Quantifying Morphological Computation based on an Information Decomposition of the Sensorimotor Loop

Keyan Ghazi-Zahedi[1], Johannes Rauh[2]

[1]Max Planck Institute for Mathematics in the Scienes, Inselstrasse 22, 04103 Leipzig, Germany
[2] Leibniz Universität Hannover, Welfengarten 1, 30167 Hannover, Germany
zahedi@mis.mpg.de, rauh@math.uni-hannover.de

Abstract

The question of how an agent is affected by its embodiment has attracted growing attention in recent years. A new field of artificial intelligence has emerged, which is based on the idea that intelligence cannot be understood without taking the embodiment into account. The contribution of an agent's embodiment to its behaviour is also known as morphological computation. In this work, we propose a quantification of morphological computation, which is based on an information decomposition of the sensorimotor loop into shared, unique and synergistic information. Using a simple model of the sensorimotor loop, we show that the unique information of the body with respect to the environment is a good measure for morphological computation.

Introduction

Morphological computation is discussed in various contexts, such as DNA computing and self-assembly (see Pfeifer et al., 2007c; Hauser et al., 2012, for an overview). In this publication, we are interested in quantifying morphological computation of embodied agents which are embedded in the sensorimotor loop. Morphological computation, in this context, is described as the trade-off between morphology and control (Pfeifer and Scheier, 1999), which means that a well-chosen morphology, if exploited, substantially reduces the amount of required control (Montúfar et al., 2014). Here, the term *morphology* refers to the agent's body, explicitly including all its physiological and physical properties (shape, sensors, actuators, friction, mass distribution, etc.) (Pfeifer, 2002). The consensus is that morphological computation is the contribution of the morphology and environment to a behaviour, that cannot be assigned to a nervous system or a controller. There are several examples from biology that demonstrate how the behaviour of an agent relies on the interaction of the body and environment. A nice example is given by Wootton (1992, see p. 188), who describes how "active muscular forces cannot entirely control the wing shape in flight. They can only interact dynamically with the aerodynamic and inertial forces that the wings experience and with the wing's own elasticity; the instantaneous results of these interactions are *essentially* determined by the architecture of the wing itself [...]"

One of the most cited example from the field of embodied artificial intelligence is the Passive Dynamic Walker by McGeer (1990). In this example, a two-legged walking machine preforms a naturally appealing walking behaviour, as a result of a well-chosen morphology and environment, without any need of control. There is simply no computation available and the walking behaviour is the result of the gravity, the slope of the ground and the specifics of the mechanical construction (weight and length of the body parts, deviation of the joints, etc.). If any parameter of the mechanics (morphology) or the slope (environment) is changed beyond a small threshold, the walking behaviour will not persist. In this context, we understand the exploitation of the body's and environment's physical properties as the embodiments effect on a behaviour.

It is important to note that talking about a system's behaviour in the context of embodied artificial intelligence does not make much sense if there is no agency involved. This means that the e.g. the Passive Dynamic Walker is basically nothing more than an interesting mechanical system. Yet, its purpose is to study how mechanical properties of the legs affect human walking, which is why it is often cited as an example for morphological computation. In the same way, the interaction of the physical properties of insect wings in the given example are meaningless, if there is no agent that exploits the properties to achieve a behaviour of interest (flying in this case of insects, walking in case of humans). In this paper, we focus on the quantification of body's physical interaction with it's environment, to which we refer as morphological computation. This does not mean that we are not aware that the flapping of the wings requires a control, i.e., that morphological computation needs to be induced by an agent. Discussing both aspects of morphological computation in detail is beyond the scope of this work, although they are both addressed by our information decomposition of the sensorimotor loop, as we discuss later.

Theoretical work on describing morphological computation in the context of embodied artificial intelligence has been conducted by (Hauser et al., 2011; Füchslin et al., 2012). In this publication, we study an information-theoretic

approach to quantifying morphological computation which is based on two of our previous publications: In (Zahedi and Ay, 2013) we have investigated different quantifications of morphological computation, which all match the general intuition, but showed different results when applied to a simple model of the sensorimotor loop. In (Bertschinger et al., 2014) we have derived a general decomposition of a mutual information of three random variables into unique, shared, and synergistic information (Bertschinger et al., 2014). Here, we apply this information decomposition to the simple model of the sensorimotor loop in order to improve our previous measures of morphological computation.

The paper is organised in the following way. The next section discusses the sensorimotor loop and its representation as a causal graph. The third section describes the bivariate information decomposition from Bertschinger et al. (2014). Based on the information decomposition, the fourth section introduces the unique information as a measure for morphological computation in the sensorimotor loop. The fifth section presents numerical results, which are then discussed in the final section. An appendix explains how we computed our measure of morphological computation.

