1142/S0219525909002258

prediction ζ of the state estimation error is minimized.

Although ζ may be derived from any feature of the learning process such as learning progress or predictability of the sensory trajectory we found it interesting to consider the integration of additional signals x^H that are available to the agent via other sensory modalities. In this way, self-organizing control can be extended to applications in sensor fusion. The additional signals can be considered as a *context* with respect to the basic learning goal of the homeokinetic controller. For example, when the homeokinetic controller is driven by errors from proprioceptive sensors then the Hebbian layer may refer to external sensors. Another option would be to associate distance information with possibly complex visual input.

If the additional input x^H to the higher layer is unavailable ζ is defined to be zero. Similarly, if x^H and the low-level error ξ are not correlated then their effect on the weights (18) will essentially average out. In this way such situations will not be recognized by the Hebbian layer and the effect of ζ onto c will also remain unsystematic. In these cases the actual behavior is produced solely by the homeokinetic controller. Otherwise the controller will adapt such that both ξ and ζ are reduced. In order to retain the full flexibility of the low-level controller we do not allow the context parameter to interfere with the activity parameter c that induces an exploratory behavior if sufficiently large. If c were modifiable by the context then the robot might stop to forage and become 'stunned', which did not seem to be desirable for a robot although it is known from behavioral studies in a number of animals [24, 25, 26]. Instead the context-dependent energy function (16) applies only in the update rule of the threshold h (15), influencing the bias of the controller rather then leaving the high-gain regime. This is also compatible with earlier studies [6, 15] that have shown that the exploratory mode is characterized by a stationary non-zero c value while the threshold h continues to fluctuate.

The Hebbian layer is realized by a leaky integrator neuron with a linear output function for each of the sensors x of the homeokinetic layer. All additional sensory information x^H available to the higher layer is used as input to each neuron and weighted by the synaptic strength w_{ij} according to

$$\zeta_i = \sum_{j=1}^{m} w_{ij} x_j^H, \quad i = 1..n, \tag{17}$$

with m being the number of context sensors and n the number of sensors available to the homeokinetic layer. The update rule for the weights is

$$\Delta w_{ij} = \varepsilon \xi_i x_j^H (1 - w_{ij}^2), \tag{18}$$

where ε is a learning rate and $\xi_i x_j^H$ realizes Hebbian learning between the low-level error ξ_i of the homeokinetic layer and the input x_j^H of the Hebbian layer. A decay term $(1 - w_{ij}^2)$ is added which restrains the weights from unlimited increase. For this weight normalization to be effective we must assure that the initial values obey $\|w_{ij}\| < 1$ and that this property is not destroyed by the time discretization. We found still an advantage of the soft threshold in Eq. 18 as compared to a hard limit. Nevertheless, a sensory input of 1 weighted with a synaptic strength of nearly 1 would result in a predicted state estimation error ζ of about 1, which can realize an immediate change of the actual behavior as intended. By using the context-dependent energy function (16) the homeokinetic controller can thus avoid situations which lack low-level predictability.

Electronic version of an article published as Advances in Complex Systems,
Vol. 12, No. 3 (2009) 273-291, DOI: 10.1142/S0219525909002258

5 Foraging in a wheeled robot

In the general case we have a vector of sensor values $x_t \in \mathbb{R}^d$ at the discrete instants of time $t = 0, 1, 2, \ldots$. By way of example we may consider a two-wheeled robot, cf. Fig. 3, where the low-level controller receives the measured wheel velocities as input. In addition infrared sensors are available as context sensors.

