Information and closure in systems theory

Nils Bertschinger, Eckehard Olbrich, Nihat Ay, Jürgen Jost

Max Planck Institute for Mathematics in the Sciences
Inselstr. 22
D 04103 Leipzig, Germany

Abstract

The notion of closure plays a prominent role in systems theory where it is used to identify or define the system in distinction from its environment and to explain the autonomy of the system. Here, we present a quantitative measure, as opposed to the already existing qualitative notions, of closure.
We shall elaborate upon the observation that cognitive systems can achieve *informational closure* by modeling their environment. Formally, then, a system is informationally closed if (almost) no information flows into it from the environment.
A system that is independent from its environment trivially achieves informational closure. Simulations of coupled hidden Markov models demonstrate that informational closure can also be realized non-trivially by modeling or controlling the environment. Our analysis of systems that actively influence their environment to achieve closure then reveals interesting connections to the related notion of autonomy.
This discussion will then call into question the system-environment distinction that seems so innocent to begin with. It turns out that the notion of autonomy depends crucially on whether, not just the state observables, but also the dynamical processes are attributed to either the system or the environment. In that manner, our conceptualization of informational closure also sheds light on other, more ambitious notions of closure, e.g. organizational closure, semantic closure, closure to efficient cause or operational closure, intended as a fundamental (defining) concept of life itself.

1 Introduction

Our theoretical interest concerns the type of system that is a unity for and by itself and not only for an external observer distinguishing some entity from the rest of the world. This requires a system that can be described as a whole without reference to its environment. In systems theory, this property is usually referred to as closure.
In this regard, one encounters several notions of closure in the literature: autopoiesis as organizational closure (Maturana and Varela [1]), closure to efficient cause (Robert Rosen [2]), semantic closure (Howard Pattee [3]), or operational closure (Niklas Luhmann [4]). These concepts of closure play an important role in the architecture of systems theory, because they are used to

1. define the system (in distinction to its environment) and to
2. explain the autonomy of the system.

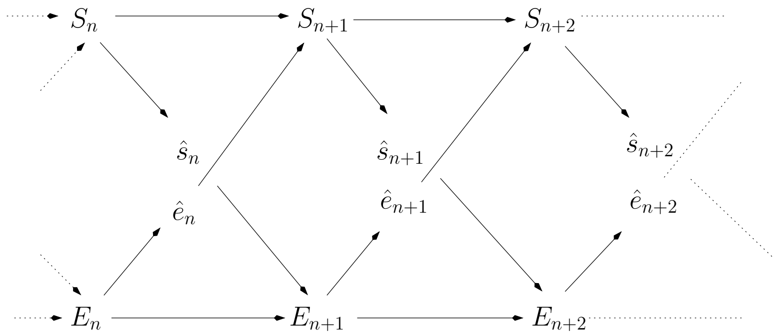


Figure 1: *The system S and the environment E interact through the channels $\hat{s}$ and $\hat{e}$. The figure shows the temporal dependencies of this interaction.*

An autopoietic system, for example, is said to be organizationally closed, because the processes that constitute the organization of the system are themselves maintaining/reproducing the conditions for their own existence. Thereby, the system also defines its own boundary that separates itself from its environment. This self-referential distinction from its environment therefore gives rise to the specific autonomy of such a system. Consequently, in systems theory, closure properties and autonomy are considered to be closely related concepts which are both at the heart of defining the system itself.

These aforementioned attempts, however, either remain vague (Luhmann, Maturana), lead to concerns about their formal consistency ([5]), or are too abstract to work with ([2]). Therefore, the formalization of closure remains a central issue in the formalization of systems theory.

As a starting point, we use a closure type phenomenon that can be observed in cognitive systems. Such systems are assumed to be capable of reducing the information flow from the environment into the system by modeling the environment. We shall call this “informational closure”. We think that this concept can also contribute an abstract notion of “modeling” that does not depend upon the identification of certain structures in the system as explicit models or representations.

Informational closure is, arguably, not enough to completely capture all properties of closure, in particular the ones required to derive the existence of the system itself. It is, however, more amenable to quantification since it should not be considered to be an all-or-nothing phenomenon. A system can only be informationally closed with respect to those features of the environment that can be modeled. Unpredictable events that nevertheless do affect the system clearly have to give rise to an information flow into the system.