Sensorimotor Loop

Our information theoretic decomposition of the mutual information requires a formal representation of the sensorimotor loop, which is introduced in this section. In our understanding, a cognitive system consists of a brain or controller, which sends signals to the system's actuators, thereby affecting the system's environment. We prefer the notion of the system's *Umwelt* (von Uexkuell, 1934; Clark, 1996; Zahedi et al., 2010), which is the part of the system's environment that can be affected by the system and which itself affects the system. The state of the actuators and the *Umwelt* are not directly accessible to the cognitive system, but the loop is closed as information about the *Umwelt* and the body is provided to the controller through the sensors. In addition to this general concept of the sensorimotor loop, which is widely used in the embodied artificial intelligence community (see e.g. Pfeifer et al., 2007a) we introduce the notion of *world* and by that we mean the system's morphology and the system's *Umwelt*. We can now distinguish between the intrinsic and extrinsic perspective in this context. The world is everything that is extrinsic from the perspective of the cognitive system, whereas the controller, sensor and actuator signals are intrinsic to the system. This is analogous to the agent-environment distinction in the context of reinforcement learning (Sutton and Barto, 1998), in which the environment is understood as everything that cannot be controlled arbitrarily by the agent.

The distinction between intrinsic and extrinsic is also captured in the representation of the sensorimotor loop as a causal or Bayesian graph (see Fig. 1). For simplicity, we only discuss the sensorimotor loop for reactive systems.

This is plausible, because behaviours which exploit the embodiment are usually better described as reactive and not as deliberative. The most prominent examples are locomotion behaviours, e.g. human walking, swimming, flying, etc., which are all well-modelled as reactive behaviours.

The random variables S, A, and W refer to sensor, actuator, and world state, and the directed edges reflect causal dependencies between the random variables (see Klyubin et al., 2004; Ay and Polani, 2008; Zahedi et al., 2010). Everything that is extrinsic is captured in the variable W, whereas S and A are intrinsic to the agent. The random variables S and A are not to be mistaken with the sensors and actuators. The variable S is the output of the sensors, which is available to the controller or brain, the action A is the input that the actuators take. Consider an artificial robotic system as an example. Then the sensor state S could be the pixel matrix delivered by some a sensor and the action A could be a numerical value that is taken by a motor controller to be converted in currents to drive a motor.

Throughout this work, capital letters (X, Y, ...) denote random variables, non-capital letters (x, y, ...) denote specific values that random variables can take, and calligraphic letters ($\mathcal{X}$, $\mathcal{Y}$, ...) denote the alphabets for the random variables. For example, the random variable X may take the value $x \in \mathcal{X}$. Greek letters (α, β, ...) refer to generative kernels, i.e. kernels which describe an actual underlying mechanism or a causal relation between random variables.

The random variables that we consider depend on time, which we model as a discrete parameter $t \in \mathbb{N}$. For example, the output of the sensors corresponds to a sequence of random variables $S_1, S_2, \ldots$, with one random variable S_t for each time step t. We are mostly interested in what happens in a single time step. Therefore, we use the following notation. Random variables without any time index refer to some fixed time t and primed variables to time $t + 1$. For example, the two variables S, S' refer to S_t and S_{t+1}.
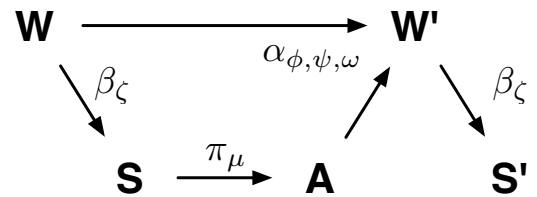

Figure 1: A formal model of the sensorimotor loop.

Formally, the sensorimotor loop is given by the probability distribution $p(w)$ and the kernels $\alpha(w'|w, a)$, $\beta(s|w)$, and $\pi(a|a)$, see Figure 1. We choose the same parameterisable binary model of the sensorimotor loop as in (Zahedi and Ay, 2013) (with an additional synergistic parameter, see Eq. (1)). It allows us to control the causal dependencies of

S, A, and W individually, and thereby enables us to evaluate the information decomposition in the sensorimotor loop and compare the result with our previous results. The model is given by the following set of equations:

$$\alpha_{\phi,\psi,\omega}(w'|w,a) = \frac{e^{\phi w'w+\psi w'a+\omega w'wa}}{\sum_{w''\in\Omega}e^{\phi w''w+\psi w''a+w''wa}} \quad (1)$$

$$\beta_\zeta(s|w) = \frac{e^{\zeta sw}}{\sum_{s''\in\Omega}e^{\zeta s''w}} \quad (2)$$