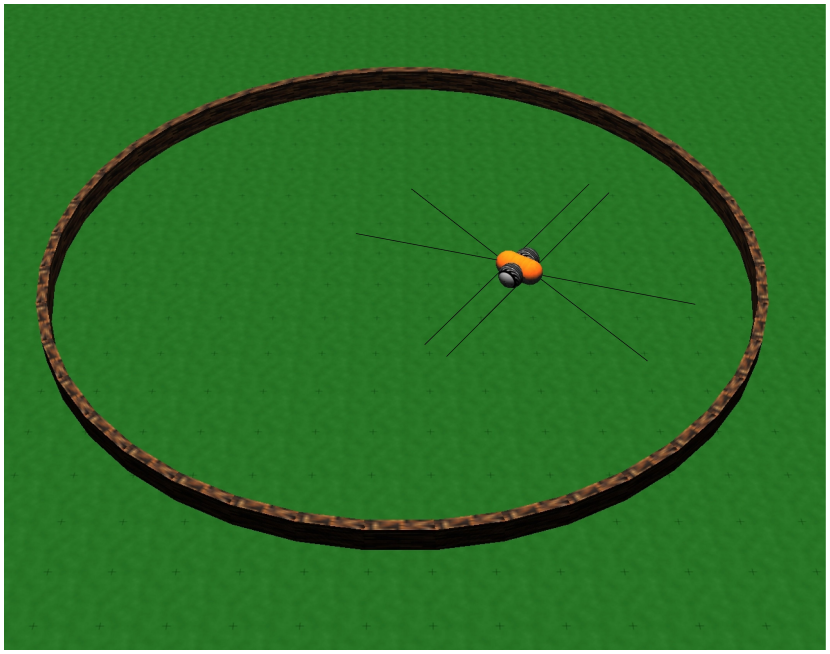

Figure 3: Experiments are performed with a two-wheeled robot in a circular arena. The robot is equipped with wheel counters and eight infrared sensors. The black lines indicate the infrared sensor orientation and range. The sensor range is three length units and the diameter of the arena is 14 length units.

The state estimation (or modeling) error ξ describes differences between predicted and measured wheel velocities. The predicted modeling error ζ is used to modulate the homeokinetic layer in order to change the actual behavior before arriving at situations with a large state estimation error, which refers to collision situations in the example. The setup of the experiments consists of a simulated two-wheeled robot with infrared sensors, placed in a circular arena, for details see Fig. 3.

In the experiments we will show that obstacle avoidance behavior of a two-wheeled robot equipped with infrared sensors can be obtained based solely on the intrinsic properties of the system. The effectiveness of the obstacle avoidance is not perfect since the system tries occasionally to explore also the regions near the boundaries. Nevertheless the time the robot spends near obstacles is strongly reduced, cf. Fig. 5. The Hebbian layer is provided with proximity information from eight infrared sensors with a sensor range of three length units. In order to suppress small noisy activity in the infrared sensors, only sensor values larger than 0.15 are considered. The synaptic strength w_{ij} of the Hebbian layer are initialized with zeros. The parameters of the homeokinetic layer are initialized with small random values.

In a first experiment only the homeokinetic layer was used while the second experiment was featuring the extended controller including the higher control layer. Each experiment runs for one hour of simulated real time in a simulation environment that is designed to

Electronic version of an article published as Advances in Complex Systems,
Vol. 12, No. 3 (2009) 273-291, DOI: 10.1142/S0219525909002258

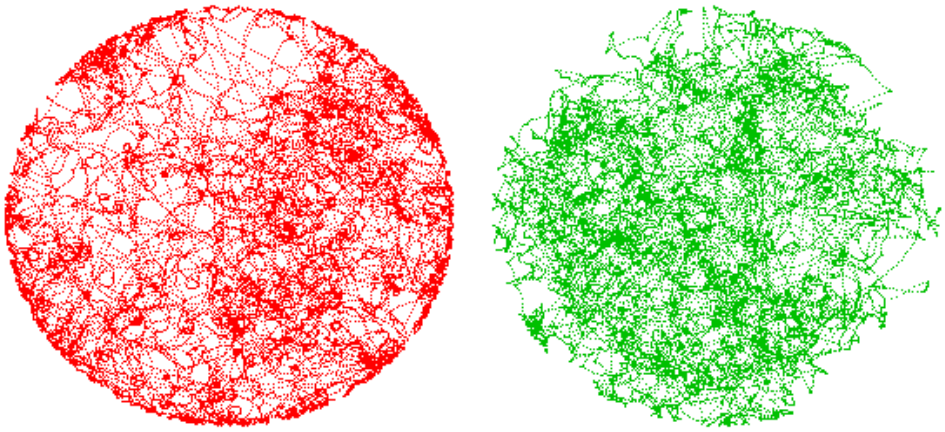

Figure 4: Trajectory of the robot using pure homeokinetic control (left) and the extended controller (right). An increasing concentration of the robots positions in the inner obstacle free part of the circular arena can be identified when using the Hebbian control layer, as compared to pure homeokinetic control.

realistically reproduce the physical interactions of the robot with the external world. In order to obtain information about long-term stability a third experiment was conducted that lasted 24 hours.

The trajectory of the robot in the two experiments is plotted in Fig. 4. The positions of the robot concentrate increasingly to the inner obstacle-free region when using the Hebbian control layer as compared to pure homeokinetic control. The histogram of the robots distance from the center of the arena illustrates the effect of the learning scheme, see Fig. 5. During the first part of the experiment (top row) the Hebbian layer started to adapt but shows hardly any effect on the robots behavior yet. Hence the histograms show similar distributions.

