Some ways to see two in one

Martin Biehl[1] and Daniel Polani[1]

[1]University of Hertforshire, Hertforshire, UK
m.biehl@herts.ac.uk

Abstract

We present work towards claryfing whether and how the idea of agents as "subsystems" of an underlying (artificial) universe can be captured formally. For this we propose formal notions of a universe, a decomposition into subsystems and a criterion to prefer some choices of such decompositions over others. Universes are modelled by finite Markov chains, a decomposition is an information conserving set of subprocesses induced by partitions of the state space and our criterion prefers decompositions that improve predictability by minimizing stochastic interaction. Using very simple examples we find three different classes of Markov chains, with respect to their "decomposability". Our approach also highlights the fact that the stochastic interaction of multivariate finite Markov chains crucially depends on the chosen multivariate structure of the state space.

Introduction

Motivation

In this publication we address two problems, how to describe a given system as a composition of subsystems (in this paper, only two), and how to pick suitable descriptions among the possible ones. Note, when we say subsystems, we use this term at first in an intuitive way, without having a rigorous definition. Furthermore, we will use "decomposition" and "description" interchangably as a shorthand when referring to a description as a composition of two subsystems.

The individual subsystems induced by a description should be suitable to represent both agents and environments. Here the agent and its environment could be from biological systems, robots or virtual creatures. This means the class of original systems and the subsystems themselves must be general enough to at least model these three scenarios. Additionally each description should conserve all information about the original system.

Concerning the criterion used to pick descriptions we are eventually seeking one that checks whether the two subsystems can be seen as an agent-environment pair (in other words, form a perception-action loop) and not just a pair of arbitrary subsystems. Note that such a criterion would have to be able to detect agency in the subsystems. We leave this problem to future work. Here we are satisfied with a weaker criterion, one that can be interpreted to measure the "twoness" of the decomposition. Several concepts can be interpreted in this way: a first candidate would be the independence of the two subsystems from each other; a second related but not necessarily identical concept would be the self-determinedness or autonomy of the individual subsystems; a third would be an increase in simplicity of some kind achievable by a decomposition.

There are two scenarios where describing a system as a composition of subsystems is of interest and our approach might be applied in the future. In the first, take the perspective of an external observer of an artificial universe (e.g. a cellular automaton like the game of life). There is to our knowledge no established way to define subsystems which represent interesting entities/objects (say gliders). In artificial life the interesting entities are those that are the most life-like. The subsystems we seek should be able to represent those and the criterion we seek should pick them out. In the second scenario, take the perspective of an agent subject to a sensor input stream. If indeed decomposing the input stream makes it simpler to process, the agent could benefit from this. Additionally, it has been argued (Salge and Polani, 2011) that detecting other agents has merits in its own right. Roughly speaking, other agents produce more information relevant to an agent than is available in other parts of the environment. This suggests, that in a very general way it is advantageous for an agent to view its sensory input stream as being composed of subsystems, some agents and one environment.

Related work

Our work is strongly influenced by Kolchinsky and Rocha (2011). They investigate how learning a model of some given system (a multivariate Markov chain) can be improved by modeling it as a composition of independent subsystems (also Markov chains). In case of a low number of samples to learn from, they find that indeed the predictive error of the model is lowest for a composite model. Their approach is especially convincing in the scenario of an agent subject to an

input stream; the agent modeling the stream as a composite system gaining predictive power in comparison to modeling it as a non-composite one. So in this sense decomposition simplifies prediction for the agent. Note that they vary the compositions of models that are being learned not the descriptions of the original system itself. As we will see the same predictive error can also serve as a useful criterion in the latter case.

Another approach is taken by Balduzzi (2011). He tries to detect emergent subsystems (stochastic processes in general) in a given system (also a multivariate stochastic process, the game of life is used as an example). Here, the subsystems are not chosen once and stay fixed over time, but are composed out of different "units" (groups of variables) at different timesteps and can even skip timesteps altogether. The criterion used to select good choices of subsystems is dependent on how much information (in a specific sense) the different units convey about each other. While this criterion is related to how predictive one unit is of another, this work is not easily comparable with the present one. It is related mainly through the perspective of viewing subsystems as phenomena of an underlying system. This view is also applied specifically to the glider in the games of life by Beer (2013). This is an in depth study of an interesting example of a subsystem in a larger system but up to now the method depends strongly on a preidentified structure.