$$\pi_\mu(a|s) = \frac{e^{\mu as}}{\sum_{a'\in\Omega}e^{\mu a's}} \quad (3)$$

$$p_\tau(w) = \frac{e^{\tau w}}{\sum_{w''\in\Omega}e^{\tau w''}}, \quad (4)$$

where $a,w,s,w'\in\Omega=\{\pm 1\}$ and $\phi,\psi,\omega,\zeta,\mu,\tau\geq 0$. As in (Zahedi and Ay, 2013), we make the following two simplifying assumptions, which do not restrict generality too much. First, we assume that all world states $w\in\Omega$ occur with equal probability, i.e. $p(w=1)=p(w=-1)=\nicefrac{1}{2}$. Second, we assume a deterministic sensor, i.e. $\zeta\gg 1\Rightarrow p(s|w)=\delta_{sw}$, i.e. the sensor is a copy of the world state. The first assumption does not reduce generality much, because it only assures that the world state itself does not already encode some structure, which is propagated through the sensorimotor loop. The second assumption does not violate the generality of the model, because in a reactive system, the sensor state S and A can be reduced to a common state, with a new generative kernel $\gamma(a|w)=\sum_s\pi(a|s)\beta(s|w)$. Hence, keeping one of the two kernels deterministic and varying the other in the experiments below, does not reduce the validity of this model. This leaves four open parameters ψ,ϕ,ω, and μ, against which the morphological computation measure is validated.

Information decomposition is most often discussed in the context of binary logical functions such as XOR, and so it is useful to think of the one-step sensorimotor loop as such a logical function. The one-step sensorimotor loop is a simplified model of the causal diagram of the sensorimotor loop (which unfolds over time) that we did not present and discuss due to spacial constraints. A more thorough discussion can be found in (Zahedi et al., 2010; Zahedi and Ay, 2013; Pfeifer et al., 2007b). The one-step sensorimotor loop is sufficient to discuss, visualise and evaluate the information decomposition for the following reason. The decomposition and quantification of morphological computation is based on the joint distribution $p(w',w,a)$ which can be obtained e.g. by observation. An example for such an application would be the recording of an animal's motion with 3D motion capturing and vibromyography (muscle activity) sensors. The joint distribution $p(w',w,a)$ can then be obtained from the recorded data (which implicitly means that we assume ergodicity). The presented model (see Eqs. (1) to (4) and Fig. 1) is chosen such that it allows us to freely parametrise the joint distribution.

Information Decomposition

Next, we introduce the information decomposition that underlies our measure of morphological computation. We first explain this information decomposition in a general information theoretic setting and later explain how we use it in the sensorimotor loop.

Consider three random variables X,Y,Z. Suppose that a system wants to predict the value of the random variable X, but it can only access the information in Y or Z. How is the information that Y and Z carry about X distributed over Y and Z? In general, there may be *redundant* or *shared* information (information contained both Y and Z), but there may also be *unique* information (information contained in only one of Y or Z). Finally, there is also the possibility of *synergystic* or *complementary* information, i.e. information that is only available when Y and Z are taken together. The classical example for synergy is the XOR function: If Y and Z are binary random variables and if $X=Y\,\text{XOR}\,Z$, then neither Y nor Z contain any information about X (in fact, X is independent of Y and X is independent of Z), but when Y and Z are taken together, they completely determine X (in particular, X is not independent from the pair (X,Y)).

The total information that (Y,Z) contains about X can be quantified by the mutual information $MI(X:(Y,Z))$. However, there is no canonical way to separate these different kinds of information. Mathematically, one would like to have four functions $SI(X:Y;Z)$ ("shared information"), $UI(X:Y\setminus Z)$ ("unique information of Y"), $UI(X:Z\setminus Y)$ ("unique information of Z"), $CI(X:Y;Z)$ ("complementary information") that satisfy

$$\begin{aligned}MI(X:(Y,Z)) = SI(X:Y;Z)+UI(X:Y\setminus Z)\\ +UI(X:Z\setminus Y)+CI(X:Y;Z).\end{aligned} \quad (5)$$

From the interpretation it is also natural to require

$$\begin{aligned}MI(X:Y)=SI(X:Y;Z)+UI(X:Y\setminus Z),\\ MI(X:Z)=SI(X:Y;Z)+UI(X:Z\setminus Y).\end{aligned} \quad (6)$$

A set of three functions SI, UI, and CI that satisfy (5) and (6) is called a *bivariate information decomposition* by Bertschinger et al. (2014). It follows from the defining equations and the chain rule of mutual information that an information decomposition always satisfies

$$MI(X:Y|Z)=UI(X:Y\setminus Z)+CI(X:Y;Z). \quad (7)$$

Equations (5) and (6) do not specify the functions SI,UI, and CI. Several different candidates have been proposed so far, for example by Williams and Beer (2010) and Harder et al. (2013). We will use the decomposition of Bertschinger et al. (2014) that is defined as follows[1]:

---

[1]The same functions were also proposed by Griffith and Koch (2014) starting from a measure for "union information" obtained from formal information-theoretic arguments.

