Emergence of Interaction Among Adaptive Agents

Georg Martius[1,2,3], Stefano Nolfi[4], and J. Michael Herrmann[1,2,5]

[1] Bernstein Center for Computational Neuroscience Göttingen
[2] Institute for Nonlinear Dynamics, University of Göttingen
[3] Max Planck Institute for Dynamics and Self-Organization
[1−3]Bunsenstraße 10, 37073 Göttingen, Germany
[4] Institute of Cognitive Sciences and Technologies, LARRAL
Via S. Martino della Battaglia 44, 00185, Roma, Italy
[5] School of Informatics, Institute for Perception, Action and Behavior
Edinburgh University, 11 Crichton Street, Edinburgh EH8, Scotland, U.K.

georg@nld.ds.mpg.de stefano.nolfi@istc.cnr.it mherrman@inf.ed.ac.uk

Abstract. Robotic agents can self-organize their interaction with the environment by an adaptive "homeokinetic" controller that simultaneously maximizes sensitivity of the behavior and predictability of sensory inputs. Based on previous work with single robots, we study the interaction of two homeokinetic agents. We show that this paradigm also produces quasi-social interactions among artificial agents. The results suggest that homeokinetic learning generates social behavior only in the the context of an actual encounter of the interaction partner while this does not happen for an identical stimulus pattern that is only replayed. This is in agreement with earlier experiments with human subjects.

1 Introduction

The concept of self-organization describes the formation of specific structures in the presence of unspecific driving forces. It is considered to be relevant for the ontogenesis of living beings, the generation of behavior in autonomous robots, and the emergence of higher functions in agents. We have shown [1, 2], that interesting behaviors can be generated in robotic agents [3] by a self-organizing controller that follows the cybernetic principle of *homeokinesis*[6] [4]. It combines the maximization of sensitivity with respect to external stimuli and the avoidance of unpredictable behaviors. Driven by this principle, an agent becomes engaged in a vivid interaction with its environment, it starts to move autonomously and escapes from blockage as well as from unpredictable i.e. quasi-random situations. In this way the agent shows a preference for states where the control actions are effective. If the environment also contains other agents then the adaptation capabilities of the agents may play a role in the emergence of communication [5].

[6] Greek: homoios (equally, likewise) and kinesis (movement), meaning the adaptive control of a kinetic quantity and thereby self-regulate internal parameters.

Whether or not the adaptivity is indeed sufficient to enable the communication will be the main question of this contribution. Although the interaction is contingent, i.e. may be negotiated to any level from no interaction to strongly correlated behavior, it turns out that the homeokinetic principle induces a bias toward intense interaction which is counterbalanced only by the self-generated instability of the agents. We will explore the characteristics of the behavior exhibited by two homeokinetic agents that interact in a shared environment, where direct internal representations of other agents are not possible. Moreover, internal states in the agents will turn out to be unnecessary. The experiments reported here are based on a variation of the perceptual crossing scenarios introduced and studied on human subject by Auvray et al. [6] in which two human subjects can move an object left or right along a single dimension and can perceive through a tactile sensor an object corresponding to the other agent or to a "shadow" image of the other agent. The interesting aspect of this scenario is that human subjects display a good ability to discriminate between sensing and non-sensing objects (corresponding to the other agent or to its shadow) which move exactly in the same way, without the need of a specific training [6]. This can be explained by considering the way in which the agents react to the perception or to the lack of perception of the other agent. The same type of ability is demonstrated by the ability of young children to spontaneously discriminate between a video showing the behavior of their mother interacting live with the child with respect to a video showing a pre-recorded interaction [7]. This discrimination skill has been reproduced in evolutionary robotics experiments where agents have been selected for the ability to discriminate between real agents and insensitive "shadows" [8] or between interacting agents and their exact behavioral replay [9]. From these experiments it has been concluded that the detection of social contingency does not require complex cognitive skills [8].

We will demonstrate how an ability to discriminate social contingency arises spontaneously in self-organizing agents which have been rewarded for paying attention to objects independently on whether objects correspond to agents or to their shadowed images. Indeed, as soon as agents develop a preference for any tactile perception, they tend to display a preference for situations which lead to a bi-directional sensation. This tendency is due to the characteristics of the homeokinetic paradigm which maximize the sensitivity with respect to external stimuli while minimizing unpredictability. These results suggest that agency detection might result indirectly from the entrainment of the behavior of the two agents which emerge spontaneously (i.e. even in absence of a direct reward) in self-organizing homeokinetic agents.