The system (univariate Markov chain) and subsystems (induced subprocesses) defined by Görnerup and Jacobi (2008) are also used by us. The subsystems are only investigated individually though and compositions that conserve information about the original system are not considered. As a criterion for choosing subsystems the Markov property is used. The Markov property can be interpreted as measuring self-determinedness. We will also employ it but instead of looking directly for Markovian subprocesses we focus on properties of the composition of the subsystems.

One of the authors has previously proposed emergent descriptions (Polani, 2004). Given a stochastic process, emergent descriptions are a set of stochastic processes that contain all information about the original process, are mutually independent and apart from a single one are information conserving. The subsystems we propose here necessarily fullfill the first requirement and a dynamic version of the second one; the third one is not taken into account here.

Once a system has been decomposed into subsystems there are further measures taking into account the interaction of processes that could be used to quantify the suitability of the decompositions. An example employing a framework similar to this paper would be the autonomy measures by Bertschinger et al. (2008).

Methods

System

As an extremely simple model for the original system we assume we are given a stationary finite univariate Markov chain[1] $\{X_t\}_{t\in I}$ (we also refer to it as the universe process) defined by the transition kernel (or Markov matrix)

$$p(x'|x) := p_{x,x'} := Pr(X_{t+1} = x_{t+1}|X_t = x_t) \qquad (1)$$

as the right hand side is independent of t in the stationary case. Our assumption is that the universe process should be Markov, as there is nothing external influencing it and therefore also nothing external to store information about past states in. In the case of the agent facing an input stream this assumption is of course a crude approximation. Choosing finiteness and time discreteness is done for the great reduction of technical issues compared to more general frameworks. Importantly, stationarity may be seen as an approximation, as different choices of subsystems may be most suitable at different times. We choose a univariate process because multivariate processes pre-impose a compositional structure. Moreover, in the finite case each of the variables of the multivariate process takes values in a finite state space; such a process can always be transformed into a univariate process.

Decomposition via coordinatizations

The decompositions we propose describe the original process via a composition of processes induced on partitions of the state space. The subsystems are then those induced processes. To ensure that the original process $\{X_t\}_{t\in I}$ can be retrieved from the subprocesses $\{A_t\}_{t\in I}, \{B_t\}_{t\in I}$ we choose the partitions (we abuse notation and also call them A, B) such that they form Cartesian coordinates of the state space $\mathcal{X} = A \times B$.[2] For each choice of two such partitions A, B we then have a bijective map $f_{(A,B)} : \mathcal{X} \to A \times B$. This map is obtained as follows, denote by $f_A(x) : \mathcal{X} \to A$ the function mapping x to the element a (also called a *block*) of the partition A that it belongs to, and analogously for B. Then define

$$f_{(A,B)}(x) := (f_A(x), f_B(x)) = (a, b). \qquad (2)$$

It is easy to see that the inverse is equal to the intersection of the blocks a and b, which is a unique state x in this case:

$$f_{(A,B)}^{-1}(a,b) := \{x : x \in a \cap b\}. \qquad (3)$$

---

[1] Either we choose the index set I as the integers or we initialize the process in its stationary distribution at some time $t = t_0$ and let it run indefinitly.

[2] This condition ensures for the purpose of this paper the more general requirement, that the subprocesses lose no information about the original process, i.e. $H(X_t|A_t, B_t) = 0$ at all times t, where H denotes the Shannon entropy.

The fact that $\mathcal{X} = A \times B$ ensures that there is always exactly one x in this intersection. For the rest of this paper we will refer to such a pair of partitions as a *coordinatization*.

In this way we obtain an alternative decription of the original process $\{X_t\}_{t \in I}$ and can set:

$$p(a', b'|a, b) := p(f_{(A,B)}^{-1}(a', b')|f_{(A,B)}^{-1}(a, b)). \quad (4)$$

Note that in general the induced subprocesses interact with each other, so that we obtain two interacting stochastic processes. This seems suitable for representing an agent-environment pair which was one of our requirements. At least it is not uncommon to model such interaction in this way; often in the context of perception-action loops (Klyubin et al., 2004; Bertschinger et al., 2008; Ay et al., 2012).

To generate coordinatizations, we note that each map $f_{(A,B)}$ an be visualized as a way of filling an $|A| \times |B|$ grid with the states of $\mathcal{X}$. We will call this grid the *coordinatization grid*. Then rows correspond to the blocks of A and columns correspond to the blocks of B. For example, if $f_{(A,B)}(x) = (a, b)$, the state x will appear in the grid at the intersection of row a and column b. Each coordinatization then corresponds to coordinatization grid and vice versa.

