Structure from Behavior in Autonomous Agents

Georg Martius, Katja Fiedler and J. Michael Herrmann

Abstract— **We describe a learning algorithm that generates behaviors by self-organization of sensorimotor loops in an autonomous robot. The behavior of the robot is analyzed by a multi-expert architecture, where a number of controllers compete for the data from the physical robot. Each expert stabilizes the representation of the acquired sensorimotor mapping in dependence of the achieved prediction error and forms eventually a behavioral primitive. The experts provide a discrete representation of the behavioral manifold of the robot and are suited to form building blocks for complex behaviors.**

I. Introduction

Autonomous robots as well as animals process sensory information for the purpose of generating appropriate behavior in their respective environment. This includes the selection of perceptual features, the storage of behavioral episodes, and the recall of earlier experiences, which are all essentially governed by the requirements of present behaviors and the planning of future tasks. A robot that serves as a model of animal behavior [13] should attain the organization of its sensorimotor control without specific programming nor an explicite task description, but rather achieve a vivid informational and physical interaction with its environment from a small set of sufficiently general instructions such as being encoded in the genes of an animal.

A suggestive approach is homeostatic control [6], [14], where behavioral affordances [7] are balanced with internal drives, i.e. an opportunity for behavior must be matched by a specific internal urge in order to be realized. At the same time opportunities are to be searched for when the drive requests, which might be difficult if the drives are less specific. Then an exploratory activity of the agent is required for searching the behavioral space, which we propose to be brought about by a homeokinetic principle [2], [3], [9], [8]. While simple behaviors have been shown to arise naturally from this learning scheme, it remains a major problem how an increase in the complexity of the obtainable behaviors can be achieved in such a system. The present contribution presents building blocks for a second-order learning algorithm which relies on the occurrence and persistence of self-organized low-level behaviors.

Simple behaviors can be obtained by a fixed mapping from sensory stimulus to motor actions in a closed-loop sense. The resulting relation of earlier to later inputs depends at least partially on the control of the agent. At short time intervals the complexity of this relation can be expected to be moderate which is additionally endorsed by the agent if its controller prefers predictable behaviors. Below we will describe a marginally stable control scheme that maximizes the sensitivity of the agent with respect to external inputs while maintaining their predictability. Through the latter requirement the exhibited behaviors become dependent on both the robot's body and its environment. Particularly affording behaviors are persistent for a longer time and are repeatedly performed in a similar way. The agent can identify these behaviors as behavioral primitives and make them available to higher-order control mechanisms that can select, initiate, modulate or combine them.

The following section describes the emergence of motoric primitives in a dynamical systems framework. Sect. III outlines the function of the multi-expert system to generate behavioral primitives for the example of spherical robots. More details on this and a second application are given in Sect. IV. Finally we discuss a extension of the present scheme.

II. Formation of elementary behaviors

Consider an autonomous agent with a controller K that generates motoric outputs $y = K(x, c)$ as a function of sensory inputs $x = \{x_1, \ldots, x_n\}$ in dependence of a set of parameters $c = \{c_0, c_1, \ldots, c_n\}$. A simple form of the controller is given by

$$K(x, c) = g\left(\sum_{i=1}^{n} c_i x_i + c_0\right), \tag{1}$$

where $g(\cdot)$ is a smooth sigmoidal function. In addition there is a predictive model $\tilde{x} = M(y, a)$ which estimates future sensory inputs from motoric outputs in dependence of a set of parameters $a = \{a_0, a_1, \ldots, a_m\}$,

$$M(y, a) = \left(\sum_{i=1}^{m} a_i y_i + a_0\right). \tag{2}$$

We define the sensory motor function with discrete time t as

$$\tilde{x}_{t+1} = \psi(x_t) = M(K(x_t, c), a). \tag{3}$$

The project received support from the Bernstein Center for Computational Neuroscience Göttingen grant number 01GQ0432.