The next section introduces homeokinetic learning. In Sections 3 and 4 we describe result from our experiments which are discussed in Section 5.

2 Self-organized Control

Based on the concept of homeokinesis [4] and the concept of self-organization, we developed in previous studies a controller [4, 1, 2] that establishes interesting sensorimotor couplings in a closed loop setup. This is achieved by the simultaneous maximization of the sensitivity with respect to sensory stimuli and the maximization of the predictability of future inputs. Both predictability and sensitivity are defined with respect to an adaptive internal representation of the sensorimotor loop, cf. Fig. 1. From the difference of the sensor readings and the model estimates an energy function is obtained that is used to modify the parameters of the controller by gradient descent. The internal model is simultaneously improved. Details can be found in the appendix.

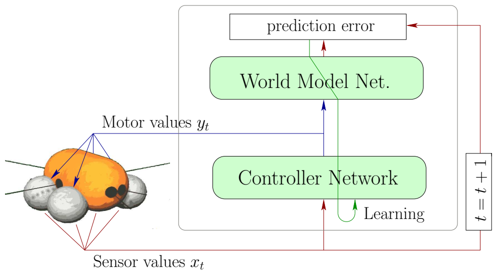


Fig. 1. Schematic view of the self-organizing controller attached to a wheeled robot.

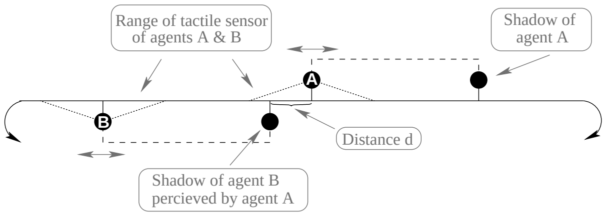
The capabilities of homeokinetic control become obvious by considering some examples that were realized on various robotic platforms [3]. The "rocking stamper" [2] consists of an inverse pendulum mounted on a bowl-like trunk. It exhibits different rocking modes, preferably at the eigenfrequencies of the system.

A more complex example for the self-organization of *natural* behaviors is provided by a spherical robot [10] which is actuated by three internal massive weights that can be moved along orthogonal axes. After an initial phase, the system prefers to keep one mass fixed at one axis while performing a coordinated movement of the other two such that the robot rotates around the first axis. Thus the robot moves forward like a wheel or sometimes turns on the spot. The behavior is changed every few tens of revolutions by an internal reorganization that occurs even in the absence of external stimuli. Furthermore, high-dimensional systems such as snake- or chain-like robots, quadrupeds, and wheeled robots [2] 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. However, it is possible to shape the development of behaviors with reinforcement [11].

3 Interaction Among Homeokinetic Agents

We use a set-up similar to Refs. [6, 12], where two agents are moving along a one-dimensional track with cyclic boundary conditions. Each agent has a copy (shadow) that moves along at a fixed distance, cf. Fig. 2 (left). The agents in our experiments have an independent parameter dynamics and are initialized randomly in order to avoid any implicit knowledge about the partner.

The controllers remain adaptive during the entire experiment, i.e. there is no training phase. The behaviors are acquired on the fly rather than being acquired during a previous evolutionary process. Each of the agents are equipped with a tactile sensor, see Fig. 2 (right), that is activated in the same way when being close to either the other agent or its copy. In contrast to Refs. [6, 12], we use a sensor with a continuous characteristic, which fits better into the paradigm of dynamical systems and does not require internal states in the agents. The sensor values depends only on the distance between the centers of the robots and objects, cf. Fig. 2. Because the robots move in a one-dimensional world, it is reasonable to assume point-like agents that are, however, assumed to obey realistic physical laws for mass, inertia, friction, and motoric forces.



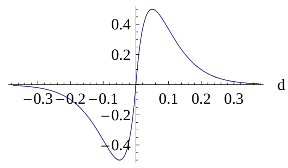


Fig. 2. (*left*) Schematic setup with two agents and the respective shadows. (*right*) Activation function of the tactile sensor with a sensor range of $R = 0.3$

The motor forces affect the agent by the following dynamical equations

$$a_{t+1} = (y_t - \mu v_t)/m \qquad a\text{: acceleration, } \mu\text{: friction, } m\text{: mass} \tag{1}$$

$$v_{t+1} = a_{t+1}\Delta t + v_t \qquad v\text{: velocity, } \Delta t\text{: time step} \tag{2}$$

$$p_{t+1} = \mathrm{Env}(v_{t+1}\Delta t + p_t) \qquad p\text{: position} \tag{3}$$

where y is the force produced by the motor and $\mathrm{Env}(\cdot)$ is a function that maps the position into the cyclic environment $[-1, 1)$. The shadow has the fixed offset position $s_t = \mathrm{Env}(p_t + 0.6)$.