We note though that if we rename the blocks in A (B) this has no influence on the properties of the process induced on A (B) or on the partitions they represent. Renaming the blocks in A (B) corresponds to permuting rows (columns) in the coordinatization grid. We therefore can reduce the number of coordinatizations that we have to investigate by choosing only one of each set of coordinatizations that can be obtained via row or column permutations of their associated grids. This can be achieved for example by always mapping state $x = 1$ to the top left corner; mapping states to the top row such that their values are increasing to the right; and to the leftmost column such that their values are increasing downwards. Instead of $|\mathcal{X}|!$ ways of mapping the states to the grid (and a resulting $|\mathcal{X}|!$ coordinatizations) there are only $|\mathcal{X}|!/|A!||B!|$ possibilities of this kind. This is still too big a number to check for larger systems, but we can already obtain some interesting insights from analysing small ones.

Coordinatizations and modularizations

Coordinatization of a univariate process as defined above results in a bivariate process. The method could easily be extended in order to obtain multivariate processes with $3, 4, 5...$ subprocesses by using $3, 4, 5, ...$-dimensional coordinatization grids. In this sense any finite multivariate process is already a generalized coordinatization of some underlying univariate finite process. Given a multivariate process of k variables it is possible to combine multiple of those variables (e.g. the first few in one group the rest in another) in order to get a bivariate process i.e. a coordinatization (this is a case of what Kolchinsky and Rocha (2011) call modularization). We want to stress that by combining variables

of a multivariate process i.e. by modularization, not all possible coordinatizations can be obtained. Only if the process is viewed as univariate and the coordinatizations are constructed from there, all possible ones are obtained. In other words modularizations result in only a subset of the possible compositional structures. This can be seen when considering a bivariate process as given, combining variables further is then not possible anymore, so only one coordinatization (the given one) can be obtained by modularization. Yet if the process is defined on state space $\mathcal{X}$ and any given structure is ignored, there are $|\mathcal{X}|!/u!v!$ possible ways of mapping the states in $\mathcal{X}$ into a $u \times v$ coordinatization grid and as many coordinatizations for each pair u, v of factors of $|\mathcal{X}|$.

Criterion

To choose the most suitable among the coordinatizations we calculate the *stochastic interaction* (Ay and Wennekers, 2003) $I_{(A,B)}$ (Eq. 5 below) with respect to the coordinatization (A, B) and look for those that minimize it. This can be motivated for both scenarios, the external obeserver and the agent perspective.

In case of the external observer, note that stochastic interaction measures in a specific sense inhowfar the system dynamics is more than a composition of independent subprocesses. Conversely, a low stochastic interaction indicates the composition of independent subprocesses. For a coordinatization a low stochastic interaction can then be interpreted as a measure of "twoness".

Next take the perspective of an agent, who assumes that its input process is a composition of independent subprocesses and models it as such. If the predictive error can be reduced by this assumption the agent would have a good reason to make it. As Kolchinsky and Rocha (2011) have argued, the predictive error of a model is the sum of two terms, the error due to the assumed composition and the error due to imperfection of the learned model. The error due to the assumed composition is quantified by the stochastic interaction and is independent of other features of the learning method or model. The second term of the predictive error (or risk) then quantifies the error due to the imperfection of the parameters of the composite model. This error decreases with the number of samples available to the learner. We do not focus on this second term here but want to find the coordinatization which minimizes the "baseline error" i.e. the stochastic interaction which a possible learned composite model can reach in the best case.

For a bivariate stochastic process like the coordinatizations the stochastic interaction is defined as[3]

$$I_{(A,B)}(X'|X) = \mathrm{KL}\left[p(A', B'|A, B) \,\|\, p(A'|A)p(B'|B)\right] \tag{5}$$

---

[3]For the general definition we refer reader to Kolchinsky and Rocha (2011) and Ay and Wennekers (2003).

KL[.||.] denotes the Kullback-Leibler divergence, which is in this case defined as

$$\text{KL}\left[\, p(A', B'|A, B) \,||\, p(A'|A)p(B'|B)\,\right] :=$$

$$\sum_{a,b} p(a, b) \sum_{a',b'} p(a', b'|a, b) \log \frac{p(a', b'|a, b)}{p(a'|a)p(b'|b)}.$$

The marginalised transition kernels $p(a'|a), p(b'|b)$ are calculated from the transition kernel of the original process and a given starting distribution (e.g. the stationary distribution) $p(x)$:

$$p(a'|a) := \frac{\sum_{x':f_A(x')=a',x:f_A(x)=a} p(x'|x)p(x)}{\sum_{x:f_A(x)=a} p(x)}, \quad (6)$$

$$p(b'|b) := \frac{\sum_{x':f_B(x')=b',x:f_B(x)=b} p(x'|x)p(x)}{\sum_{x:f_B(x)=b} p(x)}. \quad (7)$$

