Predictive information and explorative behavior of autonomous robots

Nihat Ay[*+], Nils Bertschinger[*], Ralf Der[*#], Frank Güttler[#], Eckehard Olbrich[*]
[*] Max-Planck Institute for Mathematics in the Sciences Leipzig, Germany
[#] University Leipzig Institute of Informatics, Germany
[+] Santa Fe Institute, Santa Fe, USA

March 5, 2008

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 two-wheel embodied robot moving in a rectangular arena with several obstacles. The mutual information (MI) between past and future sensor values is found empirically to have a maximum for a behavior which is both explorative and sensitive to the environment. This makes predictive information a prospective candidate as an objective function for the autonomous development of such behaviors. We derive theoretical expressions for the MI in order to obtain an explicit update rule for the gradient ascent dynamics. Interestingly, in the case of a linear or linearized model of the sensorimotor dynamics the structure of the learning rule derived depends only on the dynamical properties while the value of the MI influences only the learning rate. In this way the problem of the prohibitively large sampling times for information theoretic measures can be circumvented. This result can be generalized and may help to derive explicit learning rules from complexity theoretic measures.

1 Introduction

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, see [1]. It has been shown that predictive information, also termed excess entropy [4] and effective measure complexity [11], is the most natural complexity measure for time series. This concept is of immediate interest for the field of autonomous robots if applied to the time series of sensor values the robot produces. The

difference to classical time series analysis is in the fact that the robot generates these time series by its behavior so that behavior can be related to the complexity of the time series. Thus, on the one hand we may use complexity theory in order to classify the behavior of robots in interaction with the environment. On the other hand, once such a measure is established it can be used as an objective function for the self-organization of behavior of the robot.

The self-organization scenario we have in mind is completely based on the internal perspective of the robot i.e. the adaptation of the behavior is driven by an objective function which is based on the time series of the sensor values alone. Predictive information seems to be a good candidate for the self-organization of environment related explorative behavior. In fact, 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. The behaviors emerging from maximizing the predictive information (like any other complexity measure) depend in an essential way on the embodiment of the robot in its interaction with the environment. This paper aims at investigating, in a concrete embodied robot experiment, the link between the complexity measure in sensor space and the realization of the behavior in physical space. We use a robotic system that is simple enough to be treated analytically but reflects already much of the general case. In particular our robotic system is fully embodied in the sense that physical influences like inertia, collisions and so on play an essential role. However, we do not study the full predictive information but restrict ourselves to the mutual information (MI) between successive time steps which is equal to the predictive information in the case of Markovian systems, see below. We show by both theoretical analysis and experimental results, that the maximization of the predictive information defines a working regime of the robot where it is particularly explorative (richness in dynamics) while being in good sensor contact with the environment (high predictability of future events).

Our approach relates to other approaches of using statistical measures for robotics, a good introduction is [16] 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 [14] and [13]. 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 [25] [15]. There is a close relationship to the attempts of guiding autonomous learning by internal reinforcement signals [24] and to task independent learning [19], [21], [23]. 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 paper is organized as follows: We introduce in Sec. 2 the robot and then give a dynamical systems analysis of its behavior. In particular we introduce the concept of the effective bifurcation point (BP). This analysis is helpful in understanding the different behavioral regimes realized by the robot. Sec. 3 introduces the information theoretic measures and gives a theoretical expression

for the case at hand. After this we present in Sec. 4 the results of experiments with the simulated robot showing that the MI has a maximum close to the effective bifurcation point where the robot is seen to cover the largest distances without losing its sensitivity against collisions with the environment. Finally in Sec. 5 we formulate a general learning rule for the parameters of the controller based on the gradient ascent of the mutual information as obtained by the theory of Sec. 3. This is seen to be an appropriate way to avoid the sampling problem associated with the empirical MI measure.

2 The robot

In the present paper we are using a simple two-wheel robot simulated in the *lpzrobots* simulation tool [18] based on the physics engine ODE, see [22]. Each wheel is driven by a motor, the motor values being given by the vector $y_t \in \mathbf{R}^2$ which is the output of the controller. The only sensors are wheel counters measuring the true velocity of each of the wheels, i.e. $x_t \in \mathbf{R}^2$ is the vector of the measured wheel rotation velocities. The physics engine ODE simulates in a realistic way effects due to the inertia of the robot, slip and friction effects of the wheels with the ground and the effects of collisions. The velocities are such that the robot upon collisions may tumble so that we have a truly embodied robotic system.

2.1 The control paradigm

There are many different paradigms for the control of autonomous robots. In the present paper we consider closed loop control with a tight sensorimotor coupling. The controller is a function

$$y = K(x) \tag{1}$$

mapping sensor values $x \in \mathbf{R}^n$ to motor values $y \in \mathbf{R}^m$. We restrict ourselves in the present paper to a purely reactive controller. In more general cases the controller might additionally depend on an internal state. In the concrete setting, the sensor values are the velocities of the wheels as measured by the wheel counters, the outputs y being the target velocities of the wheels. There are a few conditions the controller must fulfill for physical reasons. On the one hand, the controller outputs must be limited by the maximum velocity the robot can realize. On the other hand, due to the directional symmetry of the robot used in the experiments, the controller should be invariant with respect to inverting the input and output velocities simultaneously. For the sake of simplicity we use a pseudo linear expression

$$y_i = g(C_{i1}x_1 + C_{i2}x_2) \tag{2}$$

where $i = 1, 2$, and require additionally that the function $g(z)$ is monotonic. Due to the symmetry and boundedness argument an antisymmetric sigmoid function

is a natural choice for $g(z)$. We use in the present paper $g(z) = \tanh(z)$. Any other sigmoid function will produce qualitatively similar results as can be seen in terms of the analysis given below.

In the present paper we want to determine empirically the predictive information over the coupling parameters C_{ij} defining the behavior of the robot. In order to keep the sampling effort manageable we omit the cross channel couplings, i.e. $C_{12} = C_{21} = 0$. Due to the right-left symmetry of the robot we also put $C_{11} = C_{22} = c$ so that our matrix C is

$$C = \begin{pmatrix} c & 0 \\ 0 & c \end{pmatrix} \tag{3}$$

and there is only one parameter determining the behavior of the robot.