In this paper, after introducing the basic setup and the notation, we propose a measure for informational closure. In section III we apply our measure to simple hidden Markov models. There, we focus on non-trivial closure, i.e. systems that feature low information flow from the environment, but nevertheless contain mutual information about their environment. We also distinguish the case of passive systems, which cannot influence the environment, from the case of active ones that are able to change the environment. This not only allows us to propose a measure of the influence a system can exert upon its environment, called self-efficacy, but also leads to a discussion of the related notion of autonomy. We conclude with a summary of our information theoretic description of closure in the context of systems theory and provide an outlook on future research directions.

2 Informational Closure

The basic setting we want to study is shown in Fig. (1). We assume that we have observables S_n describing the state of the system and observables E_n describing the state of the environment at time t_n. At the beginning, these are chosen according to the system-environment distinction of an external observer. According to our view, this distinction is justified insofar as the system, defined by that distinction, can really achieve closure.

The notion of informational closure refers to a situation where the information flow between the environment and the system tends to zero. We measure the information flow from the environment into the system $J_n(E \to S)$ by the conditional mutual information [1]:

$$\begin{aligned} J_n(E \to S) &= MI(S_{n+1}; E_n | S_n) & (1) \\ &= H(S_{n+1}|S_n) - H(S_{n+1}|S_n, E_n) & (2) \\ &= H(E_n|S_n) - H(E_n|S_n, S_{n+1}) \; . & (3) \end{aligned}$$

Here, $H(A)$ denotes the entropy of the probability distribution of the random variable A. For the sake of simplicity, we restrict ourselves to discrete observables, i.e.

$$H(A) = -\sum_{i=1}^{N} p(a_i) \log_2 p(a_i) \tag{4}$$

where the random variable A takes the value a_i with probability $p(a_i)$. For two random variables A and B, $H(A, B)$ denotes the entropy of the joint probability distribution $p(A, B)$ and

$$H(A|B) = H(A, B) - H(B) \tag{5}$$

the conditional entropy of A given B. The mutual information between two random variables A and B is defined as

$$MI(A; B) = H(A) - H(A|B) \; , \tag{6}$$

and the conditional mutual information between A and B given C as

$$MI(A; B|C) = H(A|C) - H(A|BC) \; . \tag{7}$$

The latter measures the reduction of the uncertainty of A given B if C is known additionally, which can be interpreted as the "information flow" from A to B (or in the opposite direction) not transmitted via C.

For the information flow from the environment into the system, as defined in Eq. 1, the following useful identity holds:

$$MI(S_{n+1}; E_n|S_n) = MI(S_{n+1}; E_n) - (MI(S_{n+1}; S_n) - MI(S_{n+1}; S_n|E_n)) \; . \tag{8}$$

This expresses the information flow as the difference between

- the information that S_{n+1} contains about the environment, $MI(S_{n+1}; E_n)$
- and the mutual information between consecutive system states that is related to the environment, $MI(S_{n+1}; S_n) - MI(S_{n+1}; S_n|E_n)$.

We shall now discuss different ways of how a system can achieve informational closure. First there is a trivial case:

[1] This is also known as "transfer entropy" [6] and can be considered as a quantitative expression for the Granger causality [7].

A) Independence: The system and the environment are independent stochastic processes. This implies in particular

$$\begin{aligned} MI(S_{n+1}; E_n) &= 0 \qquad \text{and} \\ MI(S_{n+1}; S_n) - MI(S_{n+1}; S_n|E_n) &= 0 \end{aligned}$$

and therefore informational closure (8).

The "informational closure" becomes non-trivial if the state contains information about the environment, i.e. $MI(S_{n+1}; E_n) \neq 0$:

B) Adaptation: The state of the system contains full information about the part of the environment that interacts with the system, i.e. $H(\hat{e}_n|S_n) = 0$. By the independence structure as stated in Fig. (1) and Eq. (3), this implies closure $J_n = 0$.

For this case, however, two subcases should be distinguished:

B1) Passive adaptation: The system is driven by the environment and adapts passively to all changes in the environment.
In the case of an environment, that appears deterministic to the system, $H(\hat{e}_{n+1}|\hat{e}_n) = 0$, this can be achieved by simply copying the observation of the environment into the system, $S_{n+1} = \hat{e}_n$.

