

Think Globally, Sense Locally: From Local Information to Global Features

Malte Harder, Daniel Polani, Chrystopher L. Nehaniv
Adaptive Systems Research Group
University of Hertfordshire
Hatfield, AL10 9AB United Kingdom
Email: {m.harder, d.polani, c.l.nehaniv}@herts.ac.uk

Abstract—Information theory can be used to quantify morphological and topological features of a collective of agents in an environment. In particular we considered chains of agents equipped with local sensors to detect neighbours. We then investigated different information theoretic quantities regarding certain global features of the collective, ranging from the information a single agent provides to the information that is contained in the sensor readings of the whole collective.

I. INTRODUCTION

Pattern formation and morphogenesis are central topics in studying the development of biological systems. How do functional differentiation, body axes, skin patterns or nervous systems emerge from a single cell?

One of the first theoretically inspired studies on morphogenesis were made by Turing in the fifties with the model of reaction-diffusion systems [1]. Reaction diffusion systems are models of the concentration of substrates distributed in space. In these models two processes can change the concentration: local “chemical” reactions and diffusion. This leads to the development of spatial patterns, like stripes, spots or spirals and explains how certain patterns can form from a homogeneous initial state [2]. Gierer and Meinhardt considered specific variants of reaction-diffusion models, activator-inhibitor systems and used them to explain the formation of organising regions and primary gradients [3]–[5].

Also robotics and especially nano-robotics are concerned with the problem of pattern formation in the design of self assembling systems. Here agent based models are generally used and there are several constructive approaches to create languages that specify formation processes. Christensen et al deployed a system to the SWARM-BOT platform that allows a collective of robots to assemble in specific shapes by passing rules between connected robots [6], [7]; De Rosa et al presented a language based on predicates that can be used to specify local interactions between modules of a modular robot and hence the shape of it [8].

For a general introduction to pattern formation in an artificial life context see [9]. Beyond that, there has been more specific work done on pattern formation within the artificial life community: a study of a model that is inspired by the chemotactic behaviour of the slime mold *Physarum polycephalum* [10], cellular models using cell-adhesion [11],

evolutionary dynamics of morphogenesis [12] and stigmergic algorithms to build artificial nest-like structures [13].

A well known benchmark for morphogenesis is the French flag problem, that is, the formation of three bands of distinctly differentiated cells using a gradient of a morphogen [14]–[16]. The challenge is that a cell on a local scale has to decide into which colour it needs to differentiate where the colour depends on the absolute position which is a global property of the system. The morphogen gradient allows a cell to sense its absolute horizontal position and hence solves the problem into which colour to differentiate.

What is common in these scenarios is the principle that global properties are sensed on a local level: From languages in robotics that are specifically designed to establish and communicate global features, over morphogens distributed by a reaction-diffusion system to stigmergy that uses the environment to manifest and exchange information that is locally otherwise not senseable for the agents.

This information is not a vague abstract quantity, it can be specifically measured using information theory. Global features of an agent collective have informational content that needs to be processed locally by an individual agent to gain a degree of certainty about the particular feature. Information about a global feature can be partially provided on a local level and we will study how much information an agent already can acquire locally about certain global spatial features.

Information theory provides conservation laws concerning the information that needs to be injected into a system to control it as well as the information that can be observed within the system [20]. This is of great interest as conservation laws make it possible to state constraints and requirements for self-organising system. Moreover, information theory allows us to be mostly agnostic towards the actual mechanics behind a system. Thus, we intentionally ignore the mechanism by which information is transferred, whether specific message passing protocols are used as in [6], local sensor readings trigger hormones to achieve long distance signalling, or timing problems need to be considered as in the firing squad synchronisation problem [17]. Here we are exclusively focusing on how agents use locally available information while constrained by information theoretic considerations.

There have been first studies on stigmergy using information theory [18], however with a focus on the information flow

between agents. As said above, we are not interested in stigmergy as a concept of information transfer, but in the general question of how can we quantify the information content in a global property and how much of this information can be acquired by local sensors?

As an example for a global feature let us consider the property that a chain of matchsticks form a loop. We can make sense of the question “How much information is contained in this feature?” by reformulating it as an information theoretic problem: Suppose the world we investigate consists of just three matchsticks on a table that are connected at their ends in a chain. Knowing whether the chain forms a loop or not reduces the number of possible ways they can be connected in either case. If it forms a loop, only a triangle is allowed, if it does not form a loop, only chains where the inner angles do not sum to 180 degree are allowed. This means that the uncertainty of the match configuration is reduced once we gain the knowledge whether the chain forms a loop. Thus we established an information theoretic formulation of a topological feature.