G. Martius and K. Fiedler are with the Bernstein Center for Computational Neuroscience Göttingen, Max-Planck-Institute for Dynamics and Self-Organization Göttingen, and the Institute for Nonlinear Dynamics, Göttingen University, Bunsenstr. 10, 37073 Göttingen, Germany georg@nld.ds.mpg.de, katja@nld.ds.mpg.de

J. M. Herrmann is with the Institute for Perception, Action and Behaviour, School of Informatics, University of Edinburgh, Edinburgh EH9 3JZ, U.K. He is a PI at the Bernstein Center for Computational Neuroscience Göttingen michael.herrmann@ed.ac.uk

The parameter vectors a and c are adapted by gradient descent with respect to an objective function E. If only deviations from a prescribed trajectory are considered, it is suggestive to define E based on the distance between the correspondingly prescribed sensor values $\hat{x}_t$ and actual sensor values x_t. If both the controller and the predictor are adapted such that the expected prediction error

$$E_0 = \|x_{t+1} - \hat{x}_{t+1}\|^2 = \|\xi_{t+1}\|^2 \tag{4}$$

is minimized, then stable but typically trivial behaviors are achieved, e.g. the robot tends to stabilize any state where $x_{t+1} = \tilde{x}_{t+1} =$ const. There are, however, examples where this principle is indeed successful, cf. [6], [14], which is usually the case if the drive for activity comes from another source.

A different objective function derived from homeokinesis [2] can be defined based on a virtual sensor value given by

$$\hat{x}_t = \arg\min_x \|x_{t+1} - \psi(x)\| \,. \tag{5}$$

This means that if $\hat{x}_t$ was observed then an optimal prediction would have been possible. The prediction of x_{t+1} based on the actual sensor value x_t cannot be better than the one based on $\hat{x}_t$. Therefore, in order to obtain better predictions in future, it makes sense to minimize instead of (4)

$$E = \|x_t - \hat{x}_t\|^2 = \|v_t\|^2 \tag{6}$$

by gradient descent for the controller parameters c. The difference $v_t = x_t - \hat{x}_t$ may be due to either the inaccuracy of the model or a perceptual error. For the agent this does not make any difference because perceptual errors are detectable only with respect to the predictive model. Interestingly, the state where internal and perceptual errors compensate turns out not to be stable. The model parameters a are simultaneously adapted based on E_0 (4).

Mathematically, Eq. (5) is a regularized solution of the possibly ill-posed equation $\hat{x}_t = \psi^{-1}(x_{t+1})$, i.e. in principle $\hat{x}_t$ is produced by the inverse of the loop function (3), cf. Fig. 1. The controller thus performs a stabilization in inverted time, i.e. destabilization in forward time. At the same time the agent does not depart much from the predicted trajectory due to (6).

Writing the dynamics of the sensor values (3) and (5) as

$$\psi(x_t) + \xi_{t+1} = \psi(x_t + v_t) \tag{7}$$

and performing a Taylor expansion to first order we obtain

$$\psi(x_t + v_t) = \psi(x_t) + L(x_t)\, v_t, \tag{8}$$

where $L_{ij} = \frac{\partial}{\partial x_j}\psi_i(x)$ is the Jacobian. v_t and ξ_{t+1} are related by

$$v_t = L^{-1}(x_t)\, \xi_{t+1}. \tag{9}$$

Furthermore, assuming the internal model ψ to change only slowly we see that ξ has a fixed distribution. Therefore, minimizing $\|v\|$ in (6) is equivalent to minimizing $\|L^{-1}\|$. In this way mainly the small eigenvalues of L are increased. The large eigenvalues of L instead experience a relative decay because of the increasing effects of non-linearities at excited eigenvalues. Ideally all eigenvalues of L approach unity, such that the system becomes critical. Practically, however, the stochasticity of the system as well as counterbalancing non-linearities cause the eigenvalues to become slightly larger than one. The resulting dynamics can be described as “bifurcation-inducing", i.e. when starting from a resting state movements in the either direction of the dominant eigenvector of L are initiated or periodic behaviors develop into a quasiperiodic state. Similar mechanisms have shown to be useful in robot control already for a pre-programmed parameter dynamics, cf. [1].

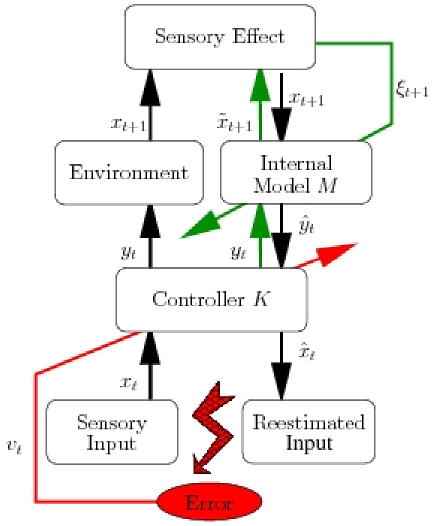

Fig. 1. The homeokinetic control algorithm. Parameters of M are adapted to minimize the difference between sensor values x_{t+1} and predictions $\tilde{x}_{t+1}$. The parameters of K are adapted to minimize the misfit between true sensory inputs x_t and their re-estimates $\hat{x}_t$.