Procedure

The procedure we use for the analysis of the examples below is then the following. We start with a given transition kernel $p(x'|x)$ (or the according Markov matrix with elements $p_{x,x'} = p(x'|x)$) of a finite Markov chain with state space $\mathcal{X} = 1, ..., |\mathcal{X}|$. Next, we calculate the stationary distribution $p(x)$ as the left eigenvector v to the eigenvalue $\lambda = 1$ of $(p_{x,x'})$. We then generate the candidate coordinatizations (A, B) via the coordinatization grid. Finally, among the coordinatizations obtained in this way, we look (brute force) for the coordinatizations that minimize the stochastic interaction, Eq.5, which represent arguably the most "natural" decompositions of $\{X_t\}_{t \in I}$ into two subsystems. For those coordinatizations we also check the Markov property for each of the two induces subprocesses $\{A_t\}_{t \in I}$ and $\{B_t\}_{t \in I}$. This can be done by checking that for any blocks $a, a' \in A$ the total probability $\sum_{x' \in a'} p(x'|x)$ to transition from any $x \in a$ to a' is independent of x (Kemeny and Snell, 1976).

Examples

We study the ideas discussed above in a class of simple systems.

Imagine a box (agent) moving on a $m \times n$ grid wrapped around at the edges. The agent can only move up, down, left, or right (not diagonally), and does so with probability $1/4$ in each direction at every timestep. This system then has mn states, so the state space $\mathcal{X} = \{1, ..., mn\}$ and every agent position corresponds to a state of the system. To make the explanations more transparent, we fix which position of the agent corresponds to which state x. Starting with $x = 1$ in the top left corner, enumerate the top row and continue always from left to right with all other rows (cf. e.g. Fig.1). Note that this can be chosen arbitrarily. All that matters is how these states are mapped to the grid which defines the coordinatization. The latter is an $|A| \times |B|$ grid, and even though the products $|A||B| = mn = |\mathcal{X}|$

are necessarily equal, $|A|, |B|$ can be a pair of factors of $|\mathcal{X}|$ different from m, n. For example, let $m = 3, n = 4$ then $\mathcal{X} = \{1, ..., 12\}$ and the tuples $(|A|, |B|)$ can be any of $(2, 6), (3, 4), (4, 3), (6, 2)$ where the last two possibilities correspond to substituting A and B and will reveal no more than the first two.

We now look at this system for different m and n.

Let $m = 3$ and $n = 3$: The world grid then looks like this:
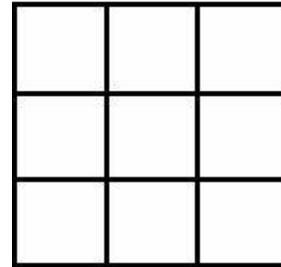
and the state space is $\mathcal{X} = \{1, ..., 9\}$. Because the states of the system are just positions on the grid, as mentioned before we label the positions with the states x in the way shown in Fig. 1. Just to give an impression, the $|\mathcal{X}| \times |\mathcal{X}| =$

Figure 1: World grid labelled with states $\{1, ..., 9\}$ in the way mentioned in the text.

9×9 Markov matrix $(p_{x,x'})$ of this system then looks like this

$$(p_{x,x'}) = \begin{pmatrix} 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & \frac{1}{4} \\ \frac{1}{4} & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & \frac{1}{4} \\ \frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & 0 & \frac{1}{4} & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 \end{pmatrix}. \quad (8)$$

Next, we choose a factor of $|\mathcal{X}|$ as the cardinality $|A|$ of the partition A which will be the number of states of the induced subprocess. This determines also the cardinality of B as $|B| = |\mathcal{X}|/|A|$. In the present case, where $|\mathcal{X}| = 9$ we
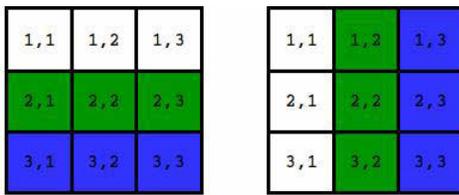
Figure 2: This illustrates the naive coordinatization of the 3×3 world. We show the world grid and overlay the coordinatization, the tupels in each grid position denote $f_{(A,B)}(x) = (f_A(x), f_B(x))$ where x is the state label fixed to the grid position as described in the text. The left (right) diagram highlights the partition A (B) with each block a (b) given the same color.

can only choose $|A| = |B| = 3$ as a nontrivial factorization. Thus, the world grid and the coordinatization grid look the same. Before we calculate the natural coordinatization, let us look at a particular candidate coordinatization, which for lack of a better term we call the *naive coordinatization*.