B2) Modeling: The system reaches synchronization and internalizes the correlations observed in the environment by building up own structures.

If the system can act in the environment there is an additional possibility to achieve closure:

C) Control: The system tries to maximize the information flow between its actions and its sensory inputs (called *empowerment* in [8]), i.e. it produces changes in the environment such that these in turn change the state of the system most efficiently. If the actions $\hat{s}$ of the system are functions of the system state, then the consecutive state of the system can anticipate the effect on the environment, which also leads to informational closure.

Since, in the following, we are mainly interested in the case of non-trivial closure we define

$$\begin{aligned} NTIC_m &:= MI(S_{n+1}; E_n, \ldots, E_{n-m}) - MI(S_{n+1}; E_n, \ldots, E_{n-m}|S_n) \\ &= MI(S_{n+1}; E_n, \ldots, E_{n-m}) - MI(S_{n+1}; E_n|S_n) \end{aligned} \tag{9}$$

as a measure for the amount of non-trivial informational closure. Note that a large value of this measure does not ensure a low information flow. It just requires that the system contains more information about the environment than it gathered at this time-step. It should therefore not be considered as a replacement for closure, i.e. low information flow, but as a complementary measure that quantifies the amount of non-trivial closure when there is closure.

3 Simulations

As a simple example we consider an environment described by the hidden Markov model shown in Fig. 2A. The environment can either be deterministic ($p = 1$) or non-deterministic ($p = 0.9$).

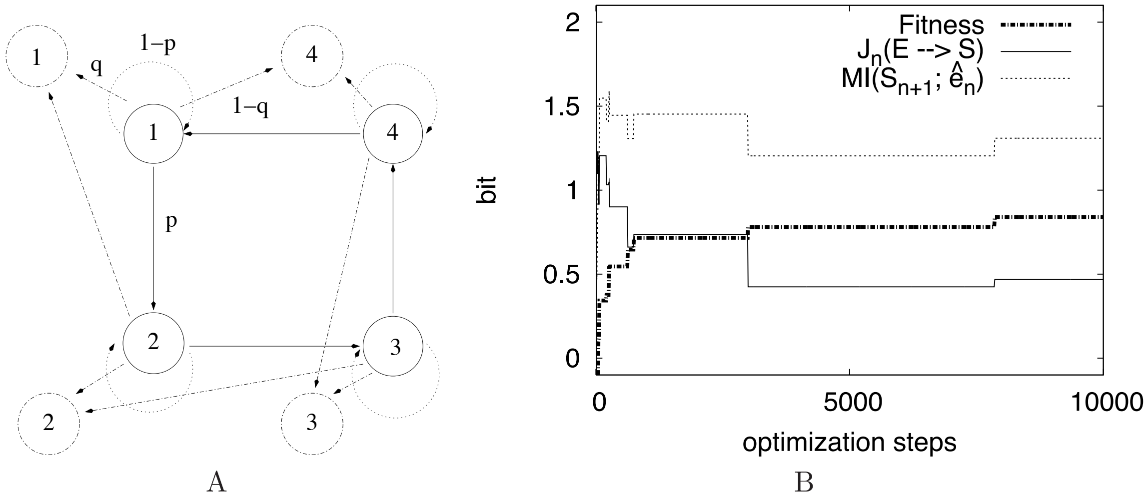


Figure 2: *A) Hidden Markov model used as environment. B) Optimization of Closure and Mutual Information by simulated annealing.*

In both cases, the state observations $\hat{e}$ are noisy ($q = 0.8$), which makes the visible dynamics quite unpredictable ($H(\hat{e}_{n+1}|\hat{e}_n) = 1.23$ bits in the case $p = 1$). This excludes passive adaptation B1 as a strategy to achieve non-trivial informational closure.

The non-deterministic environment is then used to drive a deterministic system, i.e. $S_{n+1} = F(S_n, \hat{e}_n)$. We consider systems with two and four internal states. Note that four internal states are required to fully model the hidden environmental structure. First, we do not consider the case where the system is acting upon the environment, which makes option C (control) unavailable.

