Predictive information and emergent cooperativity in a chain of mobile robots

Ralf Der[1], Frank Güttler[2], Nihat Ay[1,3]

[1] Max-Planck Institute for Mathematics in the Sciences Leipzig, Germany
[2] University Leipzig Institute of Informatics, Germany
[3] Santa Fe Institute, Santa Fe, USA

Abstract

Measures of complexity are of immediate interest for the field of autonomous robots both as a means to classify the behavior and as an objective function for the autonomous development of robot behavior. In the present paper we consider predictive information in sensor space as a measure for the behavioral complexity of a chain of two-wheel robots which are passively coupled and controlled by a closed-loop reactive controller for each of the individual robots. The predictive information, the mutual information between the past and the future of a time series, is approximated by restricting the time horizons to a single time step. This is exact for Markovian systems but seems to work well also for our robotic system which is strongly non-Markovian.When in a maze with many obstacles, the approximated predictive information of the sensor values of an individual robot is found to have a clear maximum for a controller which realizes the spontaneous cooperation of the robots in the chain so that large areas of the maze can be visited.

Introduction

Despite much progress in biologically inspired robotics, biological systems are still singled out by a high degree of self-actualisation. This phenomenon is approached by the scientific community on different levels. Concepts like autopoiesis (Maturana and Varela, 1980) try to provide a general theoretical framework for the phenomena of self-creation and self-maintenance of living beings. On the other hand, concrete modes of action are formulated by mechanisms like homeostasis as a general theory of self-regulation (Ashby, 1954). It is widely believed that the integration of self-phenomena into artificial beings would not only lead to a better understanding of living beings but also to robots with internal motivation, curiosity, the self-exploration of bodily and environmental affordances, and quite generally to creative behaviors.

There are many different approaches towards the self-actualisation of behavior in autonomous robots. Relevant for this paper is the attitude that behavior is less a sequence of actions in order to reach a prespecified goal but instead a means for (i) structuring the input information (creating statistical correlations) the robot gathers with its sensors (Lungarella and Sporns, 2005); (ii) the maximization of the information flow in the sensorimotor loop (empowerment) (Klyubin et al., 2007); (iii) the maximization of the sensorimotor coordination (Lungarella and Sporns, 2006), and others. The main question is how this can be realized. There are interesting approaches of realizing systems on the basis of concrete modes of action like homeostasis, see for instance (di Paolo, 2003), but a more systematic way is by convenient measures for the information contained in or the complexity of the sensor stream. Of methodological interest are approaches of formulating general measures for the realisation of self-organisation (Shalizi et al., 2004).

This paper tries to further develop this direction in a concrete embodied robotic system. In order to further systematize the field we introduce the notion of self-referential robotic systems – adaptive, embodied systems where the objective of adaptation is a function of the robot's sensor values alone. In particular, there is no domain specific goal or externally specified aim formulated into this function. We favor predictive information measuring the complexity in sensor space as such an objective function. The predictive information of a process quantifies the total information of past experience that can be used for predicting future events. Technically, it is defined as the mutual information between the future and the past (Bialek et al., 2001). It has been argued that predictive information, also termed excess entropy (Crutchfield and Young, 1989) and effective measure complexity (Grassberger, 1986), is the most natural complexity measure for time series. The behaviors emerging from maximizing the PI are qualified by the fact that predictive information is high if – by its behavior – the robot manages to produce a stream of sensor values with high information content under the constraint that the consequences of the actions of the robot remain still predictable. This is why we favor predictive information as an objective function.