Let Δ be the set of all possible joint distributions of X, Y, and Z. Fix an element $P \in \Delta$ (the "true" joint distribution of X, Y, and Z). Define

$$\Delta_P = \Big\{ Q \in \Delta :$$
$$Q(X = x, Y = y) = P(X = x, Y = y)$$
$$\text{and } Q(X = x, Z = z) = P(X = x, Z = z)$$
$$\text{for all } x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z} \Big\}$$

as the set of all joint distributions which have the same marginal distributions on the pairs (X, Y) and (X, Z). Then

$$UI(X : Y \setminus Z) = \min_{Q \in \Delta_P} MI_Q(X : Y | Z),$$
$$SI(X : Y ; Z) = \max_{Q \in \Delta_P} CoI_Q(X ; Y ; Z),$$
$$CI(X : Y ; Z) = MI(X : (Y, Z))$$
$$- \min_{Q \in \Delta_P} MI_Q(X : (Y, Z)),$$

where CoI denotes the interaction information (McGill, 1954), sometimes also called co-information. Here, a subscript Q in an information quantity means that the quantity is computed with respect to Q as the joint distribution.

One idea behind these functions is the following: Suppose that the joint distribution P of X, Y, and Z is not known, but that just the marginal distributions of the pairs (X, Y) and (X, Z) are known. This information is sufficient to characterize the set Δ_P, but we do not know which element of Δ_P is the true joint distribution. One can argue that the UI and SI should be constant on Δ_P; that is, shared information and unique information should depend only on the interaction of X and Y and the interaction of X and Z, but not on the threeway interaction.

The second property that characterizes the information decomposition is that the set Δ_P contains a distribution Q such that $CI_Q(X : Y ; Z) = 0$. In other words, when only the marginal distributions of the pairs (X, Y) and (X, Z) are known, then we cannot know whether there is synergy or not. See (Bertschinger et al., 2014) for a more detailed justification and a proof how these properties determine the functions UI, SI, and CI.

In Bertschinger et al. (2014), the formulas for UI, CI, and SI are derived from considerations about decision problems in which the objective is to predict the outcome of X. Here, we want to apply the information decomposition in another setting: We will set $X = W'$, $Y = W$, and $Z = A$. In our setting, W and A not only have information about W', but they actually *control* W'. However, the situation is similar: In the sensorimotor loop, we also expect to find aspects of redundant, unique, and complementary influence of W and A on W'. Formally, since everything is defined probabilistically, we can still use the same functions UI, CI, and SI. We believe that the arguments behind the definition

of UI, CI and SI remain valid in the setting of the sensorimotor loop where we need it. First, it is still plausible that unique and redundant contributions should only depend on the marginal distributions of the pairs (W, W') and (A, W'). Second, in order to decide whether W and A act synergistically, it does not suffice to know only these marginal distributions. Therefore, we believe that the functions UI, CI, and SI have a meaningful interpretation. In particular, we hope to be able to use the information decomposition in order to measure morphological computation. This view is supported by our results below, which indicate that the functions UI, CI and SI do indeed lead to a reasonable decomposition of $MI(W' : (A, W))$ and that the unique information $UI(W : W' \setminus A)$ is a reasonable measure of morphological computation, at least in our simple model of the sensorimotor loop.

The parameters of our model of the sensorimotor loop (Eqs (1) to (4)) can also be interpreted in terms of an information decomposition. Intuitively, ϕ corresponds to the unique influence of W on W', ψ corresponds to the unique influence of A on W', and ω corresponds to the complementary influence. However, the role of the other parameters ζ, μ, τ is less clear, and there is no clear correspondence for redundant information. The information decomposition has the advantage, that its definition does not depend on a parametrization. Note that if the "synergistic parameter" $\omega = 0$ vanishes, then it does not necessarily follow that $CI(W' : A; W) = 0$ (see Fig. 2). However, we do expect the complementary information to be small in this case.

Morphological computation

Morphological computation was described as the contribution of the embodiment to a behaviour. In our previous work, we derived two concepts to quantify morphological computation, which are both based on the world dynamics kernel $\alpha(w'|w, a)$.

The first concept assumes that the current action A has no influence on the next world state W', in which case the kernel $\alpha(w'|w, a)$ reduces to $\hat{\alpha}(w'|w)$. If this is the case, we would say that the systems shows maximal morphological computation, as the behaviour is completely determined by the world. To measure the amount of morphological computation present in a recorded behaviour, we calculated how much the data differed from the assumption by calculating the weighted Kullback-Leibler divergence $\sum_{w,a} p(w, a) D_{KL}(\alpha(w'|w, a) \| \hat{\alpha}(w'|w))$, which is the conditional mutual information $MI(W' : A|W)$. Because this quantity is zero if we have maximal morphological computation, we inverted and normalised in the following way: $1 - MI(W' : A|W) / \log_2 |W|$.

The second concept started with the complementary assumption that the current world state W had no influence on the next world state W', i.e., that the world dynamics kernel is given by $\tilde{\alpha}(w'|a)$. Morphological compu-

tation was then quantified as the error from the assumption, given by the weighted Kullback-Leibler divergence $\sum_{w,a} p(w,a) D_{KL}(\alpha(w'|w,a)\|\tilde{\alpha}(w'|a))$, which equals the conditional mutual information $MI(W' : W|A)$.