2.2 The sensorimotor loop

Taking the internal perspective, the only information available to the robot is the time series of its sensor values $x_t \in R^n$, $t = 1, 2, \ldots$. In order to "understand" the world (its body embedded dynamically into the environment), the robot may use the following model of the time series x_t

$$x_{t+1} = F(x_t, y_t) + \xi_{t+1} \tag{4}$$

where in general $F : R^n \times R^m \to R^n$ is a function mapping old sensor and motor values to the new sensor values with $\xi \in R^n$ being the modelling error. In practical applications F may be realized by a neural network which can be trained by supervised learning. In our simplistic case, when in unperturbed motion, the observed wheel velocities are essentially those prescribed by the controller, i.e. $x_{t+1} = Ay_t$ where the matrix A is given by $A_{ij} = a\delta_{ij}$ with a hardware constant a which we may set $a = 1$ so that eq. (4) boils down to

$$x_{t+1} = y_t + \xi_{t+1} \tag{5}$$

where 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, possible 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.

Using eq. (1) in eq. (4) we may write the sensorimotor dynamics as

$$x_{t+1} = \psi(x_t) + \xi_{t+1} \tag{6}$$

where $\psi(x) = F(x, K(x))$. In the specific case of eq. (5) we have

$$\psi(x) = G(Cx) \tag{7}$$

where G is the vector function $G : R^2 \to R^2$, $G_i(z) = g(z_i) = \tanh z_i$ with $z_i = C_{i1}x_1 + C_{i2}x_2$ for $i = 1, 2$ and thus

$$x_{t+1} = G(Cx_t) + \xi_{t+1} \tag{8}$$

Although the robot may behave in a very intricate way (see below), eq. (6) is exact, since the effects of the embodied interaction with the world are concealed in the model error ξ. In the theoretical analysis given below we will consider ξ as a random number (white Gaussian noise) in order to obtain an explicit expression for the predictive information which forms the basis of our learning rule.

2.3 Properties of the single channel dynamics

Let us now consider at first the case of identical wheel velocities, i.e. the robot is moving along a straight line. Dropping the model error (noise) for the moment, the stationary behavior of the robot is given by the fixed points (FPs) of eq. (8). We consider each loop independently (uncorrelated noise) with fixed point equation

$$x = \tanh(cx) \tag{9}$$

Standard FP analysis shows that there is a stable FP $x^* = 0$ for $0 < c < 1$. With $c > 1$ the FP $x^* = 0$ becomes unstable and there are two new, stable FPs $x^* = \pm u$ where for small u we get by means of the Taylor expansion $\tanh z \approx z - z^3/3$ in leading order the FP equation $x = cx - (cx/3)^3$ with solution

$$x^* = \pm\sqrt{3\frac{(c-1)}{c^3}} \tag{10}$$

valid for $c = 1 + \delta$ with $0 < \delta \ll 1$ in leading order of δ. On the other hand we find trivially $x^* \to \pm 1$ for $c \to \infty$ directly from eq. (9).

The discussion of the properties of the dynamics is most conveniently done by rewriting the stochastic dynamical system as a gradient descent on a potential V. In terms of the state variable $z_t = cx_t$ we have

$$\Delta z_t = -\frac{\partial V(z_t)}{\partial z_t} + c\xi_{t+1}$$

where $\Delta z_t = z_{t+1} - z_t$,

$$V(z) = \frac{z^2}{2} - c \ln\cosh z$$

and $\frac{\partial \ln\cosh z}{\partial z} = \tanh z$ was used. The potential has a single minimum at $z = 0$ for $0 < c < 1$ and it is a double well potential for $c > 1$, see Fig. 1. According to this picture, the behavior of the robot is characterized by the following three scenarios:

1. In the subcritical case, i.e. below the bifurcation point ($c = 1$), the velocity of the robot is fluctuating, due to the noise, around zero with amplitude increasing with c. Hence the robot executes a random walk with variance increasing with c. When encountering a wall it will fluctuate in front of the wall until a longer sequence of random events ξ carries it away.

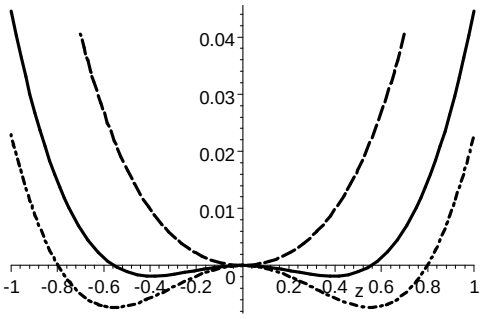

Figure 1: The potential $V(z) = \frac{z^2}{2} - c \ln \cosh z$ for $c = 0.9$ (dashed), $c = 1.05$ (solid), and $c = 1.1$ (dash-dotted). The gradient dynamcs drives the state z to the next fixed point. With noise included, the state fluctuates around the fixed point with an amplitude given by the width of the potential well. In the double well region, the noise can cause occasional switches between the wells, the switching frequency decreasing exponentially with the barrier height, see [20].

2. In the supracritical region with $c \gg 1$ the velocity is fluctuating around one of the stable FPs with amplitude being the smaller the larger c. Hence the robot is moving forever (in physical times) into one direction with more or less constant velocity. Inversion of velocity can take place only if the wheels are totally blocked, i.e. $x_t = 0$ followed by a random event ξ into the appropriate direction. The forces exerted by the robot are very high due to the strong amplification factor c (leading to $y \approx \pm 1$ even if x is already small). Movable objects do not stop the robot so that it can not discern by its behavior between light and heavy movable obstacles.

3. Eventually, there is a critical region around some value $c_{opt} > 1$ where the noise is able to switch the state between the FPs with a substantial rate. We call this (fuzzy) point the effective bifurcation point. In this region the robot executes long distance sweeps of different lengths into both directions. Due to the smaller amplification rate c, forces are more differentiated so that, by its behavior, the robot may discern between light and heavy movable objects.