Under this paradigm, behaviors are entirely contingent, depending on the physical embodiment of the robot and the starting and environmental conditions. From the point of view of applications, the question of central interest is what kind of behaviors may be expected to arise with a given embodiment and as a next step whether these behaviors are of any interest as behavioral primitives for the construction of higher-level goal-oriented strategies. It is in the nature of the question that there is not a unique answer but we are convinced that a certain systematics can be found at the level of phenomena. This paper considers the case of a chain of passively coupled two-wheel robots, each robot being controlled independently by a simple neural network under the closed-loop control paradigm. There is no central control so that a coherent motion of the chain is an emerging phenomenon based on a synchronisation of the wheels of the individual robots, the task being aggravated by the fact that the robots are moving in a maze with only narrow passages between obstacles. Nevertheless, we show that the predictive information of a single sensor value (the wheel velocity) of an individual robot is in close relationship to the ability of the chain to spontaneously self-organize into a coherent mode. In this mode, the chain may successfully navigate in the maze. This result extends the earlier finding (Ay et al., 2008) that the maximum MI in the sensor channels defines a working regime where the controller reacts in a specific way to the sensor values.

Our approach relates to other approaches of using statistical measures for robotics, a good introduction is (Lungarella et al., 2005) where a set of univariate and multivariate statistical measures are used in order to quantify the information structure in sensory and motor channels, see also (Klyubin et al., 2007) and (Klyubin et al., 2005). In particular we consider the predictive information as a prospective tool for concepts like internal motivation. Potential applications of this approach are expected in developmental robotics which has found some interest recently (Lungarella et al., 2003). There is a close relationship to the attempts of guiding autonomous learning by internal reinforcement signals (Stout et al., 2005) and to task independent learning (Oudeyer et al., 2005), (Schmidhuber, 2005), (Still, 2007). Quite generally, using a complexity measure as the objective function for the development of a robot corresponds to giving the robot an internal, task independent motivation for the development of its behavior.

The robot

In the present paper we are considering a chain of passively coupled two-wheel robots, Fig. 1, simulated in the *lpzrobots* simulation tool (Martius and Der, 2007) based on the physics engine ODE (Smith, 2005), which simulates in a realistic way effects due to the inertia of the robot, slip and friction the effects of the wheels with the ground and the effects of both the couplings and collisions. Each individual robot has a controller consisting of two neurons with the vector $x \in \mathbf{R}^2$ of the measured wheel rotation velocities as input and the vector $y \in R^2$ of nominal motor activities as output, i.e.

$$y_i = g\left(C_{i1}x_1 + C_{i2}x_2\right) \qquad (1)$$

Figure 1: In the arena the chain of passively coupled two-wheel robots is simulated in the *lpzrobots* simulation tool. Each robot is "blind" and feels the environment only by the reactions of its wheel counters on collissions with the obstacles.

where $g\left(z\right) = \tanh z$, the controller matrix C defining the behavior of the system. In the present paper we want to determine empirically the predictive information over the controller parameters C_{ij} which parameterize the behavior of the robot.

The sensorimotor loop

If the wheels are moving freely we may assume that the nominal velocity y and the measured true velocity x are equal. In a realistic situation there will be perturbations so that we write the sensorimotor dynamics

$$x_{t+1} = y_t + \xi_{t+1} \qquad (2)$$

where $x_t = \left(x_{t1}, x_{t2}\right)^T \in R^2$ and ξ contains all the effects due to friction, slip, inertia and so on which make the response of the robot to its controls uncertain. In particular, if the robot hits an obstacle, the wheels may get totally or partially blocked so that in this case ξ may be large, possibly fluctuating with a large amplitude if the wheels are not totally blocked. Moreover ξ will also reveal whether the robot hits a movable or a static object. Additional strong effects result from the couplings between the robots which exert strong forces if the robots are not in complete synchrony.