As an anecdote consider the Indian story of “the blind men and an elephant”, where a group of blind men touch an elephant at different parts of its body to sense what it is like. The first one touches the trunk and says an elephant is like a snake, the second man touches the side of the elephant and says it is like a wall and another one touches the tusk and says it is like a spear, and so on. We now consider an agent collective that is “like a group of blind men” as individual agents within a collective are usually limited by local sensors. Now, we will ask how much information do these local sensors provide about particular features and use information theory to investigate this question.

II. INFORMATION THEORY

Information theory was introduced by Shannon in 1948 [19] to study information transmission in communication networks. In recent years however, applications in control theory [20], neural networks [21] as well as artificial life [22]–[24] showed that information theory provides more than just a tool to analyse communication and information transmission from a classical engineering view.

We will give a brief introduction to the basics of information theory here, for a detailed account on information theory, see Thomas and Cover [25]. A fundamental measure in information theory is the *entropy* of a probability distribution, measured in bits. It is defined by $H(X) = -\sum_x p(x) \log_2 p(x)$ where X denotes a finite-valued random variable with values in \mathcal{X} and $p(x)$ the probability that X takes on the value $x \in \mathcal{X}$. Entropy measures the uncertainty of the outcome of a random variable. Given a second random variable Y the *conditional entropy* is

$$H(Y|X) = -\sum_{x,y} p(x)p(y|x) \log_2 p(y|x)$$

and measures the uncertainty of Y knowing the outcome of X . To relate these, *mutual information* is defined by $I(X; Y) =$

$H(Y) - H(Y|X)$. Hence, mutual information is a measure of how much the uncertainty of Y is reduced if we know the value of X . It is symmetric in X and Y and can be interpreted as the information the random variables mutually encode about each other. Again, this can be conditioned on a third random variable Z which gives the *conditional mutual information* $I(X; Y|Z) = H(Y|Z) - H(Y|X, Z)$. The conditional mutual information can be greater or less than the unconditioned mutual information.

III. FEATURES AND SENSORS

Let \mathcal{X} be an arbitrary state space, X its random variable and φ a function from the state space into some feature space \mathcal{F} . A common example is the feature space $\mathcal{F} = \{0, 1\}$ where the function φ tests whether a certain property is present (1) or not (0). The feature function then induces a random variable $F = \varphi(X)$ taking values on the feature space, that is, $p(f|x) = \delta_{\varphi(x)f}$, where δ denotes the Kronecker delta (a function that takes the value 1 if both entries are equal and 0 otherwise). The information of a feature depending on X is now given by $I(X; F) = H(F) - H(F|X)$. In this case the information is equal to the entropy of the feature $H(F)$ as the outcome of the feature is deterministic given the state, that is, $H(F|X) = 0$.

If the state space represents a collective of agents in an environment and they are equipped with sensors we can ask how much information regarding a feature can be detected by the sensors. Therefore let $\mathbf{S}^{(k)} = (S_1, \dots, S_k)$ denote the k -tuple of sensor variables which then gives the Bayesian network of Fig. 1. A Bayesian network denotes the conditional dependencies between random variables using a directed acyclic graph.

Now we can measure the information that these sensors provide about the feature; it is given by $I(F; \mathbf{S}^{(k)})$. This is the maximum amount of information the sensors can possibly gain about the value of feature F .

Assume now that readings are sent around without having information about their source attached. This for example usually happens when collectives use stigmergy to transmit information between individuals. That means, we cannot identify which sensors correspond to which sensor values, but have access to the values themselves, thus we have possibly less information about the feature. To measure the amount of information that is provided in this case let $\tilde{\mathbf{S}}^{(k)}$ denote the random variable on the set of bags of sensor values. A bag of sensor readings, denoted by $[s_1, \dots, s_k]$, is similar to a set but can contain elements more than one time and in contrast to a k -tuple without any order on the elements.

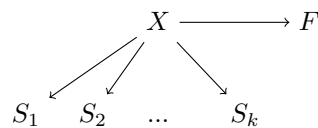


Fig. 1. Bayesian network of the feature and the sensors, determined by the conditional distributions $p(s_i|x)$, $i = 1, \dots, k$, $p(f|x)$ and $p(x)$.

