

Predictive information and explorative behavior of autonomous robots

N. Ay^{1,3}, N. Bertschinger¹, R. Der^{1,a}, F. Güttler², and E. Olbrich¹

¹ Institute for Information Theory and Robotics, Vienna University of Technology, 1040 Vienna, Austria
² Institute for Information Theory and Robotics, Vienna University of Technology, 1040 Vienna, Austria
³ Institute for Information Theory and Robotics, Vienna University of Technology, 1040 Vienna, Austria

Abstract.

We study the predictive information of a process, which 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.

PACS.

02.30.Gt, 02.30.Lv, 02.30.Mv, 02.30.Ng, 02.30.Oz, 02.30.Px, 02.30.Qz, 02.30.Rz, 02.30.Sg, 02.30.Th, 02.30.Uh, 02.30.Vj, 02.30.Wk, 02.30.Xy, 02.30.Yz, 02.30.Za, 02.30.Zb, 02.30.Zc, 02.30.Zd, 02.30.Ze, 02.30.Zf, 02.30.Zg, 02.30.Zh, 02.30.Zi, 02.30.Zj, 02.30.Zk, 02.30.Zl, 02.30.Zm, 02.30.Zn, 02.30.Zo, 02.30.Zp, 02.30.Zq, 02.30.Zr, 02.30.Zs, 02.30.Zt, 02.30.Zu, 02.30.Zv, 02.30.Zw, 02.30.Zx, 02.30.Zy, 02.30.Zz

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

^a rder@informatik.uni-leipzig.de

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 [15,25]. 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 Section 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. Section 3 introduces the information theoretic measures and gives a theoretical expression for the case at hand. After this we present in Section 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 Section 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 Section 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

$$y = K(x) \quad (1)$$

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

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

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

$$C = \begin{pmatrix} c & 0 \\ 0 & c \end{pmatrix} \quad (3)$$

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 $x_t \in \mathbf{R}^n$, $t = 1, 2, \dots$. 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

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

where in general $F : \mathbf{R}^n \times \mathbf{R}^m \rightarrow \mathbf{R}^n$ is a function mapping old sensor and motor values to the new sensor values with $\xi \in \mathbf{R}^n$ 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 equation (4) boils down to

$$x_{t+1} = y_t + \xi_{t+1} \quad (5)$$

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, 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.

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

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

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

$$\psi(x) = G(Cx) \quad (7)$$

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

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

Although the robot may behave in a very intricate way (see below), equation (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.

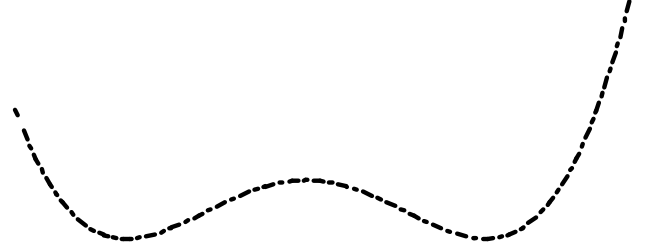
2.3 Properties of the single channel dynamics

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 equation (8). We consider each loop independently (uncorrelated noise) with fixed point equation

$$x = \tanh(cx). \quad (9)$$

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

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



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

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

3.2 Predictive information in the linear case

Our system equation (11) obeys the Markov property. Hence, as shown in Appendix A.1, the full predictive information, which relates the future to the past is given by the one-step mutual information

$$\begin{aligned} I(X_{t+1}; X_t) &= \left\langle \log_2 \frac{p(x_{t+1}, x_t)}{p(x_{t+1})p(x_t)} \right\rangle \\ &= \left\langle \log_2 \frac{p(x_{t+1}|x_t)}{p(x_{t+1})} \right\rangle \\ &= \int \int p(x, s) \log_2 \frac{p(x|s)}{p(x)} dx ds. \quad (16) \end{aligned}$$

Using equations (13) and (14) we find by elementary

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

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

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

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

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 Appendix A.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 equation (25) (with $c = L$) extends smoothly beyond $c = 1$. We write equation (6) as