In our simulations (see Fig. 2B), we optimize the transition structure F via simulated annealing using the fitness function

$$MI(S_{n+1}; \hat{e}_n) - MI(S_{n+1}; \hat{e}_n|S_n) , \tag{10}$$

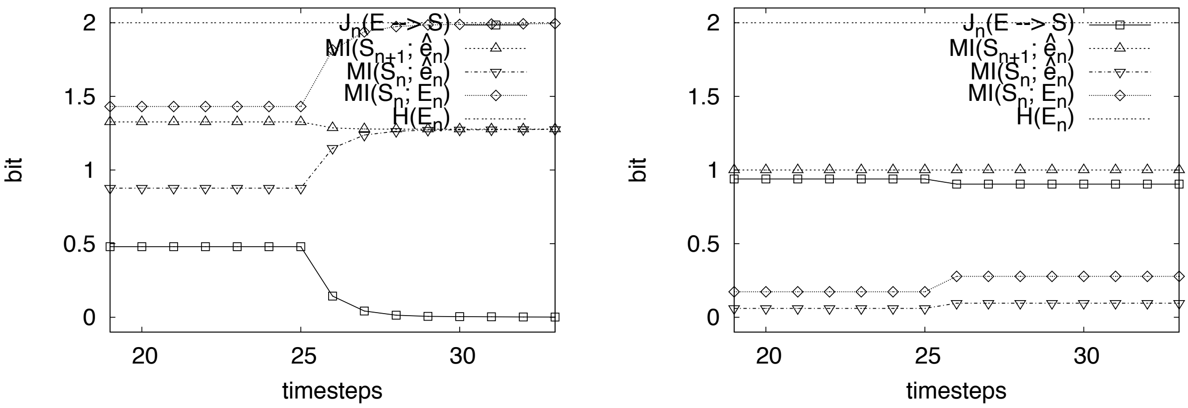
which enforces non-trivial closure through the system gaining mutual information with the environment. Note that in contrast to $NTIC_m$ for $m = 0$ as defined above, we use $\hat{e}_n$ instead of E_n, which makes this measure available to the system itself since it only depends on quantities that either belong to or are observable by the system. The original version for $NTIC_m$ is considered to be more suitable for an external observer since it fully exploits the independence structure of the system-environment interaction.

As expected, due to the non-determinism of the environment, full closure cannot be achieved, but the best system with four internal states that was found after 10000 optimization steps is able to model the hidden environmental dynamics, i.e. $MI(S_n; E_n) > MI(S_n; \hat{e}_n)$. This system achieves full closure, $J_n = 0$, and maximal mutual information (2 bit) when coupled to the deterministic environment ($p = 1$), which shows that it has extracted the deterministic part of the system dynamics (Fig. 3 A).

An optimized system with two internal states is not able to adequately represent the environmental dynamics as seen in Fig. 3 B. Even when coupled to the deterministic environment, the system is far from being informationally closed and contains only little information about the environment.

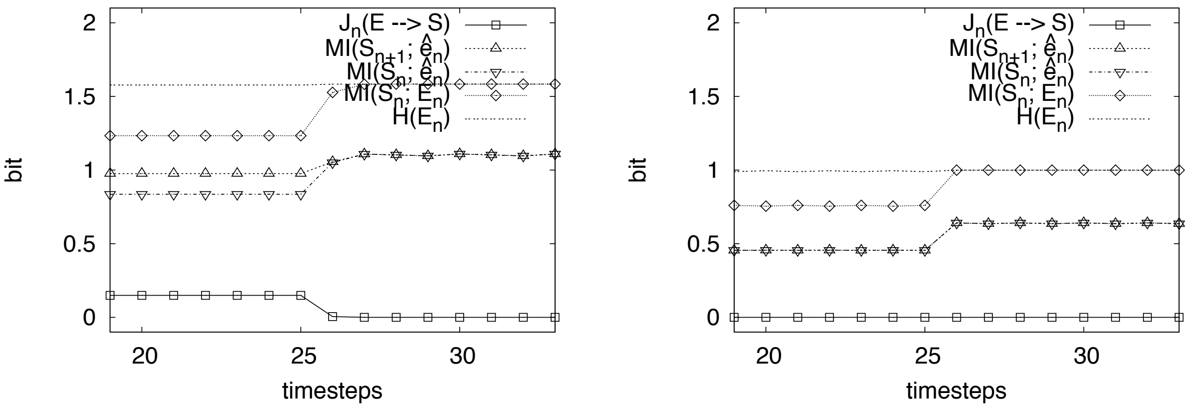
To investigate option C), where the system can control the environment, we change the environment such that:

1. it rotates stochastically, as before, if the system emits a “null” action $\hat{s} = n$
2. and is reset to state 1 if the system emits a reset action $\hat{s} = r$.