Each k -tuple of sensor readings (s_1, \dots, s_k) can be mapped to the bag that contains the same readings $[s_1, \dots, s_k]$. The probability that a bag of sensor readings is observed is now given by the sum of the probabilities of all k -tuples that are mapped to this particular bag. So, the distribution is now given by

$$p([s_1, \dots, s_k] | x) = \sum_{\sigma \in \mathcal{S}([s_1, \dots, s_k])} p(s_{\sigma(1)}, \dots, s_{\sigma(k)} | x)$$

where $x \in \mathcal{X}$ and $\mathcal{S}([s_1, \dots, s_k])$ denotes the set of all multiset permutations of $[s_1, \dots, s_k]$, that is, the set of all permutations of k indices such that for two distinct permutations $\sigma, \tau \in \text{Sym}(k)$, $\sigma \neq \tau$ the following holds $(s_{\sigma(1)}, \dots, s_{\sigma(k)}) \neq (s_{\tau(1)}, \dots, s_{\tau(k)})$. That means permutation of the indices that lead to the same k -tuple of sensor readings are only accounted for once in the sum.

The mutual information $I(F; \tilde{\mathcal{S}}^{(k)})$ now is the amount of information the sensor variables together give about the feature when we cannot identify which sensor reading comes from which sensor.

On the other hand $I(F; S_i)$ gives the amount of information a single sensor S_i provides about the given feature. The difference between $I(F; S_i)$ and $I(F; \mathcal{S}^{(k)})$ is now the information missing for each sensor to have all locally available information regarding the feature. Again, it is possible to consider the scenario where we cannot attribute a sensor reading to a specific sensor. Now, let S' denote the random variable for such a sensor reading s' , then

$$p(s' | x) = \sum_{s_1, \dots, s_k} p(s' | s_1, \dots, s_k) p(s_1, \dots, s_k | x)$$

where

$$p(s' | s_1, \dots, s_k) = \frac{1}{k} \sum_{i=1}^k \delta_{s' s_i}.$$

That means we assume that the sensor reading s' could be the reading of any of the sensors. The information an arbitrary reading provides about the feature F is now given by $I(F; S')$.

IV. GRID-WORLD TOPOLOGY

We investigated collectives of single-celled agents in a grid-world. We consider two dimensional grid-worlds as an approximation of small scale environments like those of 2D cell collectives or possible nano-robots on surfaces.

In particular we considered chains of k agents. We consider a chain as a connected set of agents such that we can locally determine up to two unique *direct neighbours* that allow to enumerate the whole chain. Therefore certain local neighbourhoods as shown in Fig. 2 are not allowed. In Fig. 2 a) it is not possible to determine whether the agent in the centre has the top left and middle left or the bottom left and middle left agents as direct neighbours which would contradict the requirement that the direct neighbours are uniquely determined. In Fig. 2 b) we have a similar situation 1 whereas in Fig. 2 c) the chain would have a branch which is also not desired in a chain. We can observe that a branch

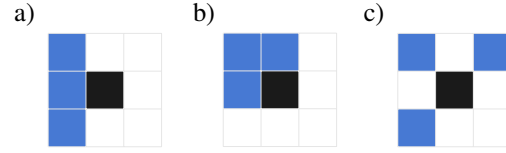


Fig. 2. Neighbourhoods that are ruled out by the chain condition. a),b) The agent has three neighbours whereas only two are permitted if the neighbourhood is only divided into one area by the neighbours. c) The neighbourhood is divided into three areas.

always appears when the one cell neighbourhood of an agent is divided into three separated areas.

We can now characterise a chain in a grid-world as a connected set of k agents that satisfies certain conditions: Each agent can have up to four neighbours of which up to two can be diagonal neighbours and up to two straight neighbours, if there are more of any type the collective automatically has a branch or non unique direct neighbours. Furthermore if the one cell neighbourhood of an agent is only separated into one area, only two neighbours are allowed. See Fig. 3 for an example of two chains consisting of five agents.