For this, map the states x into the coordinatization grid in just the same way that we mapped them into the world grid, i.e. $f_{(A,B)}(x) = (\lfloor x/|B| \rfloor, x \mod |B|)$. Then the position of the box on the grid is represented by the tuple (a, b) where a denotes the row and b the column. In Fig.2 we visualize the two partitions A and B and also indicate the resulting labelling of the states (and therefore positions) of the box.

One salient feature of this coordinatization is, that the dynamics can be specified easily:

$$p(a', b'|a, b) = \begin{cases} \frac{1}{4} & a' = a \pm 1,\ b' = b \\ \frac{1}{4} & a' = a,\ b' = b \pm 1 \\ 0 & \text{else.} \end{cases} \quad (9)$$

Surprisingly, this is not the "natural" coordinatization from our perspective here, its stochastic interaction is $I_{(A,B)}(X'|X) = 1$. Yet, there are two coordinatizations that reduce the stochastic interaction $I_{(A,B)}$ to zero. For both, the two partitions A and B are shown in Fig.3. Note that we can transform them into each other by "transposing" the gridlabelling which is due to the fact that we can just exchange A and B. To get an intuition for the solutions which let the stochastic interaction vanish, note that instead of rows and columns, now the blocks divide the world grid into its diagonals. We also found that the induced processes of each of those coordinatizations are Markov.

Let $m = 2$ and $n = 4$: Choose $|A| = 2$ so that $|B| = 4$. Similarly to the 3×3 world, the naive coordinatization leads to a stochastic interaction of $I_{(A,B)}(X'|X) = 1$ and we find 6 different coordinatizations for which it vanishes. In all of the later, partition A consists of the blocks formed by the two diagonals "winding around" the world grid (at least they can
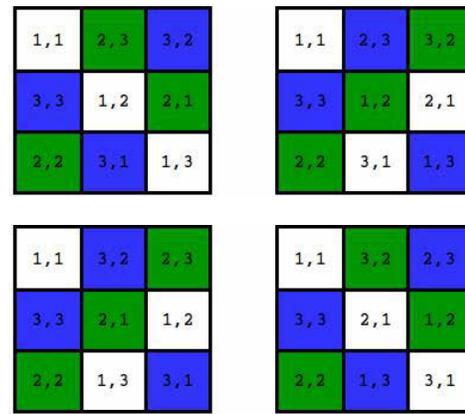

Figure 3: The two different coordinatizations of the 3×3 world process. Each row corresponds to a different coordinatization. Again we highlight partition A (B) in the left (right) diagram and indicate $f_{(A,B)}(x) = (f_A(x), f_B(x))$ at each grid position. Focussing on the first coordinatization (top row), notice that when the agent is in the central position and goes up or right, the first coordinate (the block of partition A) changes from 1 to 2 in both cases, and if it goes down or left it changes form 1 to 3 in both those cases. This means that with probabilities $p_A(2|1) = p_A(3|1) = 1/4 + 1/4 = 1/2$ the agent switches the A blocks. Similarly, for upward or leftward movement the B block changes from 1 to 3 for downward or rightward from 1 to 2, hence $p_B(2|1) = p_A(3|1) = 1/2$. Done similarly for all grid positions, we see that this is the same as if we had two independent agents switching between three positions at each timestep.

always be arranged like this via permutation of the columns of the coordinatization grid) see Fig. 4. The blocks of partition B take a variety of forms. Again all subprocesses are Markov.

Let $m = 3$ and $n = 4$: First let us choose $|A| = 3$ so that $|B| = 4$. Again the naive coordinatization gives $I_{(A,B)}(X'|X) = 1$. The minimum stochastic interaction is non-zero though, with 6 different coordinatizations reaching a value of $I_{(A,B)}(X'|X) \approx 0.43$ see Fig. 5. Notice, that partitions which group together diagonals of the world grid are impossible for the 3×4 grid, because, as a diagonal winds around the grid it only self intersects after traversing the whole state space, so that there is essentially only one diagonal in each direction. Note also that all partitions B that are part of the stochastic interaction minimizing coordinatizations do not induce Markov processes, while the partitions A still do.