The state and parameter dynamics is now described by

$$x_{t+1} = \psi(x_t) + \xi_{t+1}\,, \tag{10}$$

$$c_{t+1} = c_t - \varepsilon_c \frac{\partial}{\partial c} E\,, \tag{11}$$

where $\varepsilon_c = 0.1$ is a learning rate. The rule (11) produces an itinerant trajectory in c-space, i.e. the agent runs through a sequence of behaviors that are determined by the interaction with the environment, but are not permanently learned by the agent.

Further details of the control paradigm can be found in [3] and [9]. Nevertheless, we want to outline some characteristics and capabilities of homeokinetic control by considering some examples that were realized on various robotic platforms [10]. For example, the “rocking stamper” [4] consists of an inverse pendulum mounted on a bowl-like trunk. It exhibits different rocking modes, preferably at the eigenfrequencies of the system.

A somewhat more complex example for the self-

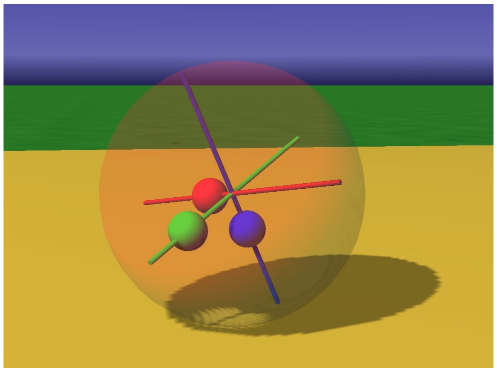

Fig. 2. A spherical robot actuated by three movable masses. It is simulated in the physically realistic simulation environment[10]. For the hardware version cf. [11].

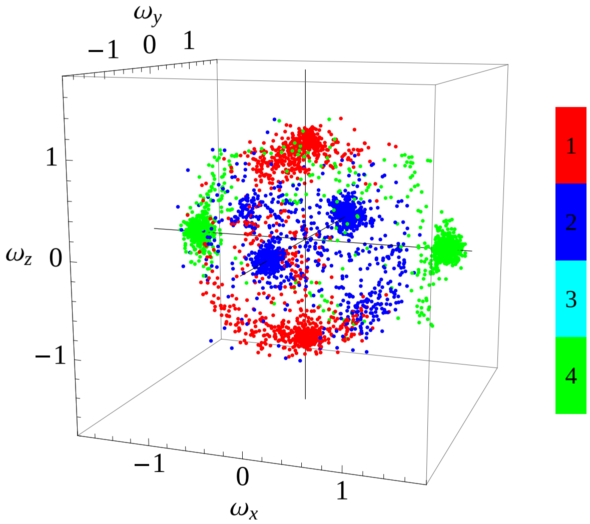

Fig. 3. Partition of the behavioral space represented by the angular velocities around the three axes of the spherical robot, cf. Fig. 2. Four experts were used during learning, but one is eventually uncommitted due to the limited number of accessible intrinsic attractors.

organization of natural behaviors is provided by a spherical robot [5] which is actuated by three internal massive weights that can be moved along orthogonal axes, see Fig. 2. After an initial phase, the system prefers to keep one mass fixed at its axis while performing a coordinated movement of the other two such that the robot rotates around the first axis. The robot thus moves forward like a wheel or sometimes turns in place. The behavior is changed every few tens of revolutions by an internal reorganization that is driven by gradient descent on the error function (6). Furthermore, high-dimensional systems such as snake- or chain-like robots, quadrupeds, and wheeled robots [4] have been successfully controlled, where it is of particular interest that the control algorithm induces a preference for movements with a high degree of coordination among the various degrees of freedom. All the robotic implementations demonstrate the emergence of play-like behavior, which strongly depend on the body and the environment.