The shape of a chain and the features we are interested in are translation invariant, therefore we only consider different shapes of chains: Our state space $\mathcal{X}^{(k)}$ consists of all different shapes of chains that are formed by k agents. We assume that the corresponding random variable $X^{(k)}$ is uniformly distributed. It is a sensible choice to assume that the space of all chains in a particular grid-world is uniformly distributed when we have no prior knowledge about a possible distribution of configurations. Here however, we are just considering the shapes of all chains for reasons of computational feasibility and not all possible ways to position a chain of given shape in a particular grid-world. Therefore, a uniform distribution of $X^{(k)}$ is only an approximation to the actual distribution of $X^{(k)}$ given a uniform distribution of chains in a grid-world of given size, but one can show that this approximation improves the larger the actual grid-world gets for a fixed number of agents.

A. Sensors

Each agent has a local sensor that can perceive other agents in the 8-connected neighbourhood of the agents location. Agents cannot sense the borders of the world and all agents are identical in terms of the sensor value they trigger given the same neighbourhood. That means each agent has a 8-bit sensor where each bit denotes whether it can see another agent in that particular field or not.

Using this sensor an agent can locally determine its direct neighbours, that is, the two agents next to it in a ordering of the chain. The direct neighbours of an agent x are then given by the following set, where $N_4(x)$ denote the neighbour agents that are within a 4-connected one cell neighbourhood of x and $N_8(x)$ the neighbour agents that are within a 8-connected one cell neighbourhood:

$$N(x) = N_4(x) \cup \{y \in N_8(x) | N_4(y) \cap N_4(x) = \emptyset\}.$$

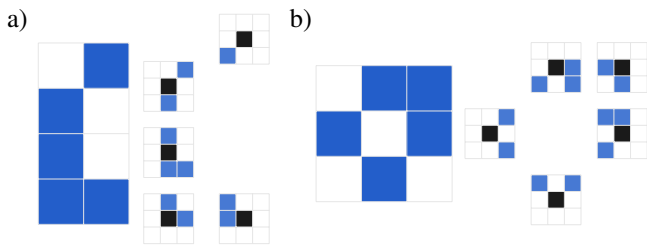


Fig. 3. Two possible shapes of chains a),b) consisting of five agents (left) with the corresponding agents' sensor readings (right, dark square denotes the position of the agent).

This means a direct neighbour is any agent in a 4-connected neighbourhood and any agent that is a diagonal neighbour that itself is not next to an agent in the 4-connected neighbourhood. I.e the non-diagonal neighbours have a higher priority to be the direct neighbour.

B. Loops and Diameter

The global configuration itself can serve as feature that we want to recover given the information of the local sensors. Furthermore we considered the “loopiness” of a chain, denoted by the random variable $L^{(k)}$ on the set $\{0, 1\}$ that is induced by a function that checks whether the chain forms a loop or not and the diameter of a chain, denoted by the random variable $D^{(k)}$ on the set $\{0, \dots, k\}$ that is induced by a function that returns the maximum of the vertical and horizontal extent of the chain.

V. INFORMATION IN CHAINS

With this setup we computed the information of the mentioned features for chains of different lengths up to $k = 9$ agents and investigated how the information is distributed among the agents. The computations were done by enumerating possible configurations of chains. It can be shown that this approximates quite well grid world environments that have a larger extent than the chains. The first observation is that the entropy $H(X^{(k)})$ grows almost linearly while the entropy of the loop random variable $H(L^{(k)})$ decreases. However this comes at no surprise as the number of chains grows almost exponentially but for more agents it gets more improbable that they form a loop.

A. Full Information Access

First we computed the information that the sensors have about the selected features $X^{(k)}, L^{(k)}$ and $D^{(k)}$ when we cannot assign a sensor reading to a particular agent, which we introduced by the random variable $\tilde{S}^{(k)}$. This information determines the shape completely for a chain of up to four agents, that is, $H(X^{(k)}) = I(X^{(k)}; \tilde{S}^{(k)})$ for $k \leq 4$. Starting with five agents we lose unambiguity (see Table I), however it is still possible to determine whether the chain forms a loop or not (see Table II). The knowledge about the diameter decreases with the number of agents as the knowledge of the whole configuration does.