Now let us choose $|A| = 2$ so that $|B| = 6$ and the world grid and the coordinatization grid are not of the same form anymore. The coordinatizations with minimum stochastic
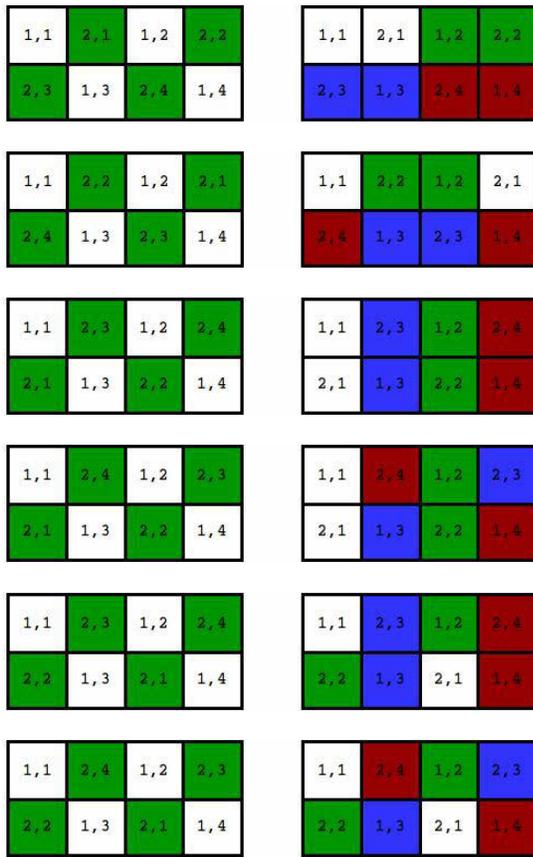
Figure 4: The coordinatizations minimizing stochastic inter-action for the 2×4 world process. The subprocesses induced by the agent dynamics on the partitions B highlighted in the right column are Markov. The same holds for the subprocess induced on the partition A shown on the left.

interaction achieve a value of $I_{(A,B)}(X'|X) \approx 0.31$ see Fig. 6, with both A and B inducing Markov processes. It can be seen that both coordinatizations use the same patterns to cover the world grid. In fact one can obtain the other by permuting the columns of the world grid. The reason the two coordinatizations are seen as different by our algorithm is that in the coordinatization grid they cannot be transformed into one another (see Fig. 7).

Let $m = 2$ **and** $n = 3$**:** In this case we can only choose $|A| = 2$ and $|B| = 3$ which leads to three different coordinatizations that achieve a minimum stochastic interaction of $I_{(A,B)}(X'|X) \approx 0.19$. The process induced on the partitions A is not Markov in theses cases while the process induced on B is.

Discussion

Formally, our examples have revealed three different classes of Markov chains. First, processes that allow coordinatiza-
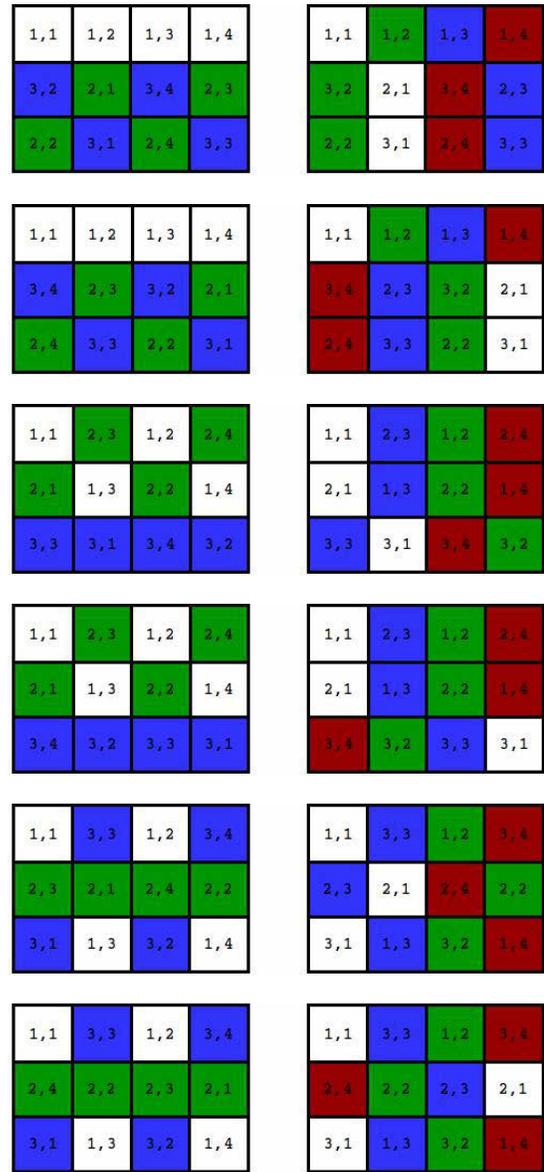
Figure 5: The coordinatizations minimizing stochastic inter-action for the 3×4 world process with $|A| = 3$ and $|B| = 4$. Here the subprocesses induced by the agent dynamics on the partitions B highlighted in the right column are not Markov. Those induced on the partitions A in the left column are.