The bottom row of Fig. 5 shows histograms of the robots position during a later part of the experiment where the influence of the Hebbian control layer is dominant. Without access to the Hebbian layer the probability of the robot to stay near the wall is approximately three times higher than being at any other distance from the center, cf. Fig. 5 (bottom left). This is caused by the fact that in the central obstacle free region of the arena behaviors are more stable due to the small state estimation error and hence larger distances are covered by the robot. Whereas in the region near the wall behaviors change more often due to a larger modeling error and the robot is not able to cover large distances. Therefore the robots probability to stay near the wall is higher. When enabling the Hebbian layer the robots probability of being near the wall is drastically reduced and the highest probability is now shifted towards the center of the arena, see Fig. 5 (bottom right).

The predicted modeling error ζ of the Hebbian layer leads to a change of the actual robot behavior before the collision region is reached. Since the selection of the following behavior is not constrained the robot can still reach the collision area, but with much less probability. This can be interpreted as a flexibility of the system which continues to explore the collision area.

The usage of the predicted modeling error in the homeokinetic layer leads to pre-collision

Electronic version of an article published as Advances in Complex Systems,
Vol. 12, No. 3 (2009) 273-291, DOI: 10.1142/S0219525909002258

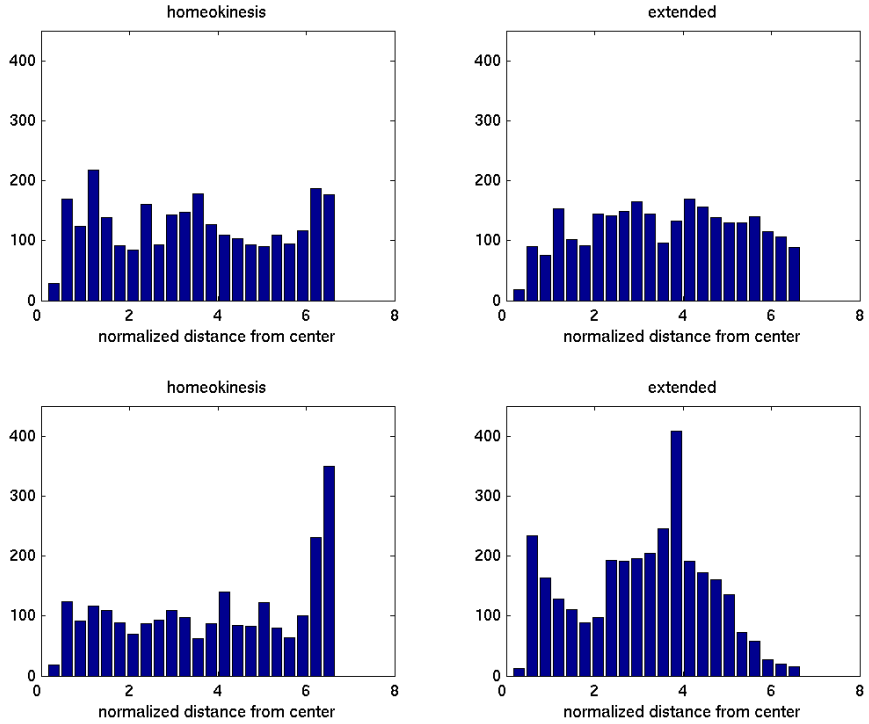

Figure 5: Histogram of the robots distance from center normalized by the respective areas. The left column presents the results for pure homeokinetic control and the right column those of the extended controller, in both columns for the first 15 minutes (top row) and the last 15 minutes (bottom row) of the experiments with a total time of 1 hour. In the initial phase the Hebbian layer is not yet functional and both controller show comparable results. In the later part of the experiment (bottom row) the mean occupancy has shifted away from the wall towards the center of the arena in the case of the extended controller.

changes of the robots behavior rather than to the trivial solution where the robot stops somewhere in the central region of the arena. In Fig. 6 the traveled distance of the robot with and without usage of the Hebbian layer is shown. Regions of inactivity are essentially absent. Also the total traveled distance is not reduced by incorporating the Hebbian layer. In the 24-hour experiments no stability problems of the system were observed, as indicated by the mean and standard deviation of the controller and internal model parameters of the homeokinetic layer in Table 1.

The weights of the Hebbian layer during the 24-h experiments show that the physical properties of the robot are reflected in the learned correlations, cf. Figs. 7. The two front infrared sensors happen to become included with a negative sign. This can be expected when considering the case of a frontal collision: The wheel counters indicate forward motion by $x_t > 0$. Then the velocity predicted by the state estimator model for the next time step will typically also be positive $\psi(x_t) > 0$. After the collision, when the front infrared sensors are still active with $x^H > 0$, the velocity sensor will yield $x_{t+1} = 0$, while the prediction is still $\psi(x_t) > 0$. Hence, the prediction error $\xi = x_{t+1} - \psi(x_t)$ will be negative. The Hebbian layer extracts this correlation between infrared sensor and modeling error by converging to negative weights for the front infrared sensors according to (18). This way the predicted