The motor value y is the controller output and hence depends on the sensory inputs obtained by a velocity sensor $x_1 = v_t$ and a tactile sensor x_2. The latter responses to the nearest object (either agent or shadow) at position o as

$$x_2 = \frac{e}{2}\gamma d\, e^{-|\gamma d|} \qquad d = \mathrm{Env}(p_t - o),\ \gamma = 6/\mathrm{R} \tag{4}$$

All sensor values are subject to a weak noise. The term “tactile” is understood as in Ref. [12], practically it could be realized by a light sensor. The sensor should however be spatially localized or provide a good signal at ranges where the robot moves in a few time steps. We have tested tactile sensor characteristics different from (4) without obtaining qualitatively different results. The function that connects the sensory inputs to the motor output y is adaptive and is updated every simulation step, cf. Appendix. For all experiments the following parameters were used: $\epsilon = 0.1$ (Eqs. 10,11), $\Delta t = 0.01$, $m = 0.1$, $\mu = 1.5$, $R = 0.3$.

Above a certain distance from the other agent the agent experiences hardly any sensory errors and increases thus its sensitivity to sensory inputs, especially to the velocity sensor which leads to high speeds. This and the fact that the presence of the other agent causes large prediction errors make extended interactions with the other agent unlikely. In order to achieve reasonable search times, we applied a reinforcement scheme [11] that introduces a preference for both the other robot and its shadow in the same way. The agents receive a reward for states with an activated tactile sensor

$$r(t) = \begin{cases} 2 & |d| < R \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

The objective function E (9) of the self-organizing controller is modulated by the reward,

$$E_r = (1 - \tanh(r(t)))E, \tag{6}$$

which may be interpreted as a learning rate modulation. Small learning rates correspond to high rewards and vice versa, such that the agent develops a tendency to stay in the rewarded areas, i.e. interact with objects in the range of the tactile sensor.

Fig. 3 displays the time that agent A received tactile input from agent B, agent B's shadow, and neither of both, respectively. The values are normalized with respect to their prior frequencies, which are given by the relative area where to encounter the events (agent and shadow have size 0.6 each, which leaves 0.8 for the free space).

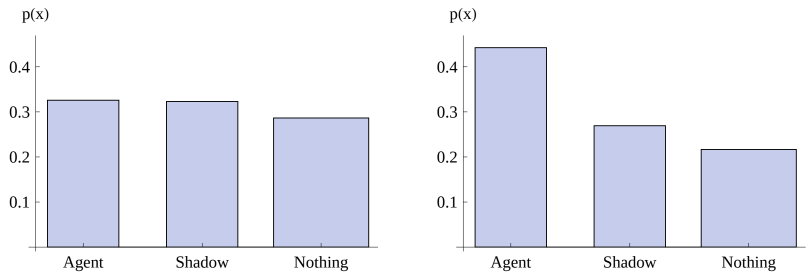


Fig. 3. Normalized interaction time of agent A with agent B, with the shadow of agent B, and with neither of the two, respectively for runs of 60 minutes simulated real time. (*left*) Self-organizing behavior according to the homeokinetic principle without rewards. (*right*) Homeokinetic learning with additional reinforcement of situations where tactile input was received.

In the case with reward the agent spends almost twice the time interacting with the other agent than with the other's shadow. This can be explained by the same reasoning that applies to the experiments with human subjects. If both agents see each other then they explore less, i.e. the changes in behavior become smaller, possibly in order to appear more predictable. However, if agent A senses the shadow of agent B then agent B does not receive any tactile input and remain explorative. In Fig. 4 the distribution of distances between both agents