tions with vanishing stochastic interaction and Markov pro-cesses as the two subprocesses. Second, processes for which no coordinatization with vanishing stochastic interaction ex-ists but whose coordinatizations that minimize stochastic in-teraction induce Markov processes (e.g. the 3×4 world with a coordinatization grid of 2×6). And third, systems where the minimum but non-zero coordinatization contains one subprocess that is not Markov (2×3 world). These re-sults are summed up in the following table ("non-M" stands
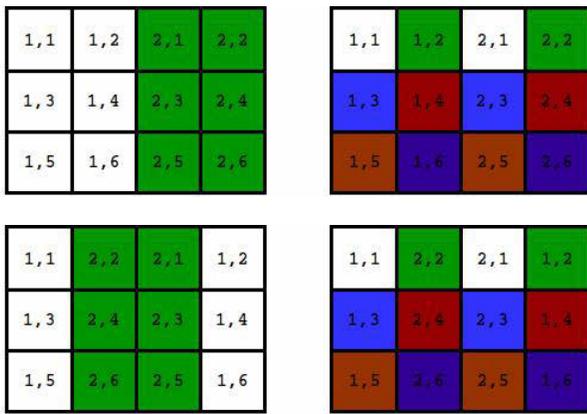
Figure 6: The coordinatizations minimizing stochastic interaction for the 3×4 world process with $|A| = 2$ and $|B| = 6$ which means the coordinatization grid is different form the world grid, see Fig.7. Note that the index pairs $f_{(A,B)}(x) = (f_A(x), f_B(x))$ reflect the 2×6 coordinatization grid. Also note that by switching columns 2 and 4 of the grid in the first row one can obtain the grid in the second row. Also note that the induced processes of all the partitions shown here are Markov.
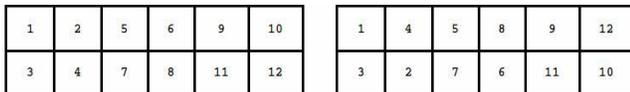
Figure 7: The coordinatization grid for the two different $|A| = 2$ and $|B| = 6$ coordinatizations that minimize stochastic interaction for the 3×4 world process. Shown is the way the states $\mathcal{X} = \{1, ..., 12\}$ are mapped to the blocks of partitions A and B. The rows are the blocks of A and the columns are the blocks of B. For example, in the left coordinatization, state $x = 5$ is mapped to $(a = 1, b = 3)$.

for not Markov and "M" for Markov):

$m \times n$	$|A| \times |B|$	$I_{(A,B)}^{\min}$	$\{A_t\}_{t \in I}$	$\{B_t\}_{t \in I}$
2×3	2×3	0.19	non-M	M
2×4	2×4	0	M	M
3×3	3×3	0	M	M
3×4	3×4	0.43	M	non-M
	2×6	0.31	M	M

The fact that the stochastic interaction vanishes for some coordinatizations and for some not, shows that the stochastic interaction depends crucially on the chosen coordinatization of a stationary finite Markov chain. Systems that might not seem decomposable, like the agent on the 3×3 grid with the naive coordinatization, can in fact still allow clean decompositions like the "diagonal" coordinatization.

On the other hand the processes for which no coordinati-

zation achieves a vanishing stochastic interaction and therefore resist clean decomposition can possibly be seen as fundamentally integrated units. Here comparisons to other measures of integration e.g. Balduzzi and Tononi (2008) suggest themselves and will be pursued in future work.

More generally, our results seem to call for interaction or integration measures that are independent of the coordinatization. A dependence on the chosen cardinalities $|A|, |B|$ (as in the 3×4 world) may still be desirable though.

With respect to the scenario of an external observer, we could show that if an artificial universe (e.g. cellular automata) is formulated in one specific coordinatization (possibly more than two dimensional, e.g. the game of life), this coordinatization might not be the one best suited for decomposing it into subprocesses. Again recall the 3×3 world grid; our system has actually been devised in the way of Eq. 9.