Electronic version of an article published as Advances in Complex Systems,
Vol. 12, No. 3 (2009) 273-291, DOI: 10.1142/S0219525909002258

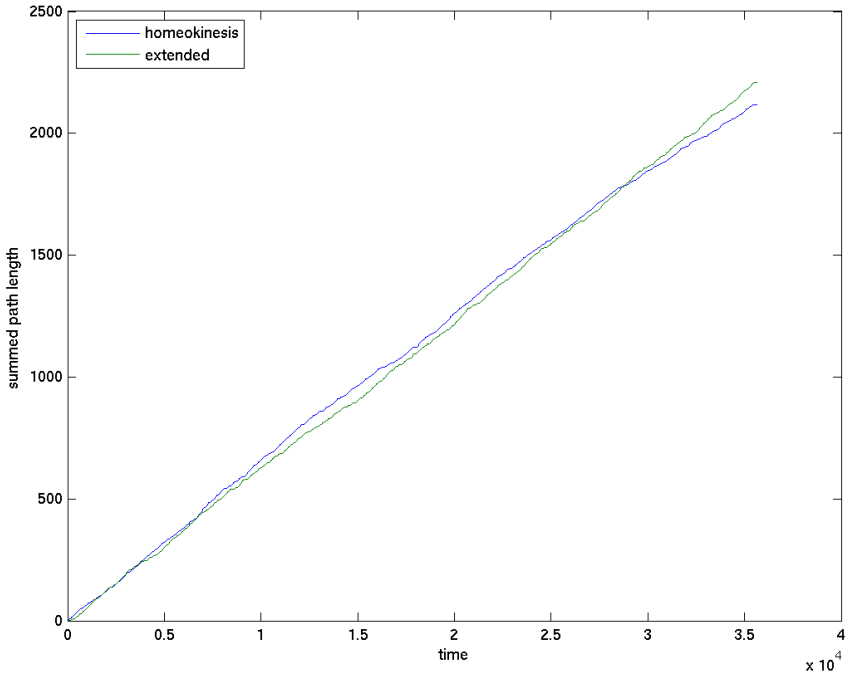

Figure 6: Cumulative distance traveled by the robot over time using pure homeokinetic control and the extended controller. The traveled distances in the two experiments are comparable, indicating that the Hebbian layer did not reduce the activity of the robot.

future modeling error ζ will be negative. The same holds true for the rear infrared sensors with inverted sign for the velocity, Hebbian weights and (predicted) modeling error. Hence, by reflecting the physical properties of the robot, the Hebbian layer provides distinct information how to react in collision situations, e.g. drive forward/backward depending on the sign of ζ. For the sidewards sensors the correlations turned out not to be significant, since they were activated during frontal as well as rear collisions. In the presented principle the directional information of the Hebbian layer is not exploited. In the context-dependent energy function E (16), which is part of the modified learning rate μ in the update rule of the bias (15), only the absolute value of the predicted modeling error ζ is relevant if ξ is assumed to be small in pre-collision situations. To be able to exploit the directional information of the Hebbian Layer we will propose a modified principle studied in a more complex hardware set-up in the following section.

We might have as well used inverted infrared sensors ($x^H \approx 0$ near the wall, $x^H \approx 1$ in free space) as context sensors. In this case the robot would rather show a tendency to stay within the vicinity of the wall. This behavior could be also interesting because the robot still retains the flexibility to adjust its internal parameters such that it is able to move freely while staying near walls. The robot's preference for wall in this modified scheme is reminiscent to a foraging rat, cf. e.g. [21].

Electronic version of an article published as Advances in Complex Systems,
Vol. 12, No. 3 (2009) 273-291, DOI: 10.1142/S0219525909002258

(a)

parameter	mean	std. deviation
$c_{0,0}$	1.2130	0.0921
$c_{0,1}$	0.0132	0.1478
$c_{1,0}$	0.0024	0.1615
$c_{1,1}$	1.2301	0.1097
h_0	-0.0090	0.2073
h_1	0.0027	0.2247

(b)

parameter	mean	std. deviation
$a_{0,0}$	0.9728	0.0311
$a_{0,1}$	-0.0044	0.0191
$a_{1,0}$	-0.0040	0.0216
$a_{1,1}$	0.9636	0.0377

Table 1: Mean value and standard deviation of (a) the controller and (b) the model parameters of a 24-hour experiment. The model parameters a converge to a unit matrix that reflects the physical properties of the robot, where each wheel is controlled by one of the motor commands. The controller parameters c reflect this structure. The bias terms h driven by the context-dependent energy function continue to oscillate about zero as indicated by the large variances.