Both concepts were analysed and quantifications were derived, which didn't require knowledge about the world, but could be calculated from intrinsically available information only. At that time, we could not determine which of the two concepts would capture morphological computation best, although both concepts and their intrinsic adaptations lead to different results in a specific configuration ($\psi = \phi \approx 0$).

Our intention in this publication is to answer this question. For this purpose, we follow a different approach to quantify morphological computation, by starting with the mutual information of $MI(W' : (W,A))$ and decompose it into the shared, unique and synergistic information, as described in the previous section. Replacing X, Y, Z by W', W, A in Eq. (5), we obtain the following decomposition:

$$MI(W' : (W,A)) = SI(W' : W; A) + UI(W' : W \setminus A)$$
$$+ UI(W' : A \setminus W) + CI(W' : W; A)$$
$$(8)$$

By Eq. (7), our previous concept two, the conditional mutual information $MI(W' : W|A)$, is given by the sum of the unique information $UI(W' : W \setminus A)$ and the synergistic information $CI(W' : W; A)$:

$$MI(W' : W|A) = UI(W' : W \setminus A) + CI(W' : W; A).$$
$$(9)$$

The examples we have discussed in the introduction (insect wing and Passive Dynamic Walker) suggest to use the unique information $UI(W' : W \setminus A)$ to quantify morphological computation, because it captures the information that the current and next world state W, W' share uniquely. The next section presents numerical results to investigate how the conditional mutual information $MI(W' : W|A)$ and the unique information $UI(W' : W \setminus W)$ compare with respect to quantifying morphological computation.

Experiments

Experiments are conducted on the parameterised model of the sensorimotor loop (see Fig. 1 and Eqs. (1) to (4)). As stated earlier, we set $\tau = 0$, i.e. the world state W is drawn with equal probability ($p(w = -1) = p(w = 1) = 1/2$), and $\zeta \gg 0$ such that the sensor state S is a copy of the world state W. This leaves four parameters for variation, namely the three world dynamics kernel parameters ϕ, ψ, ω and the policy parameter μ. We decided to plot the information theoretic quantities only for $\mu = 0$ (see Figs. 2 and 3), i.e., for the case, in which the action A is chosen independently of the current sensor value S and with equal probability. This allows us to investigate the effect of the action A on the next world state W', without any influence of W on A. We also
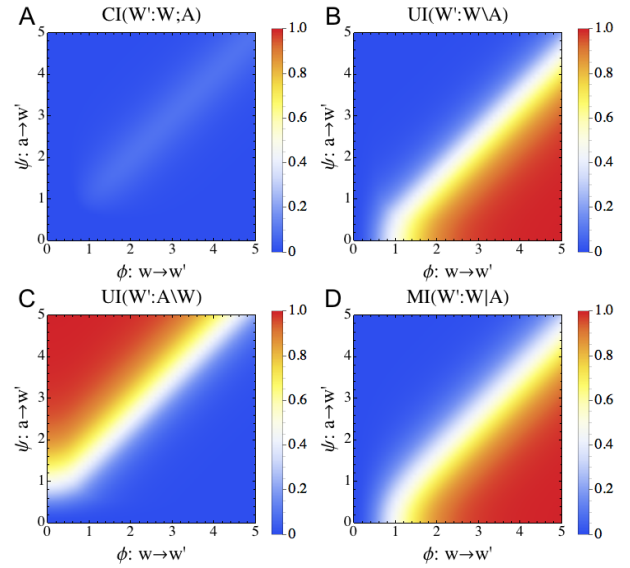

Figure 2: Information decomposition for $\mu = 0.0, \omega = 0.0$

know from previous experiments (see Zahedi and Ay, 2013), that the conditional mutual information $MI(W' : W|A)$ drops to zero for increasing μ. Thus, by Eq. (9), unique and synergistic information also decrease with increasing μ. If A is deterministically dependent on W, it also follows that the unique information $UI(W' : A \setminus W)$ is zero, because A and W are interchangeable. The only quantity that will be larger than zero is the shared information, which, by definition, is not of interest in the context of this work.

Due to spacial constraints, we decided to plot the information decomposition for varying ϕ (parameter of unique influence of W on W') and ψ (parameter of unique influence of A on W') for two different values of ω (parameter of synergistic influence of W, A on W', see Eq. (1)). This allows us to investigate the effect of the synergistic parameter on the information decomposition. Fig. 2 shows the results for $\omega = 0$, while Fig. 3 shows the results for $\omega = 2$. We will first discuss the results for $\omega = 0$, as they are best comparable with our previous results from (Zahedi and Ay, 2013).