III. Multi-expert control

If the robot encounters a similar situation in the environment it should reuse behavioral elements that have been acquired earlier. In order to guarantee a smooth transition between the elements as well as a robust and coherent activation of the appropriate behaviors it is necessary to provide the robot with a flexible representation of the behavioral elements. We suggest to use an adaptive input-output mapping for each element which then serves as an expert for the particular behavior.

$$(\hat{x}_{t+1},\hat{y}_t) = F_i(x_t,x_{t-1}),\quad i=1,\ldots,s \qquad (12)$$

The s experts are not merely open-loop controllers, but function in a closed sensorimotor loop. The function F was realized by a neural network with a controlling and a modeling layer which are connected by a small hidden layer. This bottleneck setup was chosen in order to enhance the generalization in the experts. It is possible to measure the suitability of the expert on-line in the sense of a temporal average of the quality of the predictions. This induces a competition among the experts and the expert with the maximal similarity is allowed to adapt its parameters to those of the self-organizing controller. On the other hand, if the controller visits novel regions in the parameter space, it challenges the generalization ability of the experts. For this purpose we define the prediction error as

$$\tilde{E}_t^i = \left\|(x_t,y_{t-1}) - F_i(x_{t-1},x_{t-2})\right\|^2 , \qquad (13)$$

which is modulated with a penalty for sub-optimality, i.e. experts with a prediction error far from the recently achieved optimum are deferred in favor of uncommitted experts. Each expert keeps track of the minimum of the smoothed error $\breve{E}^i$. The distance from this minimum is used as a sub-optimality term so that the winning expert w_t at time t is defined by:

$$w_t = \operatorname*{arg\,min}_{i=1,\ldots,s} \tilde{E}_t^i + p\left(\max\{0,\tilde{E}_t^i - \breve{E}_t^i\}\right)^2 , \qquad (14)$$

where $p = 5$ is the penalty factor. Only the winning expert is allowed to learn the sensorimotor mapping from the robot.

Behaviors are likely to be selected if the robot remains for a certain time interval in the vicinity of a quasi-stable behavior and if this happens repeatedly.

IV. Details of the Experiments

In the first experiment we used the spherical robot, Fig. 2 [11]. It receives as inputs the z-components of the direction in physical space of all internal axes. The controller directly affects the position of masses on each of the axes. Controlling the robot with the self-organizing controller from section II the system reached only three fixed points of the parameter dynamics (corresponding to limit cycles of the physical dynamics of the robot) such that in a system of $s = 4$ experts one remains essentially uncommitted whereas

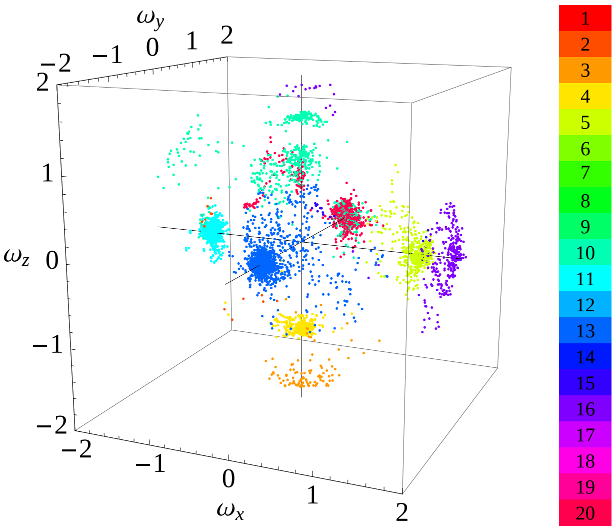

Fig. 4. Partition of the behavioral space (angular velocities around the three axes) of the spherical robot by 20 experts. In contrast to Fig. 3 now the experts have a characteristic speed which is selected in the course of learning. In addition the symmetry between forward and backward movements around the same axis is broken. For clarity only a selection of experts is displayed.

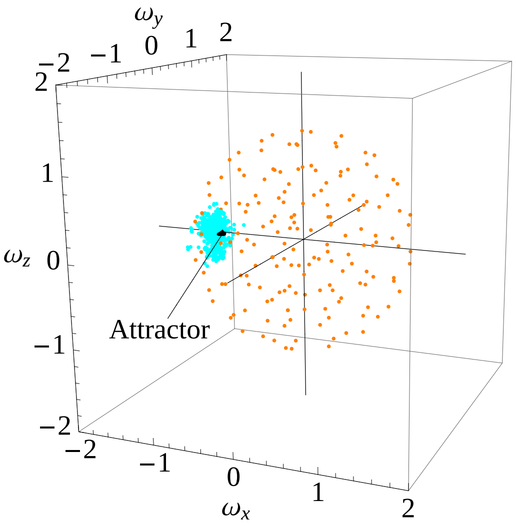

Fig. 5. The spherical robot is now controlled by one of the experts (#11) from Fig. 4. Learning is switched of here. From any initial angular speed (orange/light gray) the controller converges to the black area marked as attractor. The region where this controller was successful during learning is marked in cyan/dark gray.