A B

Figure 3: *A) Reaction of the system with four internal states when coupled to the deterministic version of the environment (p is changed from 0.9 to 1 after 25 time-steps). B) Same as A), but for the system with two internal states.*



A B

Figure 4: *A) Reaction of the active system with four internal states when coupled to the deterministic version of the resettable environment (p is changed from 0.9 to 1 after 25 time-steps). B) Same as A), but for the system with two internal states.*

As before, we optimize systems with two and four internal states to achieve non-trivial closure. But now, not just the transition structure of the system, but also the actions emitted by the system, were optimized. As can be seen from Fig. 4, this allows for a much better informational closure as compared to the corresponding passive system (Fig. 3). This difference is especially large in the case of two internal states. By using actions to restrict the environmental dynamics, even the two state system can achieve full informational closure and model the deterministic part of the restricted environment (Fig. 4 B). The system with four internal states was found to restrict the system to three states instead of modeling the uncontrolled four state rotation. To quantify the control a system exerts on its environment, we introduce the following measure for *self-efficacy*:

$$
\begin{aligned}
E_s \quad := \quad & MI(S_{n+1}; S_{n-1}|E_{n-1}) \\
& - MI(S_{n+1}; S_{n-1}|S_n, E_{n-1}) - MI(S_{n+1}; S_{n-1}|E_n, E_{n-1}) \qquad (11)
\end{aligned}
$$

Inspired by the *empowerment* measure in [8], this quantity measures the influence of the system state S_{n-1} on S_{n+1} (a generalization to longer time spans can be defined along the same lines) that is transmitted through the environment and simultaneously mediated by the system.

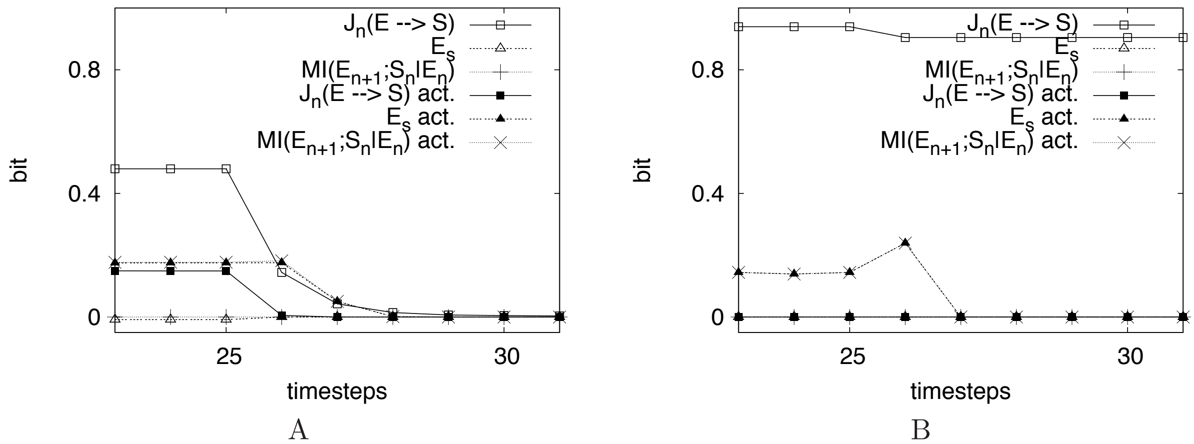


Figure 5: *A) Self-efficacy and information for the systems (passive and active) with four internal states when coupled to the deterministic version of the resettable environment (p is changed from 0.9 to 1 at time-step 25). This figure shows the adaptation to a new environment as in figures 3 and 4 A, but in more detail and compares the active and passive system also with respect to their self-efficacy. B) Same as A), but for the system with two internal states. (Note: E_s is zero in the passive case and therefore not visible in the plot.)*