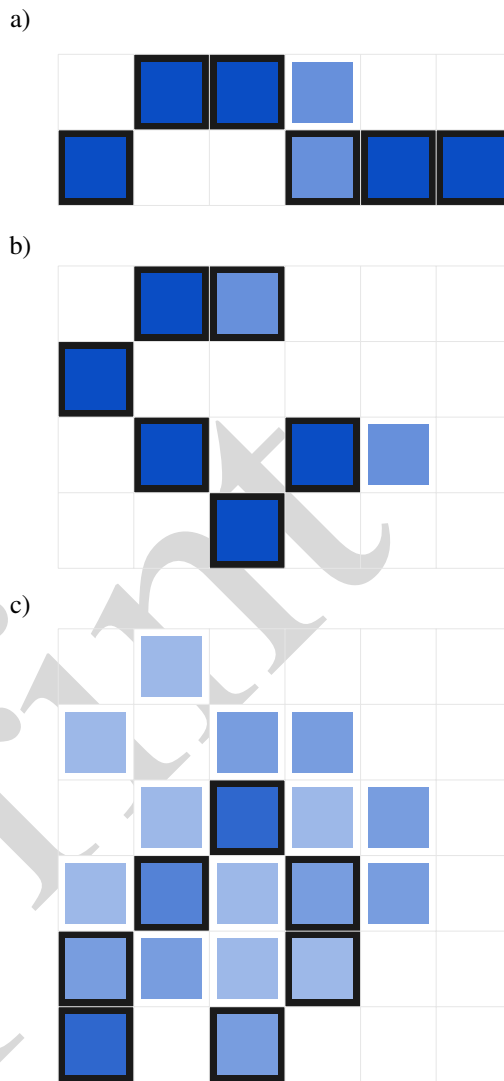


Fig. 4. Black bordered cells denote the cells actually occupied by the shape of the chain. The shaded square denotes the probability that the cell is part of the sensed configuration. Darker shading denotes higher probability. In a) and b) there is only one agent that could be located at two different positions, whereas in c) many more shapes are possible.

The missing information can be visualized and for most cases the ambiguity arises when symmetry is broken. This can be explained that under all possible shapes, chains with more symmetry do need less information about the order of the sensor inputs. See Fig. 4 for three example chains and the distributions of the reconstructed configurations.

Given that we can assign which sensor readings correspond to which sensors we get an ordering of the chain. The information the collective knows about a feature is now $I(F; S^{(k)})$. This suffices to fully determine the configuration for up to nine agents (see Table I).

B. Local Sensors and Local Communication

So far we have only considered together the information that the sensors of all agents contain. In the next step we investi-

gated what each agent individually acquires about the whole configuration and its features. For that we computed the mutual information between S' and the features $X^{(k)}, L^{(k)}, D^{(k)}$, where S' denotes as introduced earlier the random variable of a sensor reading that cannot assigned to a specific agent.

To compare the mutual information for different sizes of chains we used the *uncertainty coefficient* of the features, which is defined as

$$C_{S,F} = I(S;F)/H(F).$$

The uncertainty coefficient is a measure of relative information and allows comparison of the amount of knowledge for different sizes of chains. As expected the knowledge of each local sensor decreases the larger the chains get, this is also true for the detection of loops. While the whole collective can always detect whether it forms a loop or not, a single agent gets worse in determining this feature the more agents there are (see Table IV).

If each agent cannot only see its own sensor input but also the sensor input of its direct neighbours the situation is improved. The sensor values of the neighbour are denoted by the random variable \bar{S}' . The knowledge about the configuration is clearly higher than before when only information of one agent was available, but it is not three times as much. This can be attributed to the fact that the agents at the beginning and the end of the chain only get information from one neighbour and that certain redundancy effects come into play. However, the information of whether the chain is a loop increases by a factor larger than three (see Table III). In Figure 5 the sensor reading of an agent and its two neighbours for a loop of seven agents are shown with the corresponding distributions of configurations.

If however, we can assign a sensor reading to a specific agent but still do not know where a chain is possibly split, that is, we do not know whether an agent will always be at the same position in a chain, the situation does not change and we get the same results for the amount of information they provide regarding the features as for the case of a single unidentified agent S' .

If however we know that the agents are always ordered in the same way and we can identify which sensor readings correspond to which sensors and we can enumerate the agents such that the random variable S_i for $i = 1, \dots, k$ denotes the i -th sensor reading in a chain. If the agents form a loop the enumeration starts at an arbitrary agent, otherwise we assume that the beginning of the chain is denoted by S_1 . In this scenario we can observe that the agents at the beginning and the end of the chain, i.e. the sensor readings S_1 and S_{k-1} , individually provide all information about the feature $L^{(k)}$, that is $I(S_1; L^{(k)}) = I(S_{k-1}; L^{(k)}) = H(L^{(k)})$.