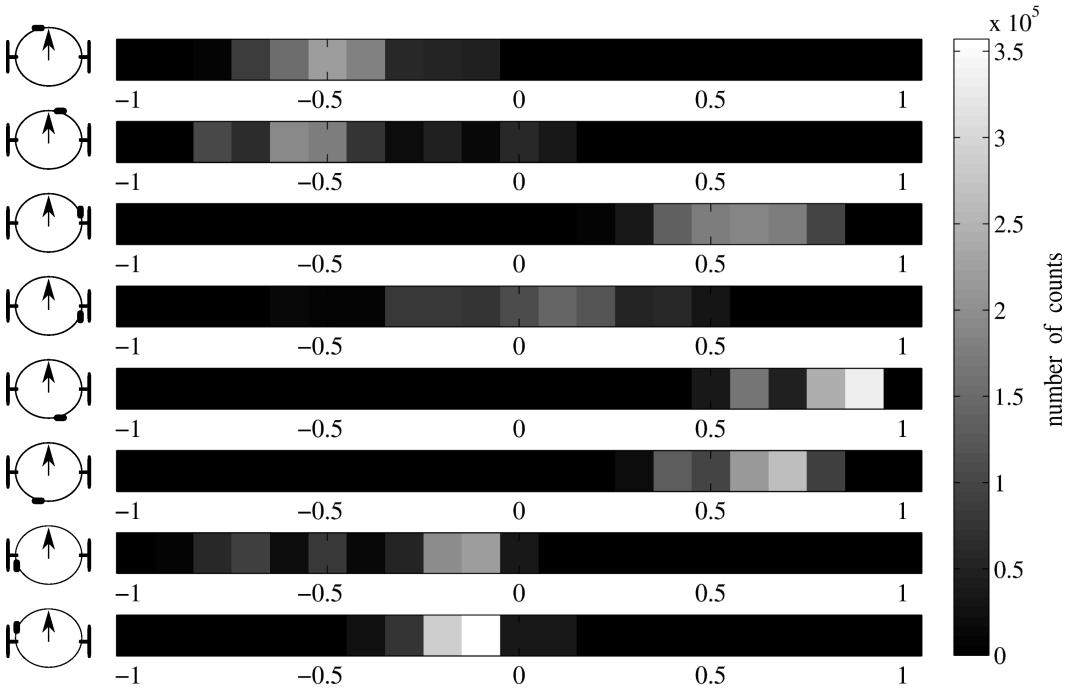

Figure 7: Histogram of the weights of the Hebbian layer contributing to ζ_1 for a long-term experiment (24 h real time) of the simulated two-wheeled robot with extended controller. The labels at the y-axis correspond to the eight infrared sensors. Front and rear sensor weights have negative and positive sign, resp., indicating the ability of the Hebbian layer to correctly extract the correlations between modeling error ξ and the activity of the proximity sensors. For details see text.

6 Gripping in a human-hand model

For the further evaluation of the context-based exploration we programmed a model of a human hand with five degrees of freedom, see Fig. 9. All joints are controlled by bidirectional motors that mimic the interplay between flexor and extensor muscles. The effect of a motor

Electronic version of an article published as Advances in Complex Systems,
Vol. 12, No. 3 (2009) 273-291, DOI: 10.1142/S0219525909002258

action is measured by motion sensors, which serve as input to the low-level homeokinetic controller. Each finger is controlled by an individual controller such that interactions between the fingers are possible only via the environment. If no object is present for manipulation the fingers become quickly engaged in vivid movements which can be interpreted as an exploration of the dynamical range. In the presence of an object the state estimation errors increase considerable when the fingers touch the object, because the information about the object is not available to the model of the dynamics of the proprioceptive sensors. It is, however, available to the Hebbian layer as context information x^H via touch sensors (realized here as infrared sensors). The Hebbian layer is implemented using (17, 18).

In this experiment we exploit the directional information of the Hebbian weights, as found and discussed in the previous section (see also Fig. 7), by directly adding the output ξ of the higher layer to the update of the threshold h so that (15) changes to

$$\Delta h = -2\mu\,(z - h) + \zeta. \tag{19}$$

This way the Hebbian layer is enabled to directly determine the direction of the actuators. The energy function is used without the context-dependent term, as given in (11). This will give the same result as applying the scheme of the previous section. The fingers will flinch when arriving close to the surface of the object but remain active otherwise like in the free case. By changing the sign of the contribution of the higher layer to the bias update in (19) we get

$$\Delta h = -2\mu\,(z - h) - \zeta.$$

Thus, the system will show the opposite reaction to the predicted modeling error ζ. This way we can shape the behavior of the system in order to show a gripping reflex. The adaptation of a Hebbian weight in dependence of the corresponding infrared sensor and state estimation error is shown in Fig. 8. Results presented in Figs. 9 and 10 show that soon, after an object