In contrast to [8], were empowerment is defined as the channel capacity from actions to future sensor inputs, we do not assume a "free will" that can freely choose actions to optimize the transmitted information. Instead, we only consider actions that can occur for the observed internal states of the system and therefore base our measure on the information that actually is transmitted from S_{n-1} to S_{n+1}.
To account for the fact that the information has to be transmitted through the environment, we also subtract the part of the mutual information that corresponds to correlations only inside the system, $MI(S_{n-1}; S_{n+1}|E_{n-1})$. We also require that the influence the system can observe at time-step $n+1$ is still known to the system, i.e. internally represented. To take this into account, we subtract the part of the mutual information that is only transmitted through the environment, i.e. $MI(S_{n-1}; S_{n+1}|S_n)$ which can be seen as a measure of "self-surprise".
The conditioning on E_{n-1} of all quantities removes any information that is shared between E and S already at time-step $n-1$ and can therefore not be attributed to the effect the system had on the environment from time-step $n-1$ to time-step n.
Considering the Markovian structure of the system-environment interaction, i.e. $MI(S_{n-1}; S_{n+1} \mid S_n, E_n) = 0$, Eq. (11) can be considered as a conditional version of the co-information (compare [9]) shared between S_{n-1}, S_n, S_{n+1} and E_n (as shown in Appendix A). Like the co-information, this quantity can become negative. We do not consider this as a problem in this case, since it fulfills the intuitive requirement

$$E_s \leq MI(E_n; S_{n-1}|E_{n-1}). \tag{12}$$

This shows that the self-efficacy of the system is bounded from above by the information flow from the system into the environment.

In Fig. 5, we show the self-efficacy (11) for the systems discussed above. As expected, in the case of purely observing systems that do not influence the environment, the information flow from the system into the environment vanishes and the self-efficacy is slightly smaller than zero. The active systems show a positive self-efficacy that is utilized to achieve better closure when coupled to the stochastic version of the environment. They react to the deterministic environment by a short (slight) increase of the self-efficacy to quickly achieve full closure. Note that the self-efficacy drops to zero when the system is fully synchronized to the environment.

This is due to the fact that in this case no information flow into the environment can be observed, similar to the case of a system that achieves closure by copying the environment as discussed above. Here, the environment can be considered to copy the deterministic dynamics of the acting system and is therefore closed with respect to the system, i.e. the influence of the system is impossible to detect since the environment never uses more states than enforced by the system.

4 Closure and Autonomy

The problem of correct attribution of control also arises in the related notion of autonomy. As described in detail in [10] under the assumption that the system cannot control the environment, corresponding to the cases A and B discussed above, a suitable autonomy measure is given by

$$\begin{aligned} A_m &= MI(S_{n+1}; S_n | E_n, \ldots, E_{n-m}) \\ &= H(S_{n+1}|E_n, \ldots, E_{n-m}) - H(S_{n+1}|S_n, E_n, \ldots, E_{n-m}) \\ &= H(S_{n+1}|E_n, \ldots, E_{n-m}) - H(S_{n+1}|S_n, E_n) \end{aligned} \tag{13}$$

This describes autonomy as the difference between non-heteronomy, measured by $H(S_{n+1}|E_n, \ldots, E_{n-m})$ as the extent to which the system state cannot be determined from the environment, and the intrinsic randomness that also the system does not control ($H(S_{n+1}|S_n, E_n)$).
If instead all mutual information shared between system and environment is attributed to the system, i.e. the system is assumed to be in maximal control of its environment, then the following autonomy measure is more adequate:

$$A^* = MI(S_{n+1}; S_n)$$

Here, the mutual information between successive system states just reflects the autonomy of the system.
These two autonomy measures are closely related to the non-trivial informational closure:

$$NTIC_m = A^* - A_m \tag{14}$$

In the case of trivial informational closure, e.g. if the system is independent of the environment, the two autonomy measures agree. When a system achieves non-trivial informational closure, for example by modeling the environment, we can observe the following

- A^* also has to be positive, actually bigger than $NTIC_m$, since all the environmental correlations that the system can model have to be reflected within the system state.
- A_m can be quite small, because a system that accurately models its environment can be predicted quite well from observation of the environment. An external observer would therefore not attribute much autonomy to such a system since he/she is able predict its state.