C. Chains of Variable Length

For all computations the number of agents was fixed. In the case that the agents have access to all sensor inputs the results would not change if we allow the length of chain to be variable. That is because the collection of all sensor inputs

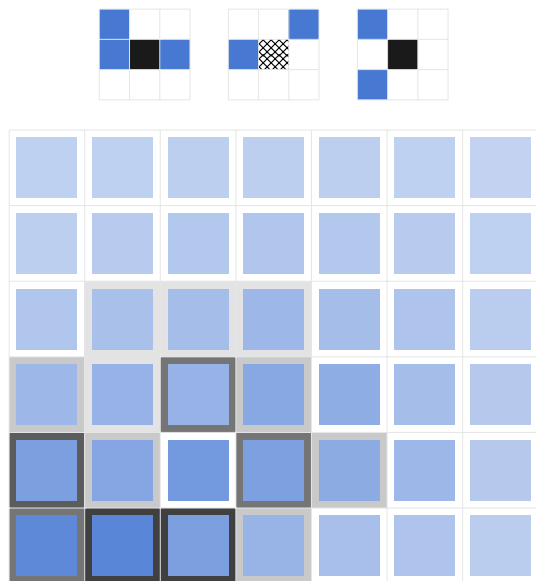


Fig. 5. The top row shows sensor readings of three consecutive agents in a chain of seven agents. Below the distribution of shapes given the sensor reading(s) is shown. The shaded squares denote the probability of a cell being occupied by any agent of the chain given only the sensor reading of the crosshatched agent (centred). The shaded borders denote the probability of a cell being occupied by any agent of the chain given that all three sensor readings are known. Darker colours denote higher probabilities.

already implicitly contains the information about the number of agents. But for local sensing varying chain length can have an impact as an agent locally cannot sense of how many other agents the chain consists.

Therefore we computed the information in these cases where the number of agents in the chain can vary. The random variable of all chains up to k agents is denoted $Y^{(k)}$. The chains are distributed such that the number of agents is uniformly distributed and given the size of a chain the chains of a specific length are also uniformly distributed. This reflects the assumption that there is neither prior knowledge about a preferred configuration nor a preferred size of the chain.

For all computed values the relative knowledge about the configuration was less than the average relative knowledge of the cases where the number of agents was known, that is,

$$C_{S',Y^{(k)}} \leq \frac{1}{k} \sum_{j=1}^k C_{S',X^j}.$$

This holds for the diameter $D^{(k)}$ and the loop condition $L^{(k)}$ as well. Moreover it holds also when the agents communicate with their direct neighbours. That means that allowing chains of variable length seems to generally increase the average uncertainty about the selected features.

VI. DISCUSSION

The approach we used to investigate how information of a global feature is provided on a local level is not limited to chains of agents in grid worlds. Here however, we focused on a specific case where the environment does not provide any

information via gradients or stigmergy, instead the local sensors can only see local neighbours and their sensor readings. We also ignored any dynamics of the collective and considered only static configurations.

We observed that on a local scale neighbourhoods of agents provided more information about a global feature than the individual agents did. That means, similar to a system of many bodies in physics, we can observe informational interaction on different levels. Also we have seen that in certain configurations individual agents are capable of completely determining a global feature. For example the loop feature where a single agent can acquire certainty about the feature if it is located at the beginning or end of the chain. Here the amount of information the first and last sensor provide about the loop feature increased when an enumeration that is consistent over all chains is given. Even without this knowledge, given that any sensor provides the reading that an agent has only one neighbour, this specific agent has locally certainty that the chain does not form a loop. At the moment our model cannot capture this property and further investigations need to address this.

Furthermore, assuming that organisms are informationally parsimonious, the result that symmetries are informationally easier to process might give an indication that there exists a certain evolutionary pressure for symmetry in living systems.

For some formation processes all information that is needed may be locally available, but one can not assume that this is the general case. Global information is often used in the process of morphogenesis, for example gradients are providing global information on a local scale. In general there are different ways of distributing information among a collective, spanning from direct communication to stigmergy, morphogens, hormones etc. However, in terms of information theory all these methods establish communication channels among the collective, sometimes incorporating the environment as a intermediate step.