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

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

$$x_{t+1} = c_{eff}x_t + \xi_{t+1} \quad (27)$$

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 equation (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, equation (27), can be used as a crude approximation for the full nonlinear dynamics around $c = 1$. Then, using in equation (17) c_{eff} instead of c immediately yields an expression

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

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 equation (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

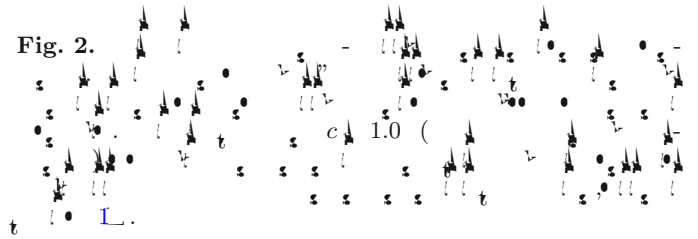


Fig. 2.

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 Section 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 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 equation (16) was replaced by the Riemannian sum. The procedure was repeated for every of the c values in the graphics, see Figure 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.

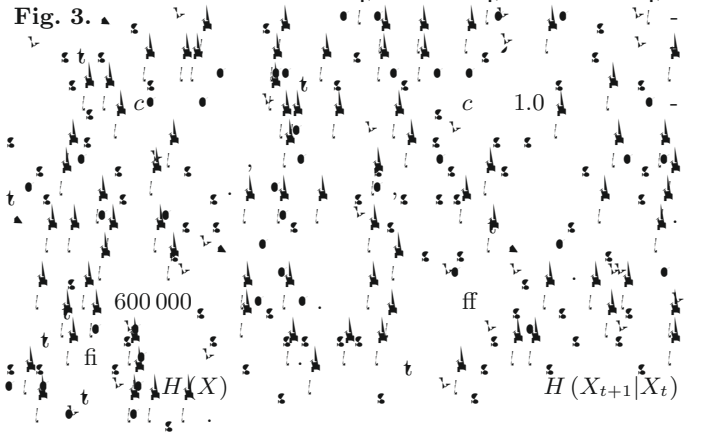
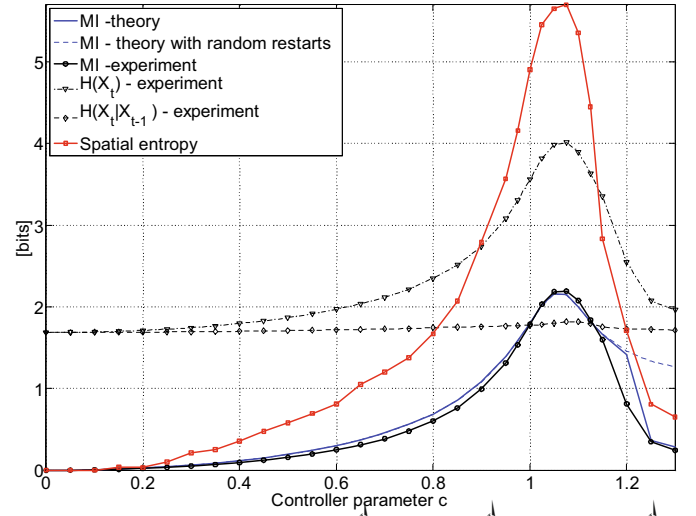
4.3 Results

The most important experimental result is the relatively sharp maximum of the empirical MI at $c_{MI} \approx 1.07$, see Figure 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 Figure 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 Sections 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.

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 Figure 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 equation (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 Appendix A.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.

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

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

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 equations (25) and (26), we obtain the explicit update rule as

$$\Delta c = \varepsilon - 2\varepsilon c x_t y_t \quad (30)$$

see Appendix A.4. 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 A) 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