In order to discuss the nature of the spontaneous cooperation phenomena observed we consider the trivial (but relevant, see below) case of a diagonal matrix C with $C_{11} = C_{22} = c$ so that the sensorimotor loop of each wheel is described by the one-dimensional system ($x_t \in R^1$)

$$x_{t+1} = g\left(cx_t\right) + \xi_{t+1}$$

the properties of which are obtained by analysing the fixed points obtained from

$$x = g(cx)$$

As discussed in earlier papers (Ay et al., 2008), the system has one stable FP for $0 < c < 1$ which becomes unstable at the bifurcation point $c = 1$ so that for $c > 1$ we have a bistable system with FPs $x = \pm q$ with q increasing for increasing c. The noise causes fluctuations around the FPs with occasional switching of the FPs, the probability of switching decreasing exponentially with increasing $c > 1$. There is however a subtlety in the fact that, under the noise, the bifurcation is effectively taking place only at the so called effective bifurcation point which is at $c = 1 + \delta$ with $0 < \delta \ll 1$ and δ increasing with the noise. This region is of particular interest since it is there that the wheel velocities are already quite large but can easily be switched by noise events caused for instance by collisions with obstacles or by the influence of the forces exerted by the other robots in the chain. In fact, due to the effect described, the wheel velocity will feel a tendency to switch sign if a torque in the opposite direction is exerted on it by the other robots in the chain. By switching the velocity, the wheel is now acting in the direction of the force exerted on it and this is the self-amplification effect necessary for the occurrence of self-organization.

The videos at (Martius and Der, 2007) demonstrate quite clearly the strength of this self-organized synchronization effect which not only makes the robot chain move into one direction but also keeps it still explorative in the sense that after some time it also inverts its direction of motion. Moreover, when colliding with a wall the chain of robots often will change velocity in an integrated manner. Finally and most importantly for the topic of the present paper, it will also effectively explore the spatial extensions of a maze the chain is put into.

Information theoretic measures

The central aim of this paper is the relation between the internal world of the robot, based on a complexity measure of its sensor values, and its relation to the external world. As motivated above, a convenient complexity measure is predictive information in sensor space, i.e. we consider the time series $S = \{X_t | t = 0, 1, 2, \ldots\}$ of the sensor values of the behaving robot.

Predictive information

The predictive information is the mutual information between the future and the past, relative to some instant of time t, of the time series S

$$I(X_{past}; X_{future}) = \left\langle \log_2 \frac{p(X_{past}, X_{future})}{p(X_{past})\, p(X_{future})} \right\rangle$$

where the averaging is over the joint probability $p(X_{past}, X_{future})$, time horizons of both past and future extending to infinity. This expression simplifies considerably if X is a Gauss-Markov process, see (Ay et al., 2008). In this case the time horizon can be restricted to just a single step so that the PI is given by the mutual information (MI) between two successive time steps, i.e.

$$I(X_{past}; X_{future}) = \left\langle \log_2 \frac{p(X_{t-1}, X_t)}{p(X_{t-1})\, p(X_t)} \right\rangle \quad (3)$$

which simplifies the sampling process considerably. Moreover, in the experiments we observed that it is sufficient to study the MI of just a single sensor, one of the wheel counters of an individual robot, and still get the full information on the behavior of the robot chain.

Of course our time series of sensor values is far from being a Gauss-Markov process. However, as shown in (Ay et al., 2008) in specific cases the full PI can very well be approximated by that of a process with white Gaussian noise of conveniently chosen strength. The reason for this agreement probably is in the fact that both in linear and in weak noise nonlinear dynamical systems the PI does not depend on the noise at all. The PI however was found to depend very sensitively on the parameters of the controller which define the behavior of the robot.

This result is very important for the practical use of the PI. In fact, it tells us that in many cases the actually infinite time horizons may be restricted to just a few steps without losing much of the information on the behavior of the system as a function of the controller parameter.

The self-referential robotic system

The PI is given in terms of the sensor values the robot produce in the course of time alone. There is no domain specific knowledge invoked into this function. We obtain a self referential robotic system when using the PI as the objective function for the adaptation of the parameters of the controller. In particular we may consider the gradient ascent on the MI as given by eq. 3

$$\Delta m_t = \varepsilon \frac{\partial I(X_t; X_{t-1})}{\partial m_t}$$

where m is any parameter of the controller of the robot. The properties of the self-referential robotic system depends also on the choice of the learning rate ε which actually has to be chosen small enough so that the time scales are well separated.