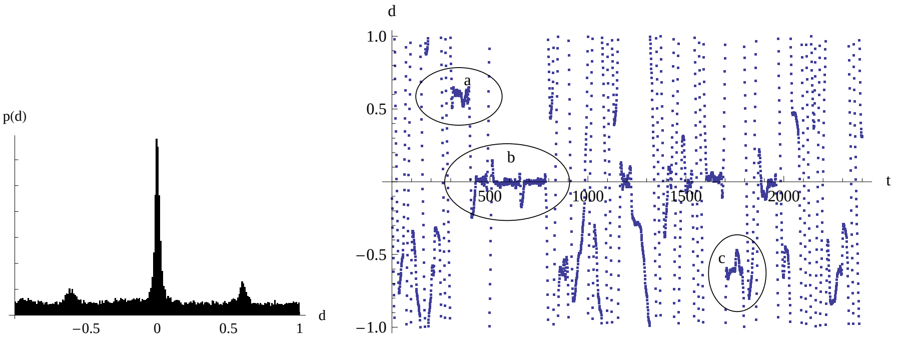


Fig. 4. Interaction of homeokinetic agents with reinforcement. (*left*) Distribution of distances between agent A and agent B. (*right*) Time course of the distance for four minutes (out of 60). (a) agent A senses shadow of agent B, (b) both agents oscillate around each other and (c) agent B senses shadow of agent A.

and the time course of the distance for a short time is displayed. The histogram of distances shows clearly the preference of a direct interaction between the two agents. Interaction with the shadow results in the the small bumps at -0.6 and 0.6.

4 Agency Detection

In order to study the detection of agency, i.e. the distinction of the live interaction with a partner and a passive replay, we have performed a variant of the above experiment where one of the agents is replaced by an exact replay of an earlier experiment. This set-up has been already studied in an evolutionary approach [9] and is actually more similar to the double TV experiment [7, 13]. In Ref. [9] the agents developed a stable interaction, which leads to a compensation of the influence of noise. In other words, the distinction of agent and replay is due to a tendency towards an adaptive dynamical interaction.

As in the previous experiment we approach the problem by an on-line learning scheme. Both the replayed and the exploring agent have again different parameter sets and behave therefore differently. The agents are moving along the same cyclic one-dimensional arena as above. Here no shadows are involved, see Fig. 5.

The agents are rewarded when sensing each other in the same way as above (5). In Fig. 6 the statistics of the behavior of the agents is plotted for both experiments. It is clearly visible that the agent spends more time interacting with the

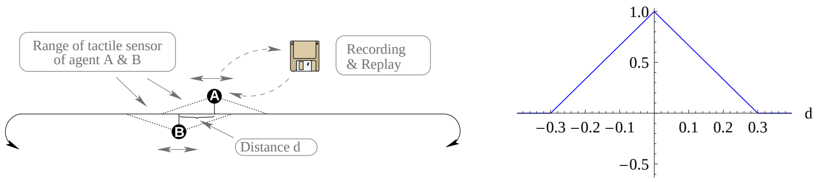


Fig. 5. (*left*) Schematic setup of the experiment with replay; (*right*) Sensor response in dependence of the distance between the agents.

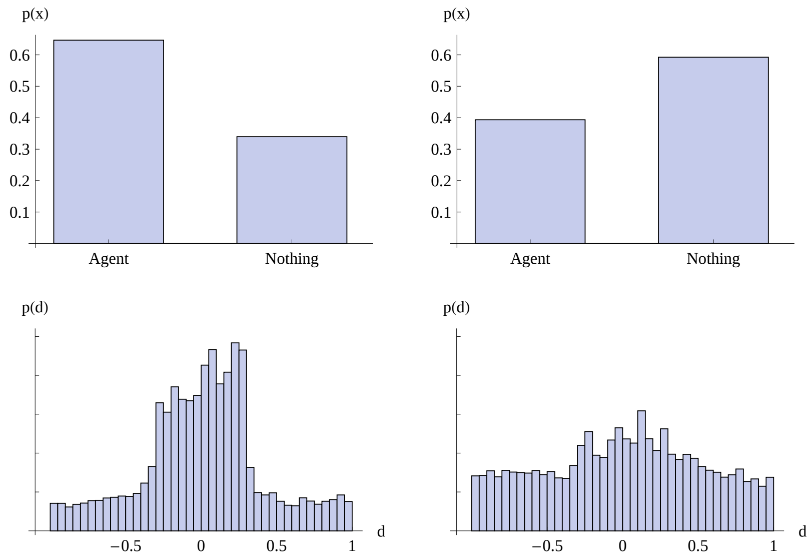


Fig. 6. Statistics of the behavior of the agents for life interaction (*left*) and for replay scenario (*right*). Top row depicts the relative time agent A spends on interacting with agent B and nothing, respectively. In the bottom row the distribution of distances between both agents is plotted.

real agent than with the non-reactive replay. This implies that it is easier to synchronize if both agents are adapting.