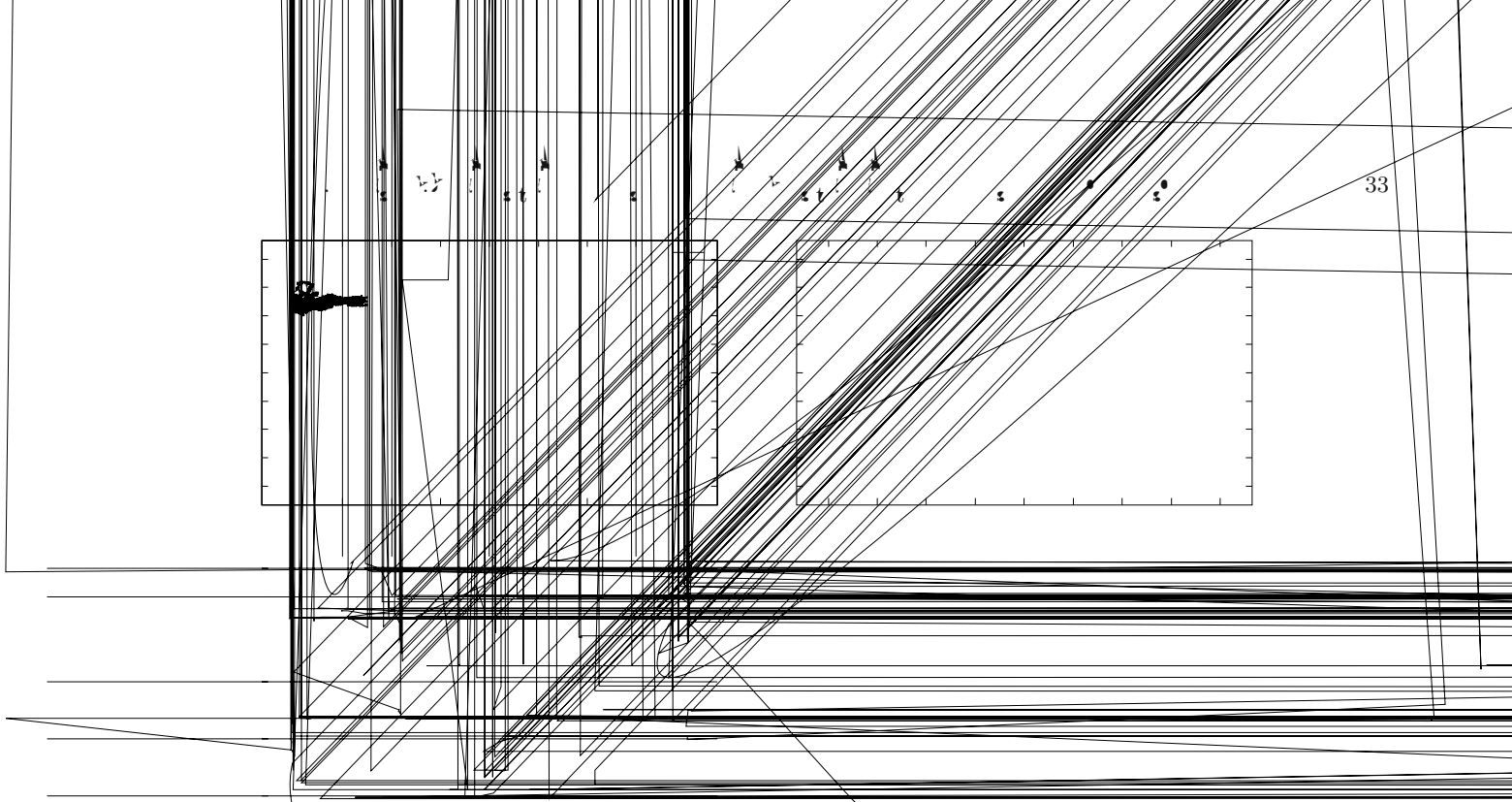
rule for all values of c since it drives c into the vicinity of the maximum mutual information. This might be appropriate for some moderate noise but is not correct if the noise is small. The derivation of a more general rule which drives c to the effective bifurcation point must be left to a later paper.

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 multi-dimensional generalizations was extensively used and observed to drive various types of robotic systems towards interesting working regimes under many different circumstances, cf. [8,9]. 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 considerations 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



In order to evaluate

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

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

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

and get finally

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

The MI is therefore

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

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

A.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

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

where

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

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

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

Altogether we have in the bimodal case approximately

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

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 equation (10) and $g(z) \approx 1 - 3\delta + O(\delta^2)$ we get $L = 1 - 2\delta + O(\delta^2)$ and

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

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

$$\begin{aligned} I(X_{t+1}; X_t) &= 1 - \frac{1}{2} \log_2(1 - L^2) \\ &\approx 1 + \frac{1}{2 \ln 2} L^2 \approx 1 + \frac{8}{\ln 2} c^2 e^{-4c}. \end{aligned}$$

Obviously, the MI decreases exponentially with increasing c .