It is mainly in the critical region that the robot covers both large distances in either direction and is sensitive to collisions with an obstacle: If the obstacle is fixed the robot will reverse its velocity (after some time) due to the noise amplification ($c > 1$). If the object is movable the robot will either retract or start moving the object depending on its weight. Due to slip and friction effects, in this critical regime the robot often stops moving the object after some time

so that a highly variate behavior of the robot is observed. It is to be noted that these properties, based on proprioceptive sensors (wheel counters) only, are a direct consequence of the closed loop control paradigm used.

2.4 The two-dimensional case

The fixed point analysis obtained for the one-dimensional case readily carries over to the two-wheel robot. Ignoring the noise, the controllers of the wheels are completely independent, each controller working only in its sensorimotor loop. Hence with $0 < c < 1$ both sensorimotor loops have FP $z = 0$ and with $c > 1$ we have two FPs for each loop corresponding to the behavior modes rotating on-site to the left or right and moving forward or backward on a straight line.

With given noise the most interesting regime is observed again about the effective bifurcation point. The robot is expected (and observed) to cover large distances but still reacts sensitively to the collisions with obstacles. In particular, by a collision it can be carried over from a straight line to a rotating behavior. The latter can be left if close to the effective bifurcation point. However for large c values, the robot will be caught for exceedingly long times in this rotational mode so that the exploration breaks down.

It is to be noted that, due to physical effects, the two sensorimotor loops are not independent since the wheels are connected by the body. Formally this is contained in the noise ξ. For instance, if the robot collides with some obstacle, the effect on the wheels is strongly correlated. In a head on collision both wheels may be blocked simultaneously which gives a large noise event in both channels simultaneously. Moreover a sudden change in the velocity of one wheel will have an effect on the other wheel due to the inertia effects mediated by the body.

3 Information theoretic measures

The aim of the present section is to derive theoretical expressions for the mutual information based on assumptions made on the noise character of the model error of eq. (6). As discussed above, ξ contains the highly nontrivial effects of the embodied robot in interaction with the environment. This may imply the presence of higher order statistics as well as strong correlations over time (colored noise) due to the inertia of the robot. Nevertheless we assume for the theory a white Gaussian noise. The justification is taken partly from the results. In fact we will see, that the empirical and theoretical results are in good qualitative agreement. This is sufficient for the present purpose since the theoretical results, besides being helpful for interpreting the empirical findings, are used mainly for the derivation of an on-line learning rule which adapts the parameters of the controller towards the maximum MI regime. Because of the sampling problem this is possible only on the basis of an estimate of the MI with explicit parameter dependence. This (crude) estimate is delivered by our theory.

3.1 The stochastic process in the linear case

Let us first consider again the case of a linear controller, i.e. $g(z) = z$. This is a correct approximation for the case of small z only, but will be seen to reveal already much of the nonlinear case. Using the decoupling of the channels, equation (8) reduces for each channel to the first order autoregressive (AR(1)) process

$$x_{t+1} = cx_t + \xi_{t+1} \tag{11}$$

where $x_t \in R^1$, $|c| < 1$, and we assume that ξ is a white Gaussian noise with mean zero and variance σ^2. As a consequence, the AR process is also Gaussian with variance [2]

$$\sigma_c^2 = \frac{\sigma^2}{1-c^2} \tag{12}$$

and stationary distribution

$$p(x) = \frac{1}{\sqrt{2\pi\sigma_c^2}} \exp\left(-\frac{x^2}{2\sigma_c^2}\right) \tag{13}$$

The conditional probability follows directly from eq. (11)

$$p(x_{t+1}|x_t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_{t+1}-cx_t)^2}{2\sigma^2}\right) \tag{14}$$

and eqs. (13) and (14) yield the joint probability in the stationary state immediately as

$$p(x_{t+1}, x_t) = \frac{\sqrt{(1-c^2)}}{2\sigma^2\pi} \exp\left(-\frac{(x_{t+1}-cx_t)^2 + (1-c^2)\,x_t^2}{2\sigma^2}\right) \tag{15}$$

3.2 Predictive information in the linear case

Our system eq. (11) obeys the Markov property. Hence, as shown in Appendix 8.1, the full predictive information, which relates the future to the past is given by the one-step mutual information

$$I(X_{t+1}; X_t) = \left\langle \log_2 \frac{p(x_{t+1}, x_t)}{p(x_{t+1})\,p(x_t)} \right\rangle = \left\langle \log_2 \frac{p(x_{t+1}|x_t)}{p(x_{t+1})} \right\rangle \tag{16}$$

$$= \int\int p(x,s) \log_2 \frac{p(x|s)}{p(x)} dx\, ds$$

Using Eqs. (13) and (14) we find by elementary means, see Appendix 8.2 or [3]

$$I(X_{t+1}; X_t) = -\frac{1}{2}\log_2\left(1-c^2\right) \tag{17}$$

Interestingly the expression does not depend on the strength of the noise. In order to understand this result we remember that the predictive information, represented by the MI in the AR process, combines the richness of the behavior with the predictability of the future. Both these quantities are driven by the noise, the variance of x_t increasing with increasing noise, see eq. (12), and the predictability deteriorating with it. The two influences balance each other so that the predictive information is depending only on the dynamical quantity c, meaning that it is increasing with increasing c, i.e. with decreasing stability of the dynamics.