5 Discussion

We have studied the social interaction of two agents that are capable of self-organizing their behavior. The experiments are inspired by psychological effects found in human subjects [6, 7, 13] and are intended to investigate the mechanisms of the experience of agency. Phenomenologically the observations are similar to those presented for the case of evolving robots [12, 9], where the two agents were assumed to have an identical configuration, while we have used an on-line learning algorithm with random initial conditions. In a sense this is similar to a real life situation where the knowledge about the partners is limited. Due to the quasi-random occurrence of external inputs and the general destabilization of the parameter dynamics (Eq. 10,11) by the homeokinetic learning algorithm, the agent's parameters do not converge.

In the setup with perceptual crossings (Sect. 3) the agents showed a clear distinction of the contingently reacting partner and the non-contingent shadow. This is achieved even though the agents were rewarded for each tactile sensation regardless of whether it was caused by a real agent or its shadow. One reason for the successful distinction is the qualitatively different behavior of an agent interacting with another agent and the agent searching quickly through an environment. These results reproduce the findings in humans subjects [6] as well as in evolving robots [12].

The second experiment (Sect. 4) demonstrates that the self-organizing agents can perform an agency detection merely by self-sufficient behavioral adaptation. The interaction with the reactive agent turned out to be more probable than with the statistically identical replay. It suggests that agency detection might result indirectly from the entrainment of the behavior of the agent within the robot-environmental interaction. This behavior emerges spontaneously i.e. even in absence of a direct reward in the self-organizing homeokinetic agents. It is thus not necessary to specify the distinction as a specific goal, it follows rather from the adaptive explorativity level which requires predictable reactions from the environment, while in the replay case a similar behavior as in the quasi-random case is generated. Overall, this result confirms and extends the evidence summarized above which indicates that detection of social contingency can be properly characterized as a property of coupled dynamical systems that are regulated by simple control rules.

In Ref. [14] we studied a different approach to the problem of self-organization of interaction among agents which was based on an explicit maximization of the learning progress. This scheme, however, relied on the accumulation of previous knowledge and a discrete representation of the environmental information and is thus not a minimal model as required in [9].

Studies of the present kind are relevant to an emerging formal theory of social interaction. Results from numerical approaches essentially show that certain assumptions about the agents complexity are not necessary. However, the precise formulation of the forces in a social interaction cannot be identified by computer simulations. Nevertheless, we can conclude from the current study that the emergence of interaction with contingent agents might by rooted in a fundamental requirement to the sensorimotor loop, namely the maximal integration of the environment into the sensorimotor flow.

A Derivation of the Learning Rule

The agents receive a vector of sensor values $x_t \in \mathbb{R}^n$ at time steps $t = 0, 1, 2, \ldots$. The actions of the robot are determined by a controller described by a function K that maps sensor values $x \in \mathbb{R}^n$ onto motor values $y = (y_1, \ldots, y_m)^T \in \mathbb{R}^m$,

$$y_i = K_i(x) = \tanh\left(\sum_j C_{ij} x_j + h_i\right), \tag{7}$$

where in terms of neural networks C_{ij} denotes the weights and h_i the biases. The robot is further equipped with an adaptive model of its environment that is also realized by a neural network. It approximates a function F that predicts new sensor values based on earlier sensor and motor values via

$$x_t = F(x_{t-1}, y_{t-1}) + \xi_t. \tag{8}$$

The world model F is realized by a one-layer feed forward neural network. The "noise" term ξ_t in Eq. 8 is the modeling error which is assumed to be of finite variance. Inserting Eq. 7 into 8, we obtain a dynamical system

$$x_t = \psi(x_{t-1}) + \xi_t$$

representing the dynamics of the sensorimotor loop. The internal representation ψ is used to define an error function. Predictability is achieved by minimizing the prediction error ξ. Sensitivity can be expressed by the Jacobian matrix,

$$L_{ij}(x) = \frac{\partial}{\partial x_j} \psi_i(x),$$

which specifies the linear response of the map ψ to a perturbation. High sensitivity requires large values of all eigenvalues of L which can be achieved by the maximization of the smallest eigenvalue of L or equivalently by a minimization of the largest eigenvalue of L^{-1}. The error function is thus defined as

$$E = \left\|L^{-1}\xi\right\|^2 = \xi^T \left(LL^T\right)^{-1} \xi. \tag{9}$$

The parameters of the controller are updated by gradient descent on E, where

$$L_{ij} = \sum_k \left(\frac{\partial}{\partial y_k} F_i(x, y)\right) \tanh'\left(\sum_l C_{kl} x_l + h_k\right) C_{kj},$$

which leads to the following learning rule for the controller parameters.