A.4 Derivation of the learning rule

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

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

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

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

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

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

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

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

and

$$\begin{aligned}\frac{\partial I}{\partial c} &= \frac{1}{\ln 2} e^{(2 \ln 2) \tilde{I}} L \frac{\partial L}{\partial c} \\ &= \frac{L g(z)}{\ln 2} e^{(2 \ln 2) \tilde{I}} (1 - 2zg(z))\end{aligned}$$

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

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

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

$$\Delta c = \varepsilon - 2\varepsilon c g(cx) \quad (35)$$

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 $g(cx)$ with $y_t = g(cx_t)$. Equation (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, equation (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, equation (35), in a later paper. In the present paper we simply use equation (35) for the full range of c , see the main text.

References

1. V. L. Ginzburg, *Phys. Rev. Lett.* **13**, 2407 (2001)
2. J. H. B. J. van der Kamp, *Time Series Analysis: Forecasting and Control* (1994)
3. A. R. B. J. van der Kamp, *Elements of Information Theory* (1991)
4. A. R. B. J. van der Kamp, *Phys. Rev. Lett.* **63**, 105 (1999)
5. A. R. B. J. van der Kamp, *Phys. Rev. Lett.* **120**, 177 (2001)
6. A. R. B. J. van der Kamp, *Proc. Workshop Biologically Inspired Robotics. The Legacy of Grey Walter 14-16 August 2002, Bristol Labs* (2002)
7. A. R. B. J. van der Kamp, *From motor babbling to purposive actions: Emerging self-exploration in a dynamical systems approach to early robot development, From Animals to Animats*, **4075** (2006), **406**
8. A. R. B. J. van der Kamp, *Let it roll – emerging sensorimotor coordination in a spherical robot, Artificial Life X*, (2006), **172**
9. A. R. B. J. van der Kamp, *Homeokinesis – a new principle to back up evolution with learning, Computational Intelligence for Modelling, Control and Automation*, **55** (1999), **43**
10. A. R. B. J. van der Kamp, *Relative Information*, **25(9)** **70** (1996)
11. A. R. B. J. van der Kamp, *Relative Information*, **25(9)** **70** (1996)
12. A. R. B. J. van der Kamp, *Empowerment: A universal agent-centric measure of control, Proc. CEC*, **2005**
13. A. R. B. J. van der Kamp, *Empowerment: A universal agent-centric measure of control, Proc. CEC*, **2005**
14. A. R. B. J. van der Kamp, *Empowerment: A universal agent-centric measure of control, Proc. CEC*, **2005**
15. A. R. B. J. van der Kamp, *Empowerment: A universal agent-centric measure of control, Proc. CEC*, **2005**
16. A. R. B. J. van der Kamp, *Empowerment: A universal agent-centric measure of control, Proc. CEC*, **2005**
17. A. R. B. J. van der Kamp, *Empowerment: A universal agent-centric measure of control, Proc. CEC*, **2005**
18. A. R. B. J. van der Kamp, *Empowerment: A universal agent-centric measure of control, Proc. CEC*, **2005**
19. A. R. B. J. van der Kamp, *Empowerment: A universal agent-centric measure of control, Proc. CEC*, **2005**
20. A. R. B. J. van der Kamp, *Empowerment: A universal agent-centric measure of control, Proc. CEC*, **2005**
21. A. R. B. J. van der Kamp, *Empowerment: A universal agent-centric measure of control, Proc. CEC*, **2005**
22. A. R. B. J. van der Kamp, *Empowerment: A universal agent-centric measure of control, Proc. CEC*, **2005**
23. A. R. B. J. van der Kamp, *Empowerment: A universal agent-centric measure of control, Proc. CEC*, **2005**
24. A. R. B. J. van der Kamp, *Empowerment: A universal agent-centric measure of control, Proc. CEC*, **2005**
25. A. R. B. J. van der Kamp, *Empowerment: A universal agent-centric measure of control, Proc. CEC*, **2005**