Vanishing synergistic parameter ($\omega = 0$): Fig. 2A shows that synergistic information $CI(W' : W; A)$ is small and only present if $\psi \approx \phi$ (diagonal of the image). This is in agreement with our intuition that ω is the synergistic parameter. The unique information of the action A and the next world state W', denoted by $UI(W' : A \setminus W)$, is shown in Figure 2C. The plot reveals that $UI(W' : A \setminus W)$ is only present when $\psi > \phi$, and it is large whenever ψ is significantly larger than ϕ. Figure 2B shows analogous results for the unique information $UI(W' : W \setminus A)$. In this case, the unique information is negligible whenever $\phi \lesssim \psi$, and
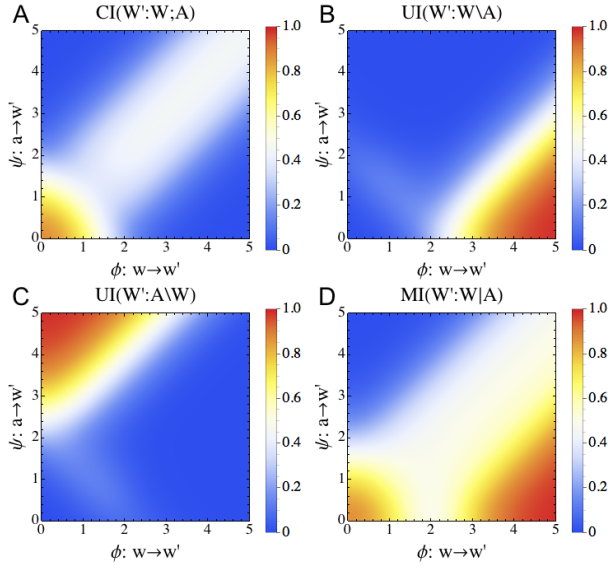
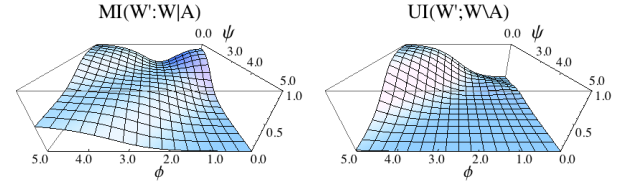Figure 3: Information decomposition for $\mu = 0.0, \omega = 2.0$

Figure 4: $MI(W' : W|A)$ and $UI(W' : W \setminus A)$ for $\omega = 2$, replotted from Figure 3 to stress the differences between the two measures. The plot on the left-hand side shows more clearly (as compared to Figure 3D) that there is a large domain in which $MI(W' : W|A)$ is indifferent.

it grows whenever ϕ is significantly larger than ψ. These two plots show that the definition of the unique information, as proposed by Bertschinger et al. (2014), is able to extract the unique influence in a setting in which two random variables actually control, i.e., causally influence a third random variable. Fig. 2D shows the conditional mutual information $MI(W' : W|A)$, which was the second concept of quantifying morphological computation in our previous work (Zahedi and Ay, 2013). As stated earlier, the conditional mutual information is given by the sum of the unique and synergistic information (Eq. (9)). Hence, there is almost no difference between Figure 2B and Figure 2D, except on the diagonal, where the unique information $UI(W' : W \setminus A)$ is slightly smaller.

Positive synergistic parameter ($\omega = 2$): To study the difference between $UI(W' : W \setminus A)$ and $MI(W' : W|A)$, and hence, to compare the new quantification with our former concept, we conducted the same experiments with a value of $\omega = 2$ (see Figs. 3 and 4). Figs. 3A-C demonstrate how the information decomposition can distinguish between the synergistic information and the unique informations, which is exactly what we need to quantify morphological computation. The unique information $UI(W' : W \setminus A)$ captures only the information that the current world state W and the next world state W' share, and therefore, captures the common understanding of morphological computation in the context of embodied artificial intelligence. In the introduction, we presented two examples of morphological computation, which described it as the contribution of the body and environment to a behaviour that cannot be assigned to any neural system or robot controller. The unique

information $UI(W' : W \setminus A)$ (see Fig. 3B) captures this notion of morphological computation best, because it vanishes if the synergistic information $CI(W' : W; A)$ (see Fig. 3A) or the unique information $UI(W' : A \setminus W)$ (see Fig. 3C) increases. Given Eq. (9), it is clear that the conditional mutual information $MI(W' : W|A)$ is positive (see Fig. 3D) whenever the unique information $UI(W' : A \setminus W)$ or the synergistic information $CI(W' : W; A)$ is positive. This is problematic for the following reason. Fig. 3D shows a large value of $MI(W' : W|A)$ also for values of $\psi > \phi$, which is counter-intuitive. Furthermore, as Fig. 4 shows (note that the $\phi\psi$ axes are rotated for better visibility), the conditional mutual information is indifferent for a large range of $|\phi - \psi| < d$. Additionally, the conditional mutual information increases for vanishing ϕ and ψ, which again is counter-intuitive, whereas $UI(W' : W \setminus A)$ (see right-hand side of Fig. 4) nicely reflects our intuition. Therefore, we conclude that the unique information $UI(W' : W \setminus A)$ is best suited to quantify morphological computation in the context of embodied artificial intelligence.