Rewriting Eq. (14), we can relate autonomy directly to informational closure as measured by the information flow $J_n(E \to S)$:

$$\begin{aligned} A^* &= A + NTIC \\ &= A + MI(S_{n+1}; E_n, \ldots, E_{n-m}) - J_n(E \to S) \\ \Rightarrow \quad J_n(E \to S) &= A^* - A - MI(S_{n+1}; E_n, \ldots, E_{n-m}) \end{aligned}$$

This demonstrates that a system exhibiting certain internal regularities as measured by $A^* = MI(S_{n+1}; S_n)$ can achieve informational closure either by gaining information about the environment or by increased autonomy, i.e. by becoming unpredictable or uncontrollable from the

environment.
Therefore, information about the environment, i.e. modeling, and autonomy can be considered as complementary strategies for achieving informational closure.

5 Discussion

Our concept of informational closure is a quantitative one. The way we introduce it, it depends on a system-environment distinction that is attributed to an external observer, but not specified further. Such a distinction might seem entirely arbitrary, but this arbitrariness can be reduced through the criterion of informational closure of the system. In other words we can, conversely, use our measure of informational closure to evaluate the employed distinction between system and environment. A higher informational closure then indicates a better identification of the system.
The closure concepts in the literature that we referred to in the introduction aim at a qualitative notion of closure. In particular, autopoiesis as organizational closure in the physico-chemical domain should yield a clear-cut distinction between life and death for organisms instead of a gradual difference. From our point of view, however, in a non-deterministic environment, full non-trivial closure cannot be achieved, since it would require mutual information about the environment and at the same time the vanishing of the information flow J_n. Ultimately, this leads us to the issue of operational closure versus thermodynamic openness in systems theory. Since our measure employs information-theoretic quantities defined in terms of entropies and because the environment always has a higher entropy than any system situated in it, we should naturally expect, that a system, which needs to act adaptively, has to get some new information from the environment.
Also, the relationship of informational closure with autonomy, as developed here, seems relevant in this regard. Our analysis shows that a crucial point for the definition of autonomy is the attribution of control to the system or to the environment. This again brings us back to the system-environment distinction, since we need to decide not only for the states, but also for the processes operating on these states whether they belong to the system or to the environment. Closure then reflects how adequately the system can be described in terms of its own observables.
We have demonstrated that our measure for informational closure gives meaningful results, at least in simple examples where it is unambiguous what the observables are and the control flow is known, e.g. the system cannot influence the environment. Therefore, we expect our closure measure to become a valuable tool for the analysis of system-environment interactions in artificial life simulations. In other cases where the control flow is not unambiguously known, we could also derive interesting conclusions that, admittedly, are open to interpretation.
In order to further investigate the issues raised above, we will develop the concept in the following directions:

- A general framework for quantifying adaptivity, closure and autonomy based on information theory.
- Including the optimization dynamics in the description of the system and describing not only the result, but also the process of adaptation by means of information theory. As long as the system is learning it should then not be informationally closed.
- Formalization of the notion of self-observation – a further internal differentiation of the system is then needed to additionally include a self-description of the system.
- Application of the measures to autonomous robots within a closed sensory-motor loop.

References

[1] H. R. Maturana, F. J. Varela, Autopoiesis and Cognition. The Realization of the Living, Reidel, Dordrecht, 1980.

[2] R. Rosen, Life itself: A comprehensive enquiry into the nature, origin and fabrication of life, Columbia University Press, New York, 1991.

[3] H. H. Pattee, Evolving self-reference: matter, symbols, and semantic closure, Communication and Cognition – Artificial Intelligence 12 (1-2) (1995) 9–28.

[4] N. Luhmann, Probleme mit operativer Schließung, in: Soziologische Aufklärung 6, Westdeutscher Verlag, Opladen, 1995, pp. 12–24.

[5] P. Bourgine, F. J. Varela, Towards a practice of autonomous systems, in: F. J. Varela, P. Bourgine (Eds.), Toward a Practice of Autonomous Systems, MIT Press, Cambridge, Massachusetts, 1991.

[6] T. Schreiber, Measuring information transfer, Phys. Rev. Lett. 85 (2000) 461.

[7] C. W. J. Granger, Investigating causal relations by econometric models and cross-spectral methods, Econometrica 37 (1969) 424–438.

[8] A. S. Klyubin, D. Polani, C. L. Nehaniv, Empowerment: An universal agent-centric measure of control, in: 2005 IEEE Congress on Evolutionary Computation, 2005, http://homepages.feis.herts.ac.uk/ comqdp1/publications.html.