The learning rule has been substantiated earlier (Ay et al., 2008) for the case of a simple sensorimotor loop and was shown to reduce to a simple synaptic dynamics consisting of a general driving term plus an anti-Hebbian learning term. The example shows that the sampling problem with the PI can be partially avoided and the gradient obtained explicitly if convenient approximations are made. We do, however, not aim to derive concrete learning rules in this paper but instead try to further elucidate the role of the PI with restricted time horizons.

The experiment

In order to keep the sampling effort manageable we use, based on symmetry arguments, two different parametrizations of the matrix C chosen such that there are only two parameters to be varied. The experiments have been carried through on a LINUX cluster of the Max Planck Institute for the Mathematics in the Sciences with about 100 nodes and have been run for $500,000$ time steps each. Results are averages over three runs for each pair of parameter values, see below.

Symmetric cross channel couplings

The first parametrization of the matrix C is given by

$$C = \begin{pmatrix} c & b \\ b & c \end{pmatrix} \tag{4}$$

We may extend the FP analysis of a single robot given above to the present case by assuming that the wheel velocities are $x_1 = x_2 = v$ (straight on motion) or $x_1 = -x_2 = v$ (on-site rotation) so that the FP is now obtained from the solution of

$$v = g\left(rv\right)$$

where $r = c + b$ or $r = c - b$ plays now the role of the feed-back strength in the loop for the straight or rotational motion, respectively. Obviously, with $b > 0$ we find that in the bifurcated region ($r > 1$) FPs are more stable for the straight on motion whereas with $b < 0$ the rotational motion is favored.

The central questions of our investigation is the behavior of the MI as a function of the behavior parameters of the robot and its relation to the behavior in physical space. Our robot chain has complicated physical properties, the videos might give an impression of the range of behavioral possibilities. When in the maze, the information transmission between the individual robots, which takes place by the physical forces transmitted via the passive coupling elements in a rather intricate way, is further corrupted by the collisions of the robots with the obstacles and the bumpers of adjacent robots in the chain, see the videos. Nevertheless, we observe for certain parameter combinations of the controller that the robot chain covers a wide area of the maze which is a clear indication of successful cooperation between the individual robots.

Figure 2 shows the MI as a landscape over the parameters c and b of the controller. We find a clear ridge structure the ridge running along the curves given by

$$c + |b| \approx 1.1$$

which means that the MI has a relative maximum close to the effective bifurcation point (now realized in the coupled system) be it either in the rotational ($b < 0$) or straight on ($b > 0$) mode. The rotation mode seems a little surprising at this point since in the chain the individual robot can not rotate. Most probably this is explained by the fact that we evaluated the MI for the first robot in the chain, which would execute an oscillatory motion by switching between the two rotational modes repeatedly. In further experiments we will evaluate the MI for the inner robots as well. The landscape moreover displays a clear local maximum which is at $b = 0$ and $c \approx 1.1$ meaning that the two channels are decoupled so that the best cooperation in the chain is if each wheel is controlled individually such that its single channel MI is maximal. This is also a little surprising since one would have expected that cooperation in the chain is best if the straight on motion is supported.

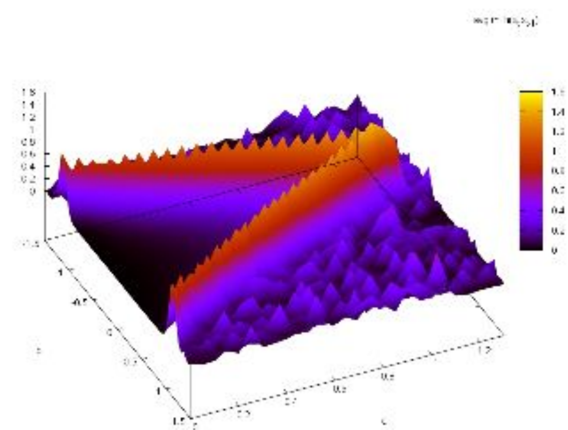

Figure 2: Th average mutual information of the sensor value for the case of symmetric cross channel couplings where $c = C_{11} = C_{22}$ and $b = C_{12} = C_{21}$.