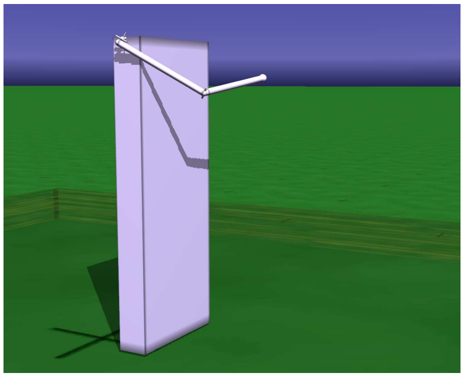

Fig. 6. Physical simulation of a human arm with four degrees of freedom.

the other three clearly partition the behavioral space, see Fig. 3. For longer runs more quasi-attractor behaviors of the robot are visited and a larger number of attractor behaviors is extracted by the experts. Figs. 3 and 4 show the space of rotation velocities about the three internal axes. Rotational velocities are used for visualization but cannot be sensed by the robot. Depending on the number of experts and the activity of the robot during learning the clustering of the behavioral space becomes more fine grained.

The frozen parameter set of an expert is not necessarily an attractor because the robot is still interacting with a potentially complex environment and the training patterns only cover a small part of the possible behavioral space. However, for the case considered here the experts determine indeed attractors as demonstrated in Fig. 5, where the robot is initialized with a set of initial conditions that cover all possible directions. For each of these initializations the robot is driven to a directed rolling mode that leaves one of the axes constant. While the set of states that the robot has visited during learning of this expert was rather fuzzy, now the states are much more focused, cf. Fig. 5. This can be considered as an extrapolation of the learned behavior. Whereas during training only states from a small cloud around the stable behavior were performed by the robot, the basin of attraction of the behavior can actually be much larger or even global as in the particular example. For the spherical robot we found that most behaviors are attractor states of the respective experts.

In addition to the spherical robot, we also considered a simulated human arm with four degrees of freedom, see Fig. 6. The arm is controlled by motor outputs that specify nominal joint angles. Due to the strong inertia effects these angles are not necessarily realized. The deviations are the essential source of information to the controller from the robot. Fig. 7 shows clustering of hand position space with 32 experts which are drawn in two coordinate frames for the sake of visibility.

V. Discussion

The controller described in Sect. II can be interpreted as a simplified neuron that interacts both with a second neuron that represents the internal model and with the environment. The output function of the controller neuron is flexible and may be brought about by the synaptic dynamics (11). Although in a biological context this reduction to control by single neuron is a gross oversimplification, it may be considered as an interesting illustration of a principle that allows a neural system to settle into a state that is characterized by an intense interaction with its environment. The immediate

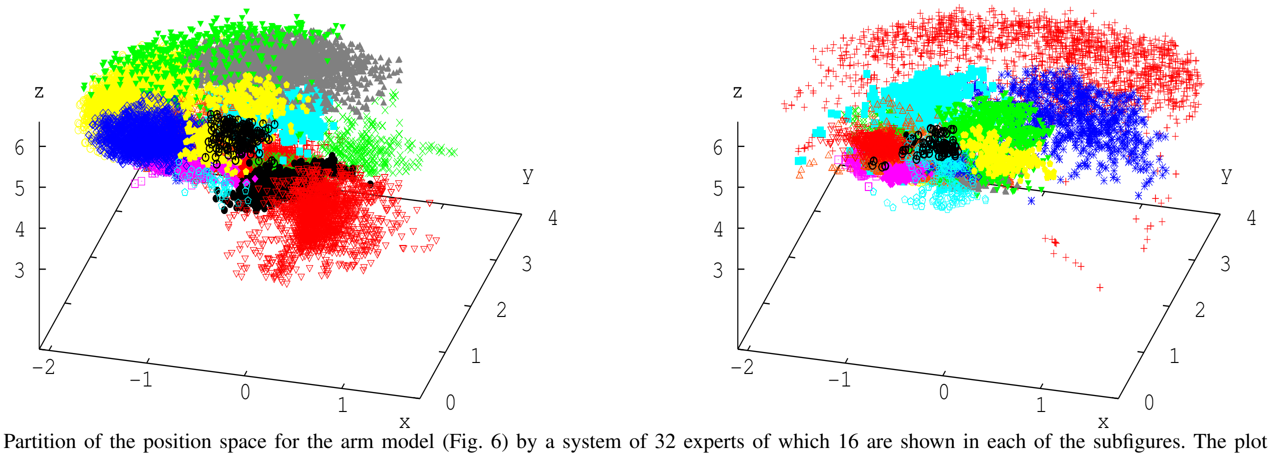

Fig. 7. Partition of the position space for the arm model (Fig. 6) by a system of 32 experts of which 16 are shown in each of the subfigures. The plot shows the coordinates of the hand position colored according to the winning expert at that point.

environment of a neuron within a larger system essentially consists of other neurons and one obtains synchronization in local networks. This extends the range spanned by the output-to-input feedback loops ultimately towards the sensory and motoric interfaces with the environment. This means only local information can be used for the self-tuning of the system to a critical state. Different groups of neurons represent different functions and are selected by activation e.g. through the dopamin system.