3.3 The nonlinear case

Instead of (11) we consider now the full nonlinear equation (6). We cannot assume anymore that the probability densities in sensor space are Gaussians. While it is not possible to write down a closed analytical expression for the mutual information as in the linear case, we can, however, use the transformation properties of the differential entropy to simplify the expression for the mutual information. We start from the representation of the mutual information $I(X_{t+1}; X_t)$ by entropies:

$$I(X_{t+1}; X_t) = H(X_t) + H(X_{t+1}) - H(X_t, X_{t+1}) \; . \tag{18}$$

with $H(X)$ denoting the differential entropy $H(X) = -\int dx p(x) \log_2 p(x)$, see for instance [3]. Now we use the fact that, if $u = f(v)$ is a vector-valued invertible function, one has quite generally

$$H(U) = H(V) + \int dv \; p(v) \; \log_2 |J(v)| \tag{19}$$

with $J(v)$ being the Jacobian of $f(v)$ [12]. By considering the transformation

$$\begin{pmatrix} \xi_{t+1} \\ x_t \end{pmatrix} \mapsto \begin{pmatrix} x_{t+1} \\ x_t \end{pmatrix}$$

provided by $x_{t+1} = \psi(x_t) + \xi_{t+1}$ (6) we get

$$H(X_t, X_{t+1}) = H(X_t, \Xi_{t+1}) \tag{20}$$

because the determinant of the Jacobian is 1 and thus the entropy does not change under this transformation. Assuming that ξ_{t+1} and x_t are statistically independent we get

$$H(X_t, X_{t+1}) = H(X_t) + H(\Xi_{t+1}) \tag{21}$$

so that finally

$$I(X_{t+1}; X_t) = H(X_{t+1}) - H(\Xi_{t+1}) \; . \tag{22}$$

by combining (18) and (21). In this approximation, the mutual information is simply given by the difference between the entropy of the sensory input, which

measures the richness of the dynamics, and the entropy of the noise which measures the unpredictability of the future. The entropy $H(X_{t+1})$ has to be evaluated by numerical simulations, the results are discussed in Sec. 4

In the model dynamics, the MI is given by the entropy of the sensor values minus that of the noise (which is constant), cf. eq. (22), so that the maximum is explained by the entropy of the sensor values alone. Hence, in this approximation the maximum MI behavior of the robot in the physical environment is the one where the robot gets maximum information in its sensor channels. This result is in nice agreement with other approaches seeing the behavior as a means of structuring input information, cf. Lungarella [16].

In order to get more explicit theoretical expressions necessary for the derivation of the learning rule below, we use linearization techniques as known from the theory of dynamical systems. If the noise is sufficiently weak, we may assume to be quite close to a stable fixed point and linearize the dynamical system

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

Writing $\delta x_t = x_t - x^*$ we get approximately

$$\delta x_{t+1} = L\delta x_t + \xi_{t+1} \tag{23}$$

where

$$L = cg'\left(cx^*\right) \tag{24}$$

obviously depends on both $x^*$and c.

The analysis below the bifurcation point (unimodal distribution) is identical to the one given in the linear case, i.e. we obtain

$$I\left(X_{t+1}; X_t\right) = -\frac{1}{2}\log_2\left(1 - L^2\right) \tag{25}$$

where actually $L = c$ since $x^* = 0$ and $g'\left(0\right) = 1$. Above the bifurcation point the distribution is bimodal, approximated by two Gaussians with equal weight. As shown in the Appendix, Sec. 8.3 we obtain

$$I\left(X_{t+1}; X_t\right) = 1 - \frac{1}{2}\log_2\left(1 - L^2\right) \tag{26}$$

The additional bit is due to the knowledge of the branch of the bimodal distribution one is in. The MI increases if approaching the bifurcation point both from below and above, see Sec. 8.3.

When approaching the bifurcation point too closely (depending on the noise) the expressions fail. However one can see by the following heuristic argument that the increase of I given by eq. (25) (with $c = L$) extends smoothly beyond $c = 1$. We write eq. (6) as

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

and note that the positive, even function $\gamma(z) = \tanh(z)/z < 1$ acts as a reduction factor on the value of c which is the smaller the larger x. Approximately we may replace $\gamma(cx)$ with its (time) average so that we get the dynamics equation

$$x_{t+1} = c_{eff} x_t + \xi_{t+1} \tag{27}$$

where $c_{eff} = \overline{\gamma(x)}c$. An explicit expression for c_{eff} can be obtained in the sense of a self-consistent mean field approach, by using the distribution $p(x)$, see eq. (13), with c replaced by c_{eff}. However we do not want to go into these details here since the main point is that $c_{eff} < c$ so that the linear dynamics, eq. (27), can be used as a crude approximation for the full nonlinear dynamics around $c = 1$. Then, using in eq. (17) c_{eff} instead of c immediately yields an expression

$$I(X_{t+1}; X_t) \approx -\frac{1}{2}\log_2\left(1 - c_{eff}^2\right) \tag{28}$$

for the MI valid approximately even for $c \gtrsim 1$. Speaking in terms of distributions, the argument relies on the fact that, with noise, the bimodality is felt only somewhat above the actual bifurcation point. Before that the distribution can be crudely approximated by a Gaussian with a width defined by c_{eff} instead of c in eq. (13).

4 An embodied robot experiment

It is one of our aims to use the information theoretic measures in realistic robotic applications putting particular emphasis on the role of the embodiment. This means that we want to discuss physical robots, be it in reality or in simulations, where the embodiment manifests itself by physical effects like inertia, slip and friction effects, uncertain sensor and actuator functioning. On the other hand we have chosen our experiments such that our theoretical expressions are still applicable.

4.1 Experiments

In the experiments, the robot is moving in an arena surrounded by walls and with several obstacles in it so that, without any proximity sensors, the robot will often collide with either the walls or the obstacles. As discussed in Sec. 2.3, this behavior is largely depending on the value c of the controller (which determines the feed-back strength of the sensorimotor loop).