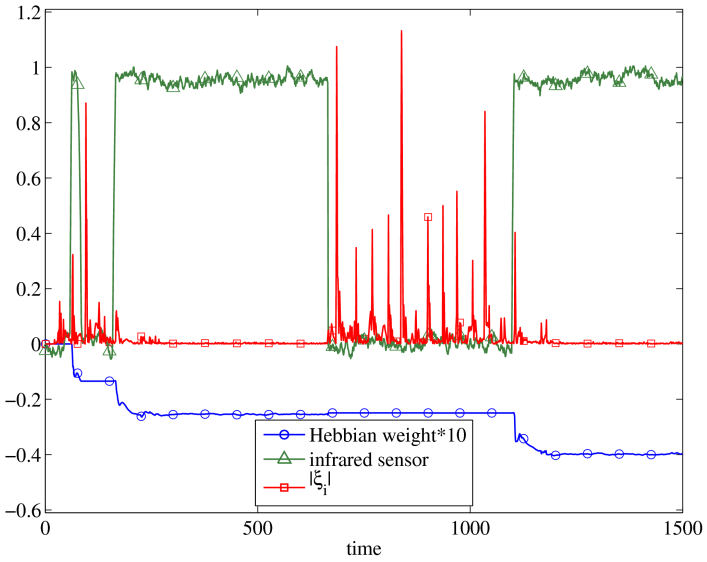

Figure 8: Adaptation of a synaptic strength of the Hebbian layer during the experiment. According to Eq. 18 the change of a weight is defined by the corresponding state estimation error ξ_i and context sensor x^H_j (here the infrared sensor). If both values are large then a change of the weight is triggered.

Electronic version of an article published as Advances in Complex Systems,
Vol. 12, No. 3 (2009) 273-291, DOI: 10.1142/S0219525909002258

is presented, a grip at the object is realized due to the domination of the Hebbian layer. If the object is removed and presented again the hand closes and the fingers grab the object.

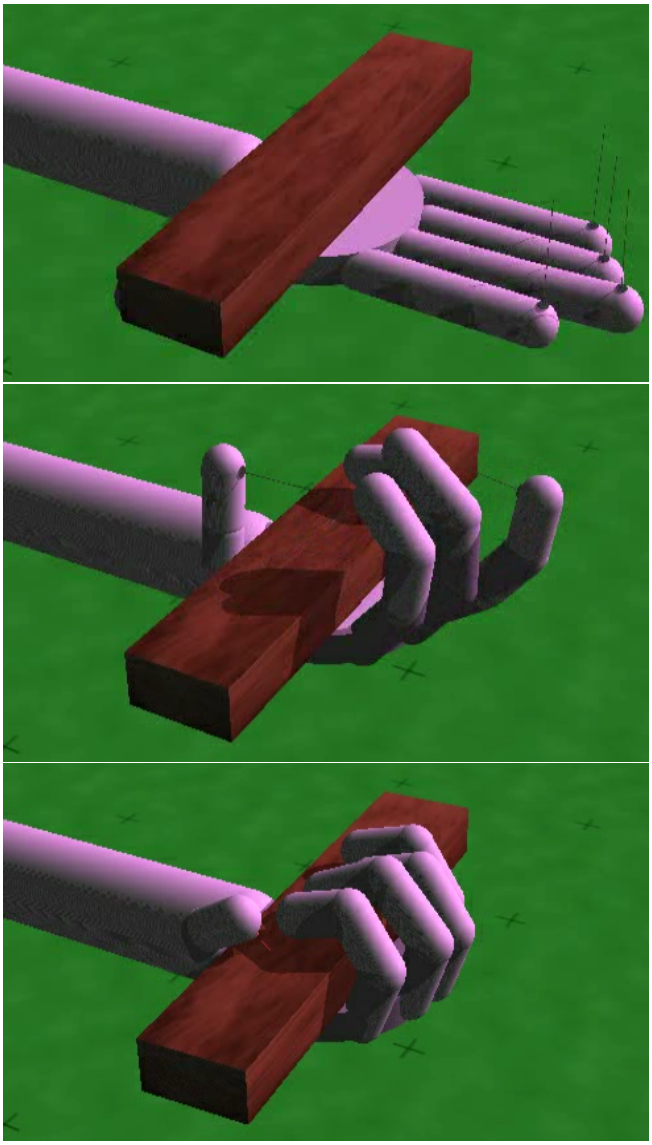

Figure 9: Simulation of a human hand with multiple degrees of freedom. The hand is equipped with motion sensors at all joins and infrared sensors at the finger tips. It is operated in a fully exploratory mode with or without a manipulated object.