We expect that future research will guide how efficient communication models for multi-agent systems can be built by stating constraints and requirements on the actual mechanism. These investigations can also lead to insights on why certain information transfer methods are used in nature. In particular the information theoretic perspective is not limited to investigate static “snapshots” of a system but can also capture the information processed in the dynamics of a collective. Thus we may be able to get a better understanding of possible mechanisms for morphogenesis or even of the fundamental principles of pattern formation.

VII. CONCLUSION

We used information theory to analyse how global information is provided by local sensors. One of the key observations is that knowing the ordering of sensor values in a chain of agents gives additional information about the global shape. As mentioned before, the additional information helps to determine where symmetries are broken in the shape of the chain. Referring to the caricature of the blind men and the elephant, one might say that it leads to a better idea what an

TABLE I
INFORMATION OF THE CONFIGURATION CAPTURED BY THE SENSORS

k	$H(X^{(k)})$	$I(X^{(k)}; \tilde{S}^{(k)})$	$I(X^{(k)}; S^{(k)})$
2	2.0000	2.000	2.000
3	4.3219	4.3219	4.3219
4	6.5392	6.5392	6.5392
5	8.7142	8.6571	8.7142
6	10.8734	10.6639	10.8734
7	13.0300	12.5793	13.0300
8	15.1831	14.3992	15.1831
9	17.3322	16.1278	17.3322

TABLE II
LOOPS AND DIAMETER CAPTURED BY THE UNORDERED SENSORS

k	$H(D^{(k)})$	$I(D^{(k)}; \tilde{S}^{(k)})$	$H(L^{(k)})$	$I(L^{(k)}; \tilde{S}^{(k)})$
2	0.0000	0.0000	0.0000	0.0000
3	0.7219	0.7219	0.0000	0.0000
4	0.9899	0.9899	0.0857	0.0857
5	1.3421	1.3421	0.0776	0.0776
6	1.5820	1.5564	0.0397	0.0397
7	1.7926	1.7301	0.0200	0.0200
8	1.9792	1.8498	0.0135	0.0135
9	2.1477	1.9298	0.0104	0.0104

elephant is like, when it is known that the trunk is not located between the legs and the tail.

In the case where we cannot assign a sensor reading to a particular sensor we observed that agents which locally share sensor readings provide more information about the “loopiness” feature than the sum of information provided by individual sensors. This effect of local “interaction” is getting stronger when the size of the chain is increased, as the information provided individually by agents decreases faster with the number of agents than the amount of information that is provided by sharing information with local neighbours.

Furthermore we were able to see that the first and last agent of the chain provide all information about whether the chain forms a loop or not when the chain is always enumerated in the same way. The case of having a beginning and an end characterising the absence or presence of the loop feature is obvious. However we can see that information theory allows us to go beyond that in that it identifies features that have the property to be detectable on a local scale.

We observed that allowing the number of agents in a chain to vary did not yield any significant changes in the observations except for some extra information required to determine the number of agents.

In summary, we have seen that we can use information theory to characterise how much regarding global topological and morphological features can be detected by local sensors thus is a first step in understanding how information is processed and distributed in morphogenesis.

TABLE III
INFORMATION CAPTURED BY LOCAL SENSORS

k	$I(X^{(k)}; S')$	$I(X^{(k)}; S', \bar{S}')$	$I(L^{(k)}; S')$	$I(L^{(k)}; S', \bar{S}')$
1	0.0000	0.0000	0.0000	0.0000
2	2.0000	2.0000	0.0000	0.0000
3	3.2310	4.3219	0.0000	0.0000
4	3.3881	5.7370	0.0370	0.0857
5	3.2947	6.5486	0.0249	0.0776
6	3.1339	6.7964	0.0087	0.0341
7	2.9720	6.8217	0.0022	0.0108
8	2.8227	6.7797	0.0008	0.0045
9	2.6856	6.7073	0.0004	0.0024

TABLE IV
NORMALIZED INFORMATION CAPTURED BY LOCAL SENSORS

k	$C_{X^{(k)}, S'}$	$C_{X^{(k)}, (S', \bar{S}')}$	$C_{L^{(k)}, S'}$	$C_{L^{(k)}, (S', \bar{S}')}$
1	1.0000	1.0000	1.0000	1.0000
2	1.0000	1.0000	1.0000	1.0000
3	0.7476	1.0000	1.0000	1.0000
4	0.5181	0.8773	0.4317	1.0000
5	0.3781	0.7515	0.3209	1.0000
6	0.2882	0.6250	0.2191	0.8589
7	0.2281	0.5235	0.1100	0.5400
8	0.1859	0.4465	0.0593	0.3333
9	0.1549	0.3870	0.0385	0.2308