4.2 The mutual information

A central aim of the present paper was to find the mutual information as a function of the behavior parameter c in the embodied robot experiment. In the experiments we evaluated the MI of each of the sensor channels independently. For this purpose we started the robot at a random position and let it run for a long time, mostly for up to one million steps with a fixed value of c. We

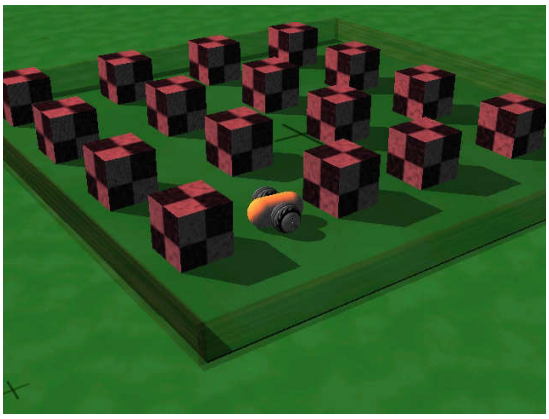

Figure 2: The arena for our two-wheel robot in the starting situation. The robot is "blind" and feels the environment only by the reactions of its wheel counters on collissions with the obstacles. The behavior with $c = 1.07$ (maximum mutual information) is singled out with the robot covering large distances while keeping maximum contact with the environment, see the videos.

discretized the interval of possible sensor values into 30 bins which proved sufficiently accurate by comparison with cases of 10, 20, and 50 bins. Probabilities $p(x)$ or $p(x_{t+1}, x_t)$ were interpreted as relative frequencies of the sensor values in each bin or pair of bins, respectively, sampled over time t. The integral in eq. (16) was replaced by the Riemannian sum. The procedure was repeated for every of the c values in the graphics, see Fig. 3.

In practice, the MI was evaluated by an update rule in order to control the convergence progress. Convergence of the MI was reached in typical runs after about 10^5 to 10^6 steps. The convergence largely depends on the value of c. In particular for $c \gg 1$ the robot may change between FPs after a very long time only and this means that the additional bit of the bimodal regime is not seen in the experiments with a finite number of steps.

4.3 Results

The most important experimental result is the relatively sharp maximum of the empirical MI at $c_{MI} \approx 1.07$, see Fig. 3. In order to relate the MI, which is taken in sensor space, to the behavior of the robot in physical space, we partitioned the maze into 10×10 cells and recorded the probability of visiting each cell. The Shannon entropy of this spatial distribution is a convenient measure of the exploration of the maze by the robot. From Fig. 4 which is depicting the trajectories of the robot we see that at the maximum of the MI the robot visits much more different sites in the maze than away from it.

The result indicates a close link between the mutual information in sensor

space and the behavior of the robot in physical space, i.e. in the specific environment. In order to discuss this point let us start with considering the behavior of the robot in terms of the dynamical system analysis given in Secs. 2.3 and 2.4. Obviously, in the experiment, the robot behaves most effectively in the region around the effective bifurcation point (critical region). This is not surprising given that the robot is blind and feels the environment only by the reactions of its wheel counters on collisions with the obstacles. In fact, in this region the robot deploys already its modes (rotation or straight) which are however both softened and occasionally swapped by the noise. Moreover, collisions with obstacles are soft and lead to immediate switching in the modes so that in the maze environment the robot seems to develop a kind of controlled bouncing strategy.

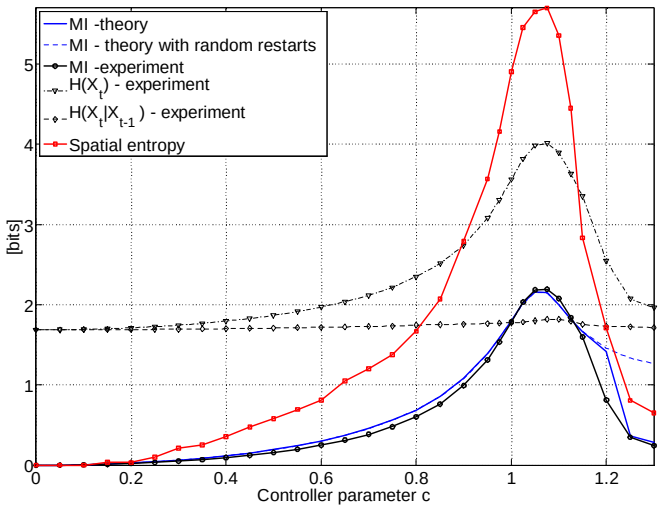

Figure 3: Mutual information in sensor channels and spatial exploravity in an embodied robot experiment: The mutual information between successive time steps as a function of the parameter c shows a clear maximum at $c = 1.07$. The position of the maximum agrees nearly exactly with the maximum of the spatial entropy, measuring the distribution of the sites visited by the robot. This indicates, that the maximum of the MI corresponds to the best exploration behavior in the maze. The experimental MI is compared with the MI as obtaind from the model dynamics by numerical simulation. All runs are over 600000 time steps. The drop off of the theoretical curve results from the fact that the bimodality is not felt due to the finite sampling time. The behavior of the entropy of the sensor values $H(X)$ and the conditional entropy $H(X_{t+1}|X_t)$ are also presented.

This is a mechanistic explanation based on the specific attractor landscape of the sensorimotor dynamics. What is the relation to the MI? Coarsely speaking the predictive information (the MI in our case) is large if the behavior is rich (so that much information from the past is necessary in order to describe the future) but still as predictable as possible. The soft mode scenario at the effective BP seems to fit well into this picture since behavior in stable modes is well predictable but not rich in dynamics whereas a behavior fluctuating around and jumping between fixed points is much more rich while retaining still some amount of predictability. Thus, in the specific setting considered, the phenomenon of an effective bifurcation point may be considered as the link between the behavior in physical and the complexity measure in sensor space.

In Fig. 3 we also present the MI as obtained from the model dynamics. In the interpretation of the result we have to consider that in the embodied robot experiments we used a certain amount of sensor noise (white Gaussian noise with $\sigma = 0.06$) which is essential for the behavior of the robot under our closed loop control paradigm. The nice agreement with the experiment seems to indicate that the model with the white Gaussian noise accounts already for most of the empirical behavior of the robot in the maze. The drop off of the theoretical curve at $c \approx 1.2$ is due to the fact that, given the finite sampling time, the system does not switch between the modes any more. In order to test this hypothesis we used random restarts of the system repeatedly. This introduces the additional bit of information, see eq. (26). The faster decay of the empirical MI probably is due to the fact that the robot has a rather large mass which stabilizes any rotational mode against being switched by the noise. Thus, once the robot has entered a rotational mode (by a collision with an obstacle) it will stay in it for the rest of the sampling time. The dependence of the MI on the sampling time in the bimodal region might seem dissatisfying. However, the difference is just the additional bit of information which is independent on the parameter c. Hence, for the derivation of the learning rule this effect is of no relevance, see Sec. 8.4 below.