$$\varepsilon^{-1} \Delta C_{ij} = \zeta_i v_j - 2\zeta_i \rho_i y_i x_j \tag{10}$$

$$\varepsilon^{-1} \Delta h_i = \qquad - 2\zeta_i \rho_i y_i \tag{11}$$

where $\rho = Cv$, $v = L^{-1}\xi$, $\zeta_i = g_i' \mu_i$, $\mu = (\frac{\partial}{\partial y} F(x,y))^\top (LL^\top)^\dagger \xi$ and $M^\dagger$ denotes the Moore-Penrose inverse of M. Note that the learning rate ε is chosen such that the parameters change at the same time scale as the behavior, sometimes even synchronously. The interplay between synaptic and state dynamics of the controller induces the potential for a high dynamical complexity of the sensorimotor loop.

Acknowledgments. G.M. thanks the LARAL for their hospitality during his stay in Rome in October 2007. This work was partially supported by a grant (01GQ0432) from the BMBF in the NNCS program and the ECAGENTS project funded by the Future and Emerging Technologies programme (IST-FET) of the European Community under EU R&D contract IST-1940

References

1. Der, R., Hesse, F., Martius, G.: Learning to feel the physics of a body. In: Proc. CIMCA-IAWTIC'06, Washington, DC, IEEE Computer Society (2005) 252–257
2. Der, R., Hesse, F., Martius, G.: Rocking stamper and jumping snake from a dynamical system approach to artificial life. Adaptive Behavior **14** (2006) 105–115
3. Der, R., Martius, G., Hesse, F.: Research webpage. `http://robot.informatik.uni-leipzig.de/Videos` (2008)
4. Der, R., Steinmetz, U., Pasemann, F.: Homeokinesis - a new principle to back up evolution with learning. In: Computational Intelligence for Modelling, Control, and Automation. Volume 55., IOS Press (1999) 43–47
5. Nolfi, S.: Emergence of Communication in Embodied Agents: Co-Adapting Communicative and Non-Communicative Behaviours. Connection Science **17** (2005) 231–248
6. Auvray, M., Lenay, C., Stewart, J.: The attribution of intentionality in a simulated environment: the case of minimalist devices. In: Tenth Meeting of the Association for the Scientific Study of Consciousness, Oxford (2006)
7. Trevarthen, C.: The self born in intersubjectivity: The psychology of an infant communicating. In: Neisser, U. (ed.) The perceived Self. Cambridge Univerity Press, Cambridge (1993) 121–173
8. Di Paolo, E.A., Rohde, M., Iizuka, H.: Sensitivity to social contingency or stability of interaction? modelling the dynamics of perceptual crossing. New Ideas in Psychology, Special Issue on Dynamics and Psychology (2008 (in press))
9. Iizuka, H., Di Paolo, E.A.: Minimal agency detection of embodied agents. In: Proc. 9th European Conference on Artificial Life ECAL 2007, Springer (2007)
10. Der, R., Martius, G., Hesse, F.: Let it roll – emerging sensorimotor coordination in a spherical robot. In L. M. Rocha et al., ed.: Artificial Life X : 10th Int. Conf. on the Simulation and Synthesis of Living Systems, MIT Press (2006) 192–198
11. Martius, G., Herrmann, J.M., Der, R.: Guided self-organisation for autonomous robot development. In Almeida e Costa, F., ed.: Advances in Artificial Life. Volume 4648 of Lecture Notes in Computer Science., Springer (2007) 766–775
12. Rohde, M., Di Paolo, E.: An evolutionary robotics simulation of human minimal social interaction. In: From Animals to Animats 9. Volume 4095 of Lecture Notes in Computer Science., Springer (2006)
13. Stormark, K.M., Braarud, H.C.: Infants' sensitivity to social contingency. Infant Behavior & Development **27** (2004) 195–203
14. Herrmann, J.M.: Dynamical systems for predictive control of autonomous robots. Theory in Biosciences **120** (2001) 241 – 252