The diversity of sites visited by the robot in the course of a fixed time is measured by the entropy of the probability distribution over the sites. As seen from Fig. 3, the landscape is very similar to the that of the MI. We have the absolute maximum at the same position so that the main message of the MI (individual control of the wheels) is corroborated. However there are also differences. On the one hand we see that the ridge corresponding to a preferred rotational motion ($c < 0$) is not so high as the one for the straight on motion. Moreover we find that there is a second clear maximum at around $c = 0$ and $b = 1.1$ corresponding to the case that the direct coupling is zero so that the control of the wheel is completely based on the angular velocity of the opposite wheel. The MI seems to have also a small local maximum in this region but this needs corroboration by a better statistics.

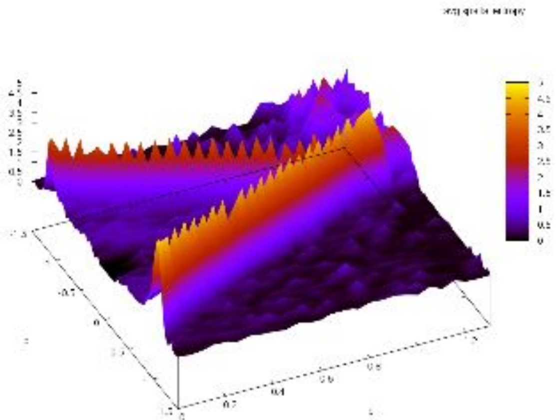

Figure 3: The entropy of the probability distribution over the sites that could be visited by the chain of robots in the maze. The entropy is over the parameters c and b of the couplings in and across channels. The entropy is maximal if all sites are visited by the chain with equal probability and is zero if the robot remains in its starting position.

Nevertheless there is a strong correlation between the MI in sensor space and the behavior of the robot as measured in physical space. This has an even stronger implications than in the single robot case considered in (Ay et al., 2008). Noting that the MI is taken by considering just one of the sensor values (wheel velocities) of an individual robot (the first one in the chain) we may conclude that the adaptation according to the maximum MI principle makes the robot capable of effectively cooperating in a collective of robots without any central control.

Antisymmetric cross channel couplings

Our second parametrization is taken as

$$C = \begin{pmatrix} c & -b \\ b & c \end{pmatrix} = \alpha \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}$$

which is known to support a Neimark-Sacker bifurcation if α exceeds 1 into an oscillatory regime (Pasemann et al., 2003), with frequency roughly given by $f \sim \phi/2\pi$. The landscape of the MI and the spatial entropy now are even more similar (although the ridge of the spatial entropy landscape is more pronounced) so that we depict only that of the MI. We see again the maximum of the single channel control ($\alpha = 1.1$, $\phi = 0$) as observed above. However, surprisingly there is clear second maximum at $\alpha \approx 2.35$ which corresponds to a very strong direct coupling combined with a very high frequency of about 3 Hz of the oscillations of the wheel velocities. This maximum is clearly seen for the MI and is even more pronounced for the entropy of the spatial distribution. In order to understand the phenomenon that the chain well manages to navigate in the maze with the strategy of rapidly (but still sensitively since under the closed loop control paradigm) switching wheel velocities we have to note that in our simulations we use values for the friction and slip parameters corresponding to a snow underground. This setting was chosen with the intention that emerging cooperation is possible best for a chain of sensitive "drivers". The strange high frequency regime is counterintuitive to this argument. Possibly this strategy is useful since it may excite a kind of navigation by controlled skidding but this will need further investigations.