The results obtained may form the basis for future generalizations of the present findings to more complex systems. We have seen that there is a direct relation of the MI in each sensor channel with the behavior of the robot in the world although the sensor values (wheel velocities) are related only in a very indirect way to the navigation behavior (bouncing strategy) in the maze. However, we studied only the one-step predictive information. The generalization therefore has to go into the direction of (i) taking a larger time horizon for both past and future since the physical is non-Markovian, (ii) include proximity sensors so that the obstacles can be seen beforehand, and (iii) using a more complex controller including internal states. It is our strong believe, that in this setting the maximum predictive information will correspond to a smooth but explorative navigation behavior in the maze with strategies for circumventing the obstacles. In fact, it is only in this way that the predictability can be made large. In future work we will also observe further characteristics of the robot behavior like the distances covered versus the damage probability (overload of the motors, e.g.) and compare those with the predictive information.

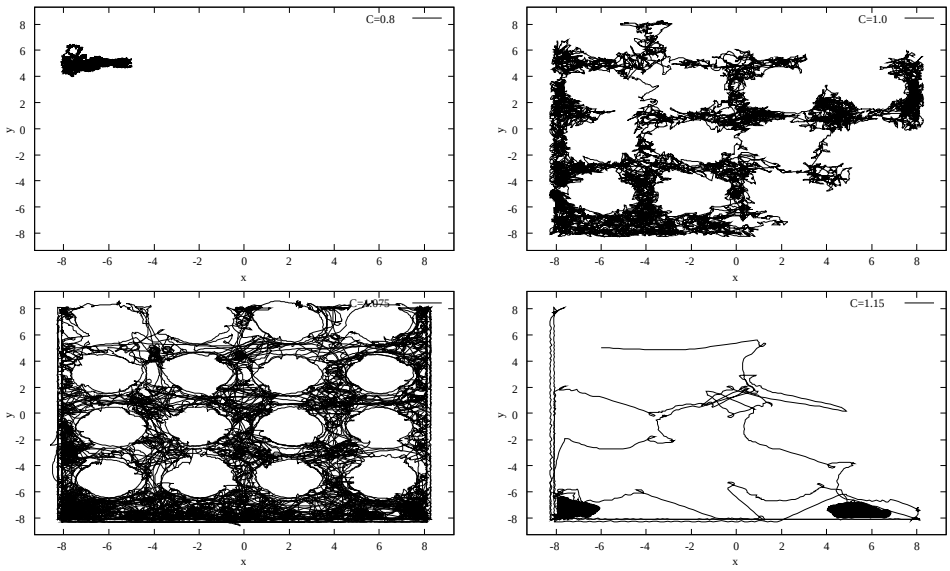

Figure 4: Trajectories of the robot in the maze for different values of the behavior parameter c. Runs are over 600.000 time steps each. With $c = 0.8$ the robot is seen to essentially fluctuate on site whereas for $c = 1.15$ the robot is caught two times in a dead lock. The runs for $c = 1.0$ and $c = 1.075$ show the sensitive dependence of the behavior on the controller parameter c.

5 Learning rules based on information measures

By our experiments we may conclude that the maximum of the mutual information defines a working regime where the robot is both explorative and sensitive to the environment. This can be used for the construction of a learning rule for the behavioral development of the robot, i.e. we define an update rule for the parameter c as

$$\Delta c = \varepsilon_0 \frac{\partial I\left(X_{t+1}; X_t\right)}{\partial c} \tag{29}$$

We have seen above that the sampling times for the MI are very long so that an on-line learning seems difficult to be realized. On the other hand, when using the theoretical expressions given by eqs. (25) and (26), we obtain the explicit update rule as

$$\Delta c = \varepsilon - 2\varepsilon c x_t y_t \tag{30}$$

see Sec. 8.4 (Appendix). This learning rule has some nice features. In particular it is extremely simple in structure (ε may be kept constant since it does influence only the learning speed, see Appendix) and moreover, besides the constant driving term it has an anti-Hebbian structure. This is interesting in the context of neural realizations of the controller.

However as explained in the Appendix, the learning rule involves approximations valid only sufficiently far away from the bifurcation point. In order to find the learning behavior around the bifurcation point we discuss at first the stationary point of the rule (30). Learning stops if (assuming the state is at the fixed point) $1 = 2cxy = 2cx^2$ according to the sensorimotor dynamics. On the other hand, the FP condition is $x = \tanh(cx)$. The numerical solution of these two equations yields $c = 1.191$ which is in the region of the effective bifurcation point (which is dependent on the noise, see above) found in the experiments. As a consequence we argue to use the learning rule for all values of c since it drives c into the vicinity of the maximum mutual information. This might be appropriate for some moderate noise but is not correct if the noise is small. The derivation of a more general rule which drives c to the effective bifurcation point must be left to a later paper.

The learning rule (30) (apart from the effective learning rate) has also been derived by minimizing the so called time loop error in the context of homeokinesis and was discussed in detail elsewhere, cf. [6]. This rule and its multidimensional generalizations was extensively used and observed to drive various types of robotic systems towards interesting working regimes under many different circumstances, cf. [9], [8]. It is interesting to see that the present approach also leads to this rule (albeit with a different prefactor) relating the concept of the time loop error with complexity measures like the predictive information.

6 Concluding remarks

The aim of the present paper has been twofold. On the one hand we have investigated, in an embodied robot experiment, the role of predictive information as a tool for quantifying the behavior of an autonomous robot. Predictive information has been shown to reduce to the mutual information (MI) between time points in the case of Markovian systems so that the MI may be used as a first step towards the full predictive information. The MI of the sensor values over time has been determined empirically in embodied robot experiments. The main result is that the MI shows a clear maximum in the working regime where, from the point of view of an external observer, the robot may be said to develop a kind of effective strategy for navigating the environment. The latter result is not trivial since, without any proximity sensors, the robot feels the environment only via its wheel counters in a very implicit way. It remains to be seen in future experiments whether this link between the information measure in sensor channels and the strategy of the robot is of a more fundamental nature, as claimed for instance in [17].