7 Conclusion

In the experiments realistically simulated robots were shown to acquire low-level behaviors which are characterized by simultaneous sensitivity and controllability. The basic behaviors are obtained from an interplay of a mildly destabilizing controller with the environment which is constrained by the prediction quality achieved by an internal model. In unforeseen situations, i.e. near 'obstacles', parameter changes are triggered which are time-consuming

Electronic version of an article published as Advances in Complex Systems,
Vol. 12, No. 3 (2009) 273-291, DOI: 10.1142/S0219525909002258

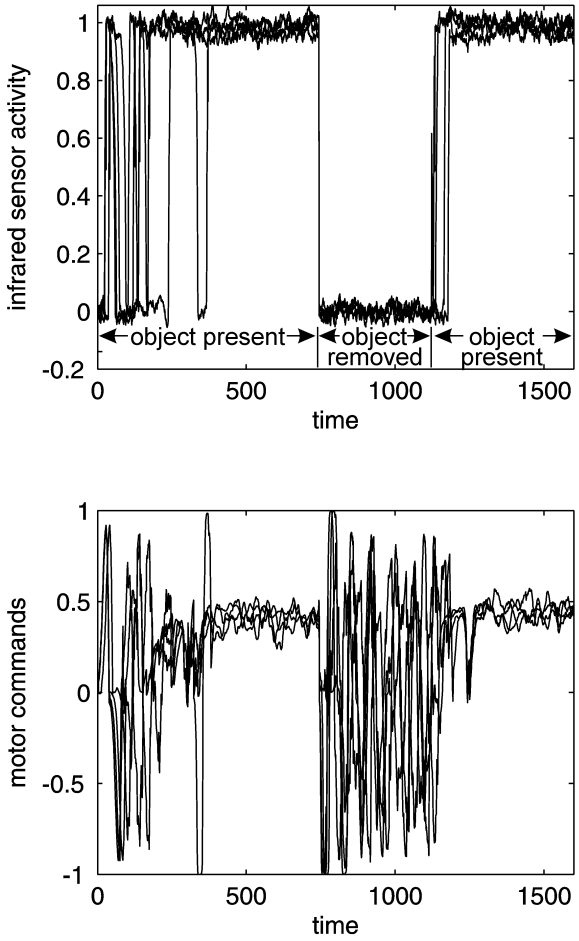

Figure 10: The finger movements that are initiated by the self-organizing controller soon converge to a grip at the object (high infrared sensor activity) with only small deviation of single fingers from the surface. When the object is removed the exploratory movements restart. If the object is present the fingers will grip it again since the Hebbian layer already learned this reflex.

and may even cause unlearning of previously acquired behaviors. The proposed second-order learning schemes are coping with such a situation in different ways [2]: Either the robot is controlled such as to avoid these situation which generates an interpretation of additional sensory inputs in terms of the low-level affordances, or it is guided towards these situations in order to further improve its prediction quality. The decision which mode of operation of the second-order learning is to be activated is to be taken in dependence of the quality of the internal model such that increasing prediction quality should favor the exploratory mode while insurmountable errors should lead to a preference of the avoidance behavior. The exploratory character of the low-level self-organizing controller is retained in both cases and the robot still explores occasionally risky regions and is hence able to adapt to slow changes in the environment. The work shows also parallels to the early motor development in biology, cf. e.g. [17], and provides a scheme for the formation of reflexes based on an approach to the self-organization of autonomous behavior.

Electronic version of an article published as Advances in Complex Systems,
Vol. 12, No. 3 (2009) 273-291, DOI: 10.1142/S0219525909002258

Acknowledgment

This work was supported by the BMBF in the framework of the Bernstein Centers for Computational Neuroscience, grant number 01GQ0432. Discussions with C. Kolodziejski and G. Martius are gratefully acknowledged.

References

[1] Ashby, W. R., *Design for a Brain* (Chapman and Hill, London, 1954).

[2] Berthouze, L. and Lungarella, M., Motor skill acquisition under environmental perturbations: on the necessity of alternate freezing and freeing, *Adaptive Behavior* **12** (2004) 47–631.

[3] Cannon, W. B., *The Wisdom of the Body* (Norton, New York, 1939).

[4] Cooper, S. J., Donald O. Hebb's synapse and learning rule: a history and commentary, *Neuroscience & Biobehavioral Reviews* **28** (2005) 851–874.

[5] Der, R., Herrmann, M., and Liebscher, R., Homeokinetic approach to autonomous learning in mobile robots, in *Robotik 2002*, eds. Dillman, R., Schraft, R. D., and Wörn, H., number 1679 in VDI-Berichte (VDI, 2002), pp. 301–306.

[6] Der, R., Hesse, F., and Liebscher, R., Self-organized exploration and automatic sensor integration from the homeokinetic principle., in *Proc. SOAVE 04* (Ilmenau, 2004), pp. 220 – 230.