As already mentioned, the decomposition into stochastic processes seems, if only through its generality, capable of accomodating agents and environments, but our examples are inconclusive on this matter.

From the perspective of an agent subject to an input stream, it had been known that the assumption of composed processes can help reduce predictive error. Kolchinsky and Rocha (2011) have shown that given a multivariate process, assuming that groups of variables form independent processes can improve model learning. Our examples show that if the multivariate structure of the process is ignored more compositions of subprocesses can be found and might reduce the predictive error further. We can interpret this in the following ways: The multivariate structure might be imposed by the agents sensory apparatus. Then the modularizations of Kolchinsky and Rocha (2011) could be seen as the best an agent can do. From our perspective the multivariate structure is not fixed though. This corresponds to the situation where the sensory apparatus of the agent is not fixed but can still be adapted (e.g. by evolution). A third more speculative interpretation might be that the agent has to "see through" the multivariate structure and actively conceive the stochastic interaction minimizing decomposition to optimize its predictive power.

Conclusion and outlook

Mathematically, we have proposed an approach for the decomposition of stationary finite Markov chains into pairs of subprocesses which retain all information about the original process, we refer to this as coordinatization. Minimization of stochastic interaction was used to determine "natural" coordinatizations. Three different classes of finite Markov chains showing different kinds of "decomposability" were found in the simple examples treated here.

Importantly, our approach reveals that stochastic interaction depends crucially on the chosen coordinatization. This implies that for a given multivariate Markov chain, con-

structing coordinatizations that ignore the given multivariate structure might achieve cleaner decompositions than any grouping together of the given variables (modularization) can achieve. Such coordinatizations might in fact reduce stochastic interaction to zero. Since stochastic interaction is a lower bound for the predictive error of composite models (Kolchinsky and Rocha, 2011), our approach can in the best case be used to boost the predictive performance of such models.

In practice though the naive search method for the coordinatizations minimizing the stochastic interaction is computationally unfeasible. It remains to be seen if improved search methods can move the approach into the efficient realm.

From an artificial life perspective, we have argued for a natural choice of decompositions of a system into two subsystems for the large class of finite Markov chains. This choice can reveal ways to view and describe systems which might otherwise be overlooked. In principle the subsystems resulting from our decomposition (interacting stochastic processes) seem suitable to represent an agent and its environment or two interacting agents. Whether our criterion can be used to identify them remains an open question.

In the future we hope to extent our approach so that it can serve artificial life researchers as an analytical tool in the context of agent-environment systems or perception-action loops.

References

Ay, N., Bernigau, H., Der, R., and Prokopenko, M. (2012). Information-driven self-organization: the dynamical system approach to autonomous robot behavior. *Theory in Biosciences*, 131(3):161–179.

Ay, N. and Wennekers, T. (2003). Dynamical properties of strongly interacting markov chains. *Neural Netw.*, 16(10):1483–1497.

Balduzzi, D. (2011). Detecting emergent processes in cellular automata with excess information. *Advances in Artificial Life, ECAL*, abs/1105.0158.

Balduzzi, D. and Tononi, G. (2008). Integrated information in discrete dynamical systems: Motivation and theoretical framework. *PLoS Comput Biol*, 4(6):e1000091.

Beer, R. D. (2013). The cognitive domain of a glider in the game of life. *Submitted*.

Bertschinger, N., Olbrich, E., Ay, N., and Jost, J. (2008). Autonomy: An information theoretic perspective. *Biosystems*, 91(2):331–345.

Görnerup, O. and Jacobi, M. N. (2008). A Method for Inferring Hierarchical Dynamics in Stochastic Processes. *Advances in Complex Systems*, 11(01):1–16.

Kemeny, J. G. and Snell, J. L. (1976). *Finite Markov Chains: With a New Appendix "Generalization of a Fundamental Matrix"*. Springer.

Klyubin, A., Polani, D., and Nehaniv, C. (2004). Organization of the information flow in the perception-action loop of evolved agents. In *2004 NASA/DoD Conference on Evolvable Hardware, 2004. Proceedings*, pages 177–180.

Kolchinsky, A. and Rocha, L. M. (2011). Prediction and modularity in dynamical systems. *Advances in Artificial Life, ECAL*, pages 423–430.

Polani, D. (2004). Defining emergent descriptions by information preservation. *InterJournal Complex Systems*, (1102).

Salge, C. and Polani, D. (2011). Digested information as an information theoretic motivation for social interaction. *Journal of Artificial Societies and Social Simulation*, 14(1):5.