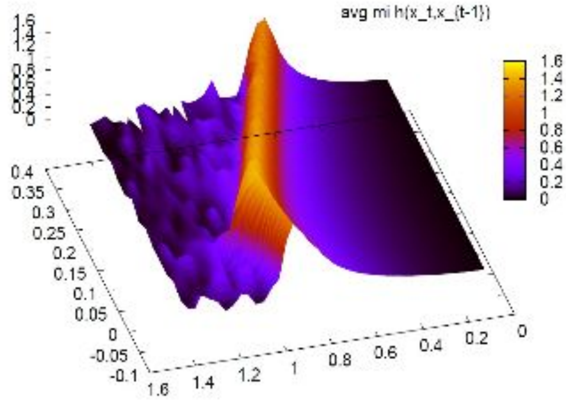

Figure 4: The MI for the case of the antisymmetric cross channel coupling. The landscape is only for positive ϕ because of the symmetry against sign inversion of ϕ. There are two local maxima of about the same height at ($\alpha = 1.1, \phi = 0$) and ($\alpha = 2.35, \phi = .35$).

The decisive point however is that obviously under the present parametrization the MI singles out specific nontrivial behavior modes which unexpectedly represent an effective control strategy. This is one further hint for the usefulness of the predictive information as a general tool for the self-organisation of behavior.

Concluding remarks

This paper has investigated the usefulness of predictive information for the self-organisation of behaviour in a chain of passively coupled robots. Predictive information has been approximated by the mutual information of sensor values over one time step which is much better accessible in real systems. Despite of this drastic simplification, we have shown that the maximum of the MI specifies a working regime of the single robot where it can effectively cooperate in the chain under difficult environmental conditions. Thus, the predictive information of even a single sensor channel is seen to be a far reaching indicator of the global behavior in physical space. In other words it is a link between the internal world of the robot (sensor space) and the behavior in the external world which is maximum if the behavior of the robot is "rich" but with a high degree of self-established sensorimotor coordination.

This concept will be continued in further work where we will in particular investigate in how far the extension of the time horizon, in particular into the past, will give measures which are more discriminative. For instance, we want to understand if such a more extended measure is able of discriminating between the two control modes of preferentially straight or rotational modes which are different in their spatial behavior but not so much in the MI. The present results clearly support the point of view that the link between the information measure in sensor channels and the behavior of the robot is of a more fundamental nature, as claimed for instance in (Lungarella and Sporns, 2006). This suggests, as a possible application, the use of the PI as an auxiliary fitness function in artificial evolution which helps driving agents into working regimes with high prospectives for emerging functionalities. This will be one of our future projects.

Another focus is on the relation of the PI to another complexity measure, the so called time loop error, and the principle of homeokinesis (Der and Liebscher, 2002), (Der et al., 1999), (Der, 2001), which has been the basis for concrete learning rules leading to the self-organization of explorative behaviors in complex robots with many degrees of freedom in dynamic, unstructured environments, see (Der et al., 2006), (Der and Martius, 2006), (Der et al., 2005) and the videos on `http://robot.informatik.uni-leipzig.de/`. We hope in the near future to produce similar results on the basis of information theoretic measures. Preliminary results indicate that the gradients of the time loop error and the mutual information can be related to each other by a change in the metric of the parameter space.

Acknowledgements

Nihat Ay thanks the Santa Fe Institute for supporting him as an external professor. The facilities provided by the LINUX cluster of the Max Planck Institute for the Mathematics in the Sciences are gratefully acknowledged.

References

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

Ay, N., Bertschinger, N., Der, R., Güttler, F., and Olbrich, E. (2008). Predictive information and explorative behavior of autonomous robots. *European Physical Journal: Complex Systems*. to appear.

Bialek, W., Nemenman, I., and Tishby, N. (2001). Predictability, complexity and learning. *Neural Computation*, 13:2409.

Crutchfield, J. P. and Young, K. (1989). Inferring statistical complexity. *Phys. Rev. Lett.*, 63:105–108.

Der, R. (2001). Self-organized acquisition of situated behavior. *Theory in Biosciences*, 120:179–187.

Der, R., Hesse, F., and Martius, G. (2005). Rocking stamper and jumping snake from a dynamical system approach to artificial life. *J. Adaptive Behavior*, 14:105 – 116.