[7] Der, R., Hesse, F., and Martius, G., Learning to feel the physics of a body, in *Proc. Intl. Conf. Computational Intelligence for Modelling, Control and Automation and Intl. Conf. Intelligent Agents, Web Technologies and Internet Commerce Vol-2 (CIMCA-IAWTIC'06)* (IEEE Computer Society, Washington, DC, USA, 2005), pp. 252–257.

[8] Der, R., Hesse, F., and Martius, G., Rocking stamper and jumping snake from a dynamical system approach to artificial life, *Adaptive Behavior* **14** (2006) 105–115.

[9] Der, R., Hesse, F., and Martius, G., Videos of self-organised robot behavior, `http://robot.informatik.uni-leipzig.de/Videos` (2008).

[10] Der, R., Martius, G., and Hesse, F., Let it roll – emerging sensorimotor coordination in a spherical robot, in *Artificial Life X : Proc. Tenth Intl. Conf. Simulation and Synthesis of Living Systems*, eds. Rocha, L. M., Yaeger, L. S., Bedau, M. A., Floreano, D., Goldstone, R. L., and Vespignani, A. (MIT Press, 2006), pp. 192–198.

[11] Der, R., Steinmetz, U., and Pasemann, F., Homeokinesis - a new principle to back up evolution with learning, in *Computational Intelligence for Modelling, Control, and Automation*, *Concurrent Systems Engineering Series*, Vol. 55 (IOS Press, Amsterdam, 1999), pp. 43–47.

[12] Di Paolo, E., Organismically-inspired robotics: Homeostatic adaptation and natural teleology beyond the closed sensorimotor loop, in *Dynamical Systems Approach to Embodiment and Sociality*, eds. Murase, K. and Asakura, T. (2003), pp. 19 – 42.

[13] Herrmann, J. M., Dynamical systems for predictive control of autonomous robots. Theory in Biosciences 120:3-4 (2001), 241-253.

[14] Hebb, D., *The organization of behavior: a neuropsychological theory* (Wiley, New York, 1949).

[15] Hesse, F., *Self-Organizing control for autonomous robots – a dynamical systems approach based on the principle of Homeokinesis*, Dissertation, (Georg-August-Universität, Göttingen, 2009).

[16] Kaelbling, L. P., Littman, M. L., and Moore, A. W., Reinforcement learning: A survey, *Journal of Artificial Intelligence Research* **4** (1996) 237–285.

[17] Kuniyoshi, Y. and Sangawa, S., Early motor development from partially ordered neural-body dynamics: experiments with a cortico-spinal-musculo-skeletal model, *Biological Cybernetics* **95** (2006) 589–605.

[18] Schöner, G. and dose, M., A dynamical systems approach to task-level system integration used to plan and control autonomous vehicle motion., *Robotics and autonomous systems* **10** (1992) 253–267.

[19] Steels, L., A case study in the behavior-oriented design of autonomous agents, in *SAB94: Proceedings of the third international conference on Simulation of adaptive behavior : from animals to animats 3* (MIT Press, Cambridge, MA, USA, 1994), ISBN 0-262-53122-4, pp. 445–452.

[20] Sutton, R. S. and Barto, A. G., *Reinforcement Learning: An Introduction* (MIT Press/Bradford Books, 1998).

[21] Tamosiunaite, M., Ainge, J., Kulvicius, T., Porr, B., Dudchenko, P., and Wörgötter, F., Path-finding in real and simulated rats: assessing the influence of path characteristics on navigation learning, *J. Comp. Nsci.* **25** (2008) 562–582.

[22] Tani, J. and Ito, M., Self-organization of behavioral primitives as multiple attractor dynamics: A robot experiment, *IEEE Transactions of on Systems, Man, and Cybernetics Part A: Systems and Humans* **33** (2003) 481–488.

[23] Williams, H., Homeostatic plasticity in recurrent neural networks, in *From Animals to Animats: Proceedings of the 8th Intl. Conf. On Simulation of Adaptive Behavior*, eds. Schaal, S. and Ispeert, A., *8*, Vol. 8 (MIT Press, Cambridge MA, 2004), pp. 344–353.

[24] Elizabeth Pennisi, In nature, animals that stop and start win the race, *Science* **288** (2000) 83–85.

[25] Rodrigo A. Vásqueza, Luis A. Ebenspergerb and Francisco Bozinovic, The influence of habitat on travel speed, intermittent locomotion, and vigilance in a diurnal rodent, *Behavioral Ecology* **13** (2002) 182–187.

[26] Terrie M. Williams, R. W. Davis, L. A. Fuiman, J. Francis, B. J. Le Boeuf , M. Horning, J. Calambokidis, and D. A. Croll, Sink or swim: strategies for cost-efficient diving by marine mammals, *Science* **288** (2000) 133–136.