REFERENCES

- [1] A. M. Turing, "The Chemical Basis of Morphogenesis," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 237, no. 641, pp. 37–72, Aug. 1952. [Online]. Available: <http://rstb.royalsocietypublishing.org/cgi/doi/10.1098/rstb.1952.0012>
- [2] L. G. Harrison, "Kinetic theory of living pattern." *Endeavour*, vol. 18, no. 4, pp. 130–136, 1994.
- [3] A. Gierer and H. Meinhardt, "A theory of biological pattern formation," *Biological Cybernetics*, 1972. [Online]. Available: <http://www.springerlink.com/index/T541854V22417131.pdf>
- [4] H. Meinhardt, *Models of biological pattern formation*. Academic Press London, 1982. [Online]. Available: http://www.eb.tuebingen.mpg.de/departments/former-departments/h-meinhardt/kurs/a_summary.pdf
- [5] —, "Primary body axes of vertebrates: generation of a near-Cartesian coordinate system and the role of Spemann-type organizer." *Developmental Dynamics*, vol. 235, no. 11, pp. 2907–2919, 2006. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/16958119>
- [6] A. L. Christensen, R. OGrady, and M. Dorigo, "SWARMORPH-script: a language for arbitrary morphology generation in self-assembling robots," *Swarm Intelligence*, vol. 2, no. 2, pp. 143–165, Jul. 2008. [Online]. Available: <http://www.springerlink.com/index/10.1007/s11721-008-0012-6>
- [7] E. Sahin, T. Labella, V. Trianni, J.-L. Deneubourg, P. Rasse, D. Floreano, L. Gambardella, F. Mondada, S. Nolfi, and M. Dorigo, "SWARM-BOT: pattern formation in a swarm of self-assembling mobile robots," *IEEE International Conference on Systems, Man and Cybernetics*, vol. 4, p. 6, 2003. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1173259>
- [8] M. D. Rosa, S. C. Goldstein, P. Lee, J. D. Campbell, and P. Pillai, "Programming modular robots with locally distributed predicates," in *Proceedings of the IEEE ICRA*, 2008, pp. 3156–3162.
- [9] E. Bonabeau, "From classical models of morphogenesis to agent-based models of pattern formation." *Artificial life*, vol. 3, no. 3, pp. 191–211, Jan. 1997. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/9385734>
- [10] J. Jones, "Characteristics of pattern formation and evolution in approximations of physarum transport networks." *Artificial Life*, vol. 16, no. 2, pp. 127–53, Jan. 2010. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/20067403>
- [11] J. Glazier and F. Graner, "Simulation of the differential adhesion driven rearrangement of biological cells," *Physical Review E*, 1993. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevE.47.2128>
- [12] P. Hogeweg, "Shapes in the shadow: evolutionary dynamics of morphogenesis," *Artificial Life*, vol. 6, no. 1, pp. 85–101, Jan. 2000. [Online]. Available: <http://www.mitpressjournals.org/doi/abs/10.1162/106454600568339>
- [13] G. Theraulaz and E. Bonabeau, "Modelling the Collective Building of Complex Architectures in Social Insects with Lattice Swarms," *Journal of Theoretical Biology*, vol. 177, no. 4, pp. 381–400, Dec. 1995. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S002251938570255X>
- [14] L. Wolpert, "Positional information and the spatial pattern of cellular differentiation," *Journal of Theoretical Biology*, vol. 25, no. 1, p. 46, 1969. [Online]. Available: [http://dx.doi.org/10.1016/S0022-5193\(69\)80016-0](http://dx.doi.org/10.1016/S0022-5193(69)80016-0)
- [15] J. Miller, "Evolving a self-repairing, self-regulating, French flag organism," *Genetic and Evolutionary Computation — GECCO*, Jan. 2004. [Online]. Available: <http://www.springerlink.com/index/U7EX2REYCALMJU46.pdf>
- [16] J. Knabe, M. Schilstra, and C. Nehaniv, "Evolution and morphogenesis of differentiated multicellular organisms: autonomously generated diffusion gradients for positional information." in *Artificial Life XI*, vol. 11, Jan. 2