Der, R. and Liebscher, R. (2002). True autonomy from self-organized adaptivity. In *Proc. Workshop Biologically Inspired Robotics. The Legacy of Grey Walter 14-16 August 2002, Bristol Labs*, Bristol.

Der, R. and Martius, G. (2006). From motor babbling to purposive actions: Emerging self-exploration in a dynamical systems approach to early robot development. In Nolfi, S., editor, *From Animals to Animats*, volume 4095 of *Lecture Notes in Computer Science*, pages 406–421. Springer.

Der, R., Martius, G., and Hesse, F. (2006). Let it roll – emerging sensorimotor coordination in a spherical robot. In Rocha, L. M., editor, *Artificial Life X*, pages 192–198. MIT Press.

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

di Paolo, E. (2003). Organismically-inspired robotics: Homeostatic adaptation and natural teleology beyond the closed sensorimotor loop. In Murase, K. and Asakura, T., editors, *Dynamical Systems Approach to Embodiment and Sociality*, pages 19 – 42, Adelaide. Advanced Knowledge International.

Grassberger, P. (1986). Toward a quantitative theory of self-generated complexity. *Int. J. Theor. Phys.*, 25(9):907–938.

Klyubin, A. S., Polani, D., and Nehaniv, C. L. (2005). Empowerment: A universal agent-centric measure of control. In *Proc. CEC*. IEEE.

Klyubin, A. S., Polani, D., and Nehaniv, C. L. (2007). Representations of space and time in the maximization of information flow in the perception-action loop. *Neural Computation*, 19:2387–2432.

Lungarella, M., Metta, G., Pfeifer, R., and Sandini, G. (2003). Developmental robotics: a survey. *Connect. Sci.*, 15(4):151–190.

Lungarella, M., Pegors, T., Bulwinkle, D., and Sporns, O. (2005). Methods for quantifying the informational structure of sensory and motor data. *Neuroinformatics*, 3(3):243–262.

Lungarella, M. and Sporns, O. (2005). Information self-structuring: Key principle for learning and development. In *Proceedings 2005 IEEE Intern. Conf. Development and Learning*, pages 25–30.

Lungarella, M. and Sporns, O. (2006). Mapping information flow in sensorimotor networks. *Comput Biol*, 2(10):e144.

Martius, G. and Der, R. (2007). lpzrobots - simulation tool for autonomous robots. `http://robot.informatik.uni-leipzig.de/`.

Maturana, H. R. and Varela, F. J. (1980). *Autopoiesis and Cognition. The Realization of the Living*. Reidel, Dordrecht.

Oudeyer, P.-Y., Kaplan, F., Hafner, V. V., and Whyte, A. (2005). The playground experiment: Task-independent development of a curious robot. In Bank, D. and Meeden, L., editors, *Proceedings of the AAAI Spring Symposium on Developmental Robotics, 2005, Pages 42-47, Stanford, California, 2005.*

Pasemann, F., Hild, M., and Zahedi, K. (2003). SO(2)-networks as neural oscillators. In Mira, J. and Alvarez, J., editors, *Computational Methods in Neural Modeling*, pages 144–151, Berlin, Heidelberg, New York. Springer.

Schmidhuber, J. (2005). Completely self-referential optimal reinforcement learners. In *ICANN (2)*, pages 223–233.

Shalizi, C. R., Shalizi, K. L., and Haslinger, R. (2004). Quantifying self-organization with optimal predictors. *Phys Rev Lett*, 93(11):118701.

Smith, R. (2005). Open dynamics engine. `http://ode.org/`.

Still, S. (2007). Statistical mechanics approach to interactive learning. *arXiv:0709.1948v1 [physics.data-an]*. submitted.

Stout, A., Konidaris, G., and Barto, A. (2005). Iintrinsically motivated reinforcement learning: A promising framework for developmental robotics. In *The AAAI Spring Symposium on Developmental Robotics*.