[9] A. J. Bell, The co-information lattice, in: S. Amari, A. Cichocki, S. Makino, N. Murata (Eds.), Proceedings of the Fifth International Workshop on Independent Component Analysis and Blind Signal Separation: ICA 2003, 2003.

[10] N. Bertschinger, E. Olbrich, N. Ay, J. Jost, Autonomy: an information theoretic perspective, for the Workshop on autonomy, ALifeX conference, to appear (2006).

Appendix A

In terms of entropies, Eq. (11) can be rewritten as follows

$$
\begin{aligned}
& MI(S_{n+1};S_{n-1}|E_{n-1}) - MI(S_{n+1};S_{n-1}|S_n,E_{n-1}) - MI(S_{n+1};S_{n-1}|E_n,E_{n-1}) \\
= \; & MI(S_{n+1};S_{n-1}|E_{n-1}) - MI(S_{n+1};S_{n-1}|S_n,E_{n-1}) - MI(S_{n+1};S_{n-1}|E_n,E_{n-1}) \\
& + MI(S_{n+1};S_{n-1}|S_n,E_n,E_{n-1}) \\
= \; & H(S_{n+1}|E_{n-1}) + H(S_{n-1}|E_{n-1}) - H(S_{n+1},S_{n-1}|E_{n-1}) \\
& - [H(S_{n+1},S_n|E_{n-1}) - H(S_n|E_{n-1}) + H(S_{n-1},S_n|E_{n-1}) - H(S_n|E_{n-1}) \\
& - H(S_{n+1},S_n,S_{n-1}|E_{n-1}) + H(S_n|E_{n-1})] \\
& - [H(S_{n+1},E_n|E_{n-1}) - H(E_n|E_{n-1}) + H(S_{n-1},E_n|E_{n-1}) - H(E_n|E_{n-1}) \\
& - H(S_{n+1},E_n,S_{n-1}|E_{n-1}) + H(E_n|E_{n-1})] \\
& + H(S_{n+1},S_n,E_n|E_{n-1}) - H(S_n,E_n|E_{n-1}) \\
& + H(S_{n-1},S_n,E_n|E_{n-1}) - H(S_n,E_n|E_{n-1}) \\
& - H(S_{n+1},S_n,E_n,S_{n-1}|E_{n-1}) + H(S_n,E_n|E_{n-1}) \\
= \; & H(S_{n+1}|E_{n-1}) + H(S_{n-1}|E_{n-1}) + H(S_n|E_{n-1}) + H(E_n|E_{n-1}) \\
& - H(S_{n+1},S_{n-1}|E_{n-1}) - H(S_{n+1},S_n|E_{n-1}) - H(S_{n+1},E_n|E_{n-1}) \\
& - H(S_n,E_n|E_{n-1}) - H(S_{n-1},S_n|E_{n-1}) - H(S_{n-1},E_n|E_{n-1}) \\
& + H(S_{n+1},S_n,S_{n-1}|E_{n-1}) + H(S_{n+1},E_n,S_{n-1}|E_{n-1}) \\
& + H(S_{n+1},S_n,E_n|E_{n-1}) + H(S_n,E_n,S_{n-1}|E_{n-1}) \\
& - H(S_{n+1},S_n,E_n,S_{n-1}|E_{n-1})
\end{aligned}
$$

which is the co-information between S_{n-1}, S_n, E_n and S_{n+1} (conditioned on E_{n-1}).

Appendix B

Proof of Eq. (12):

$$
\begin{aligned}
E_s &= MI(S_{n+1};S_{n-1}|E_{n-1}) - MI(S_{n+1};S_{n-1}|S_n,E_{n-1}) - MI(S_{n+1};S_{n-1}|E_n,E_{n-1}) \\
&= MI(S_{n+1};S_{n-1}|E_{n-1}) - MI(S_{n+1};S_{n-1}|S_n,E_{n-1}) \\
&\quad - (MI(S_{n+1};S_{n-1}|E_{n-1}) + MI(S_{n-1};E_n|S_{n+1},E_{n-1}) - MI(S_{n-1};E_n|E_{n-1})) \\
&= MI(S_{n-1};E_n|E_{n-1}) - MI(S_{n-1};E_n|S_{n+1},E_{n-1}) - MI(S_{n+1};S_{n-1}|S_n,E_{n-1}) \\
&\leq MI(S_{n-1};E_n|E_{n-1})
\end{aligned}
$$