Discussion

This work proposes a quantification of morphological computation based on an information decomposition in the sensorimotor loop. In the introduction, morphological computation was described as the contribution of an agent's body and agent's *Umwelt* to its behaviour. Important to note is that both mentioned examples highlighted the contribution of the embodiment that resulted solely from interactions of the body and environment and that cannot be attributed to any type of control by the agent. This is why we propose to use a decomposition of the mutual information $MI(W' : (W, A))$ into shared, unique and synergistic information. This allows us to separate contributions of the embodiment from contributions of the controller (via its actions A) and contributions of both, controller and embodiment.

We showed that the information decomposition is related to our previous work in the following way. The sum of the unique information $UI(W' : W \setminus A)$ and the synergistic information $CI(W' : W; A)$ is equal to the conditional mu-

tual information $MI(W' : W|A)$, which is one of our two earlier concepts for morphological computation. This relation shows the difference of this work compared to our former results. We are now able to quantify exactly how much of the next world state W' is determined by the current world state W, thereby excluding any influence of the action A. Therefore, we propose $UI(W' : W \setminus A)$ as a quantification of morphological computation.

We evaluated the decomposition in a parametrised, binary model of the sensorimotor loop. The world dynamics kernel $\alpha(w'|w, a)$ was parametrised with three parameters, ϕ, ψ, and ω, which roughly relate to the unique information $UI(W' : W \setminus A)$, the unique information $UI(W' : A \setminus W)$, and the synergistic information $CI(W' : W; A)$. For a fixed value of ω, the two parameters ϕ and ϕ were varied to evaluate the information decomposition in the sensorimotor loop. We showed that when the synergistic parameter vanishes ($\omega = 0$), synergistic information is present only for $\phi \approx \psi$. This explains why there is only a marginal difference between $UI(W' : W \setminus A)$ and $MI(W' : W|A)$ in this setting. For a positive synergistic parameter $\omega = 2$, we showed that the synergistic information was positive for a much larger domain, which led to a significant difference between $UI(W' : W \setminus A)$ and $MI(W' : W|A)$. In particular, the condition mutual information $MI(W' : W|A)$ was positive for a larger range of parameter values ψ and ϕ. There is a domain $|\phi - \psi| < d$, for which the conditional mutual information $MI(W' : W|A)$ is positive and indifferent. One would expect to see a higher morphological computation mostly when $\phi > \psi$, despite the fact that synergistic information is present. This shows that $UI(W' : W \setminus A)$ is better suited to quantify morphological computation.

We mentioned in the introduction of this paper that e.g. the flapping of the wing has a component of morphological computation that is independent of any control (the architecture of the wing interacting with the environment) and a component which is induced by the action (flapping of the wings). The synergistic information $CI(W' : W; A)$ captures the second part and should be investigated as a measure for a different type of morphological computation on its own. Unfortunately, this is beyond the scope of this work.

Zahedi and Ay (2013) proposed that a measure of morphological computation could be used as a guiding principle in an open-ended self-organised learning setting. For this purpose, the measure should only depend on information that is intrinsically available to the system. Clearly, this is not the case for $UI(W' : W \setminus A)$. Therefore, future work will include derivations of the information decomposition, which only include intrinsically available information. It would also be interesting to investigate how much a formalisation of the information decomposition can benefit from a consideration of the causal information flow (Ay and Polani, 2008; Ay and Zahedi, 2014). The starting point for our decomposition was the mutual information $MI(W' : (W, A))$, which is a correlational measure and not a measure of causal dependence, as e.g. proposed by Pearl (2000). In currently ongoing work, we are applying the quantification to motion capturing data of real robots.

References

Ay, N. and Polani, D. (2008). Information flows in causal networks. *Advances in Complex Systems*, 11(1):17–41.

Ay, N. and Zahedi, K. (2014). On the causal structure of the sensorimotor loop. In Prokopenko, M., editor, *Guided Self-Organization: Inception*, volume 9 of *Emergence, Complexity and Computation*. Springer.

Bertschinger, N., Rauh, J., Olbrich, E., Jost, J., and Ay, N. (2014). Quantifying unique information. *Entropy*, 16(4):2161–2183.

Clark, A. (1996). *Being There: Putting Brain, Body, and World Together Again*. MIT Press, Cambridge, MA, USA.

Füchslin, R. M., Dzyakanchuk, A., Flumini, D., Hauser, H., Hunt, K. J., Luchsinger, R. H., Reller, B., Scheidegger, S., and Walker, R. (2012). Morphological computation and morphological control: Steps toward a formal theory and applications. *Artificial Life*, 19(1):9–34.

Griffith, V. and Koch, C. (2014). Quantifying synergistic mutual information. In Prokopenko, M., editor, *Guided Self-Organization: Inception*, volume 9 of *Emergence, Complexity and Computation*, pages 159–190. Springer Berlin Heidelberg.