On the other hand we discussed the complexity measure as the basis for the self-organization of robot behavior by using the measure as an objective function for a gradient following learning rule. The main obstacle in such an attempt are the large sampling times until convergence is reached. In our case we needed 10^5 to 10^6 time steps. Since behavior changes by the learning process, this is prohibitive for any on-line learning scenario. However, our theoretical consider-

ations have shown that, at least in the present case, the structure of the learning rule can be obtained by using a simple model of the sensorimotor loop (which can be learned on-line by any of the known supervised learning procedures) with the mutual information featuring only as some parameter in this rule (here in the effective learning rate). Therefore it seems appropriate to use the crude estimate of the current value of the mutual information given by the theory in order to move, in an on-line learning scenario, towards the maximum of the MI. Once in that region, behavior is changing only slowly so that sampling of the MI will converge partially and may be used for improvements over the estimate.

The generalization of our results to more complicated cases is based on the close relationship of the information theoretic measure to the so called time loop error and the principle of homeokinesis, cf. [7], [10], [5], 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, cf. [9], [8], [6] 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.

7 Acknowledgements

The authors thank Michael Herrmann and Susanne Still for helpful discussions. Nihat Ay thanks the Santa Fe Institute for supporting him as an external professor.

8 Appendix

8.1 Predictive information for Markovian systems

Consider a Markov transition kernel $p(x'|x)$ and a corresponding stationary probability distribution $p(x)$, that is $\sum_x p(x)p(x'|x) = p(x')$. This defines a stationary Markov process X_t, $t \in \mathbb{Z}$, with distribution

$$\begin{aligned} &\Pr\{X_r = x_r,\, X_{r+1} = x_{r+1}, \ldots,\, X_s = x_s\} \\ &= p(x_r)\, p(x_{r+1}|x_r) \cdots p(x_s|x_{s-1}), \qquad r < s. \end{aligned} \tag{31}$$

We use the abbreviation $X_{[r,s]}$ for the random vector $X_r, X_{r+1}, \ldots, X_s$. The conditional independence structure of the distribution (31) implies that for times $r \le r' < s \le s' < t \le t'$ the conditional mutual information $I(X_{[r,r']}; X_{[t',t]}|X_{[s,s']})$ vanishes. With the chain rule for mutual information, this finally implies for

$m \geq 1$ and $n \geq 2$

$$\begin{aligned} & I(X_{[-m,0]}; X_{[1,n]}) \\ & = I(X_1; X_0) + I(X_1; X_{[-m,-1]}|X_0) + I(X_{[-m,0]}; X_{[2,n]}|X_1) \\ & = I(X_1; X_0). \end{aligned}$$

so that the predictive information as mutual information between the past and the future has a finite value which coincides with the one-step mutual information $I(X_{t+1}; X_t)$. This the quantity that we use in this paper.

8.2 Evaluation of the MI in the linear case

We derive here the MI directly on the basis of the distributions in order to get some additional insight into the process. We use

$$I\left(X_{t+1}; X_t\right) = H\left(X_{t+1}\right) + H\left(X_t\right) - H\left(X_{t+1}, X_t\right)$$

where $H\left(X\right)$ is the entropy of the stationary process X. With the Gaussian distribution of both X_{t+1} and X_t we find immediately

$$H\left(X_{t+1}\right) = H\left(X_t\right) = \frac{1}{2}\log_2\left(2\pi\frac{\sigma^2}{1-c^2}\right) + \frac{1}{2} \tag{32}$$

In order to evaluate

$$H\left(X, S\right) = -\int\int dx\ ds\ p\left(x, s\right)\log_2 p\left(x, s\right)$$

we use the joint distribution given by eq. (15), find

$$\begin{aligned} & \frac{1}{2}\int\int dx\ ds\ p\left(x, s\right)\left(\frac{\left(x - cs\right)^2 + \left(1 - c^2\right)s^2}{\sigma^2}\right) \\ & = \frac{1}{2}\int ds \frac{1}{\sqrt{2\sigma^2\pi}}\sqrt{\left(1 - c^2\right)}\exp\left(-\frac{\left(1 - c^2\right)s^2}{2\sigma^2}\right)\ \left(1 + \frac{\left(1 - c^2\right)s^2}{\sigma^2}\right) \\ & = 1 \end{aligned}$$

and get finally

$$H\left(X, S\right) = \log_2\left(2\pi\sigma^2\right) - \frac{1}{2}\log_2\left(1 - c^2\right) + 1$$

The MI is therefore

$$\begin{aligned} I\left(X_{t+1}; X_t\right) &= 2H\left(X_t\right) - H\left(X_{t+1}; X_t\right) \\ &= -\frac{1}{2}\log_2\left(1 - c^2\right) \end{aligned}$$

which is the result used in the main text. Note that this result is obtained also more elegantly from the general expression given by eq. (22) using eq. (32). In the linearized but still unimodal case we have to replace c with L.

8.3 MI in the bimodal regime

Let us assume that we are sufficiently far from the bifurcation point so that the distribution can be approximated as

$$p(x) = \frac{1}{2}\left(p_+(x) + p_-(x)\right)$$

where

$$p_\pm(x) = \sqrt{\frac{1-L^2}{2\pi\sigma^2}}\exp\left(-\left(\frac{\left(x \pm |x^*|\right)^2}{2\sigma^2}\left(1-L^2\right)\right)\right)$$

are two normalized Gaussians with negligible overlap. Using eq. (22) we have to calculate

$$\begin{aligned} H(X) &= -\int_{-\infty}^{\infty} p(x)\log_2 p(x) \\ &= -2\int_{-\infty}^{\infty}\frac{p_+(x)}{2}\log_2\left(\frac{p_+(x)}{2}\right) = 1 + \int_{-\infty}^{\infty} p_+(x)\log_2 p_+(x) \\ &= 1 + \frac{1}{2} + \frac{1}{2}\log_2\left(2\pi\frac{\sigma^2}{1-L^2}\right) \end{aligned}$$

Altogether we have in the bimodal case approximately

$$I(X_{t+1}; X_t) = 1 - \frac{1}{2}\log_2\left(1-L^2\right)$$

so that, as compared to the unimodal case, we have an additional bit of information which is clear since we now have the freedom to choose between two states.