The experts in the present scheme can be considered as a symbolic representation of behavioral modes [12] such as rotations about an axis or arm movements in a certain region of the grasping space. Because the clustering of the behaviorally relevant space is not done with respect to the statistics of the sensory inputs, but based on behavioral success, we may assume that homogeneity of the movement is the main criterion for the formation of elementary behaviors. Furthermore the clustering it is not limited simple to geometric partitioning but can also distinguish limit cycles of different frequency as seen Fig. 4.

A second order learning based on the elementary behaviors might be implemented with reinforcement learning. As most successful applications of reinforcement learning are implemented in discrete state spaces, it is worth to consider the presented learning scheme as a precondition for reinforcement learning in continuous domains. The behavioral clustering extracts repeatable and predictable movement patterns from a large behavioral manifold. If in this way all relevant behaviors are acquired by the multi-expert system then a reward based schemes can be expected to connect and select these as implied by an external source of rewards.

References

[1] A. Bredenfeld, H. Jaeger, and T. Christaller. Mobile robots with dual dynamics. *ERCIM News*, 42, 2001.

[2] R. Der. Self-organized acquisition of situated behavior. *Theory Biosci.*, 120:179–187, 2001, cf. also http://robot.informatik.uni-leipzig.de.

[3] R. Der, M. Herrmann, and R. Liebscher. Homeokinetic approach to autonomous learning in mobile robots. In R. Dillman, R. D. Schraft, and H. Wörn, editors, *Robotik 2002*, number 1679 in Berichte, pages 301–306. VDI, 2002.

[4] R. Der, F. Hesse, and G. Martius. Rocking stamper and jumping snake from a dynamical system approach to artificial life. *Adaptive Behavior*, 14(2):105–115, 2006.

[5] R. Der, G. Martius, and F. Hesse. Let it roll – emerging sensorimotor coordination in a spherical robot. In L. M. Rocha, L. S. Yaeger, M. A. Bedau, D. Floreano, R. L. Goldstone, and A. Vespignani, editors, *Proc. ALife X*, pages 192–198. MIT Press, 2006.

[6] E. Di Paolo. Organismically-inspired robotics: Homeostatic adaptation and natural teleology beyond the closed sensorimotor loop. In K. Murase and T. Asakura, editors, *Dynamical Systems Approach to Embodiment and Sociality*, pages 19 – 42, 2003.

[7] J. J. Gibson. *Perceiving, Acting and Knowing: Toward an Ecological Psychology*, chapter The Theory of Affordances, pages 67–82. Hillsdale, 1977.

[8] J. M. Herrmann. Dynamical systems for predictive control of autonomous robots. *Theory Biosci.*, 120:241–252, 2001.

[9] M. Herrmann, M. Holicki, and R. Der. On Ashby's homeostat: A formal model of adaptive regulation. In S. Schaal, editor, *From Animals to Animats*, pages 324 – 333. MIT Press, 2004.

[10] G. Martius, F. Güttler, F. Hesse, and R. Der. Videos of self-organising robot behavior and lpzrobots simulator 0.4. http://robot.informatik.uni-leipzig.de, 2008.

[11] J. Popp. http://www.sphericalrobots.com, 2004.

[12] J. Tani. Symbols and dynamics in embodied cognition: Revisiting a robot experiment. In M. V. Butz, O. Sigaud, and P. Gerard, editors, *Anticipatory Behavior in Adaptive Learning Systems*, pages 167–178. Springer-Verlag, 2004.

[13] B. Webb. Can robots make good models of biological behaviour? *Behavioral and Brain Sciences*, 24:1033–1050, 2001.

[14] H. Williams. Homeostatic plasticity in recurrent neural networks. In S. Schaal and A. Ispeert, editors, *From Animals to Animats: Proc. 8th Intl. Conf. on Simulation of Adaptive Behavior*, volume 8. MIT Press, 2004.