Harder, M., Salge, C., and Polani, D. (2013). Bivariate measure of redundant information. *Phys. Rev. E*, 87:012130.

Hauser, H., Ijspeert, A., Füchslin, R., Pfeifer, R., and Maass, W. (2011). Towards a theoretical foundation for morphological computation with compliant bodies. *Biological Cybernetics*, 105:355–370.

Hauser, H., Sumioka, H., Füchslin, R. M., and Pfeifer, R. (2012). Introduction to the special issue on morphological computation. *Artificial Life*, pages 1–8.

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

McGeer, T. (1990). Passive dynamic walking. *International Journal of Robotic Research*, 9(2):62–82.

McGill, W. (1954). Multivariate information transmission. *Information Theory, Transactions of the IRE Professional Group on*, 4(4):93–111.

Montúfar, G., Ay, N., and Ghazi-Zahedi, K. (2014). A framework for cheap universal approximation in embodied systems. *CoRR (submitted)*, abs/1407.6836.

Pearl, J. (2000). *Causality: Models, Reasoning and Inference*. Cambridge University Press.

Pfeifer, R. (2002). Embodied artificial intelligence - on the role of morphology and materials in the emergence of cognition. In Schubert, S. E., Reusch, B., and Jesse, N., editors, *Informatik bewegt: Informatik 2002 - 32. Jahrestagung der Gesellschaft für Informatik e.v. (GI)*, volume 19, Bonn. GI.

Pfeifer, R., Lungarella, M., and Iida, F. (2007a). Self-organization, embodiment, and biologically inspired robotics. *Science*, 318(5853):1088–1093.

Pfeifer, R., Lungarella, M., and Iida, F. (2007b). Self-organization, embodiment, and biologically inspired robotics. *Science*, 318(5853):1088–1093.

Pfeifer, R., Packard, N., Bedau, M., and Iida, F., editors (2007c). *Proceedings of the International Conference on Morphological Computation*.

Pfeifer, R. and Scheier, C. (1999). *Understanding intelligence*. MIT Press, Cambridge, MA, USA.

Sutton, R. S. and Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. MIT Press.

von Uexkuell, J. (1957 (1934)). A stroll through the worlds of animals and men. In Schiller, C. H., editor, *Instinctive Behavior*, pages 5–80. International Universities Press, New York.

Williams, P. and Beer, R. (2010). Nonnegative decomposition of multivariate information. *arXiv:1004.2515v1*.

Wootton, R. J. (1992). Functional morphology of insect wings. *Annual Review of Entomology*, 37(1):113–140.

Zahedi, K. and Ay, N. (2013). Quantifying morphological computation. *Entropy*, 15(5):1887–1915.

Zahedi, K., Ay, N., and Der, R. (2010). Higher coordination with less control – a result of information maximization in the sensori-motor loop. *Adaptive Behavior*, 18(3–4):338–355.

Appendix: Computing UI, SI, and CI.

In this appendix we shortly explain how we computed the functions UI and CI. The appendix of (Bertschinger et al., 2014) explains how to parametrize the set Δ_P and how to solve the optimization problems in the definitions of UI, CI, and SI. In our case, where all variables are binary, Δ_P consists of all probability distributions $Q_{\gamma_{-1},\gamma_{+1}}$ with

w'	w	a	$Q_{\gamma_{-1},\gamma_{+1}}(w',w,a)$
-1	-1	-1	$P(w',w,a) + \gamma_{-1}$
-1	-1	+1	$P(w',w,a) - \gamma_{-1}$
-1	+1	-1	$P(w',w,a) - \gamma_{-1}$
-1	+1	+1	$P(w',w,a) + \gamma_{-1}$
+1	-1	-1	$P(w',w,a) + \gamma_{+1}$
+1	-1	+1	$P(w',w,a) - \gamma_{+1}$
+1	+1	-1	$P(w',w,a) - \gamma_{+1}$
+1	+1	+1	$P(w',w,a) + \gamma_{+1}$

The range of the two parameters $\gamma_{\pm 1}$ is restricted in such a way that $Q_{\gamma_1,\gamma_{+1}}$ has no negative entries. Since every entry $Q_{\gamma_{-1},\gamma_{+1}}(w',w,a)$ involves only one of the two parameters, Δ_P is a rectangle, bounded by the inequalities

$$\max\{-P(-1,-1,-1), -P(-1,+1,+1)\} \le \gamma_{-1},$$
$$\min\{P(-1,-1,+1), P(-1,+1,-1)\} \ge \gamma_{-1},$$
$$\max\{-P(+1,-1,-1), -P(+1,+1,+1)\} \le \gamma_{+1},$$
$$\min\{P(+1,-1,+1), P(+1,+1,-1)\} \ge \gamma_{+1}.$$

To approximately solve the optimization problem we computed the values on a grid and took the optimal value. This simple procedure yields an approximation that is good enough for our purposes.