The relations reveal that the MI increases when approaching the bifurcation point both from below and above. This is obvious for the unimodal region. In the bimodal region we can use approximate expressions valid on the one hand if $c = 1 + \delta$ with $0 < \delta \ll 1$. Using eq. (10) and $g'(z) \approx 1 - 3\delta + O\left(\delta^2\right)$ we get $L = 1 - 2\delta + O\left(\delta^2\right)$ and

$$I = -\frac{\ln\delta}{2\ln 2} + O(\delta) \tag{33}$$

which decreases logarithmically for sufficiently small δ. On the other hand, with sufficiently large c we may write approximately $g'(z) = 4e^{-2|z|}$ and $z^* \approx c$ so that

$$L = 4ce^{-2c}$$

and

$$I(X_{t+1}; X_t) = 1 - \frac{1}{2}\log_2\left(1-L^2\right) \approx 1 + \frac{1}{2\ln 2}L^2 \approx 1 + \frac{8}{\ln 2}c^2e^{-4c}$$

Obviously, the MI decreases exponentially with increasing c.

8.4 Derivation of the learning rule

Let us write the two expressions for the MI below and above the BP as

$$I\left(X_{t+1};X_t\right)=\theta+\widetilde{I}\left(X_{t+1};X_t\right)=\theta-\frac{1}{2}\log_2\left(1-L^2\right)$$

where $\theta=0$ below and $\theta=1$ above the BP. The derivative in eq. (29) is taken by the chain rule, i.e. consider first

$$\frac{\partial I}{\partial L}=\frac{\partial\widetilde{I}}{\partial L}=\frac{L}{\left(1-L^2\right)\ln 2}=\frac{1}{\ln 2}e^{(2\ln 2)\tilde{I}}L$$

Using eq. (24) we have, neglecting the dependence of the fixed point on c (see below),

$$\frac{\partial L}{\partial c}=\frac{\partial}{\partial c}\left(cg'\left(z\right)\right)=g'\left(z\right)+cxg''\left(z\right)$$

With $g\left(z\right)=\tanh z$ we get in particular $g'\left(z\right)=1-g^2\left(z\right)$ and $g''\left(z\right)=-2g\left(z\right)g'\left(z\right)$ so that

$$\frac{\partial L}{\partial c}=\left(1-2zg\left(z\right)\right)g'\left(z\right)$$

and

$$\frac{\partial I}{\partial c}=\frac{1}{\ln 2}e^{(2\ln 2)\tilde{I}}L\frac{\partial L}{\partial c}=\frac{Lg'\left(z\right)}{\ln 2}e^{(2\ln 2)\tilde{I}}\left(1-2zg\left(z\right)\right)$$

which has been written in such a way that the MI is figuring explicitly. Introducing (absorbing constants into ε_0)

$$\varepsilon=\varepsilon_0 e^{(2\ln 2)\tilde{I}\left(X_{t+1};X_t\right)}\left(1-g^2\left(z\right)\right)^2 c \tag{34}$$

eq. (29) leads to the learning rule valid in the region where the linearization is valid

$$\Delta c=\varepsilon-2\varepsilon cxg\left(cx\right) \tag{35}$$

Δc denoting the increment of c in the learning step and $\varepsilon>0$ is an effective learning rate which may be taken constant in practical applications since it influences only the magnitude but not the direction of the gradient.

So far, x is the fixed point around which the linearization was taken. However if sufficiently far away from the bifurcation point, x stays close to its fixed point value so that we may replace x with its current value x_t and in the same sense $g\left(cx\right)$ with $y_t=g\left(cx_t\right)$. Eq. (35) is remarkable because of its simplicity. However, it is so far valid only far away from the BP. In order to derive a learning rule for the full range of c we have to consider several points. On the one hand, eq. (35) has been obtained by taking the derivative of I only with respect to the explicit c dependence. Including the dependence of x on c the gradient descent is seen to drive c to the BP at $c=1$. However, this is valid only in the limit of vanishing noise where the linearization is valid for all values of c. With finite noise the rule is to converge towards the effective bifurcation point and we hope to present a correction term to the above learning rule, eq. (35), in a later paper. In the present paper we simply use eq. (35) for the full range of c, see the main text.

References

[1] W. Bialek, I. Nemenman, and N. Tishby. Predictability, complexity and learning. *Neural Computation*, 13:2409, 2001.

[2] G. Box, G. M. Jenkins, and G. C. Reinsel. *Time Series Analysis: Forecasting and Control*. Prentice Hall, 1994.

[3] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley series in telecommunications. Wiley, New York, 1991.

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

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

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

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

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

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

[10] R. Der, U. Steinmetz, and F. Pasemann. 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, 1999. IOS Press.

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

[12] G. Jumary. *Relative Information*, volume 47 of *Springer Series in Synergetics*. Springer-Verlag, Berlin Heidelberg, 1990.

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

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

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

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

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

[18] G. Martius and R. Der. Lpzrobots – simulation tool for autonomous robots. http://robot.informatik.uni-leipzig.de/, 2007.

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

[20] H. Risken. *The Fokker-Planck equation*. Springer, 1989.

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

[22] R. Smith. Open dynamics engine. http://ode.org/, 2005.

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

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

[25] J. Weng, J. McClelland, A. Pentland, O. Sporns, I. Stockman, M. Sur, and E. Thelen. Autonomous mental development by robots and animals. *Science*, 291:599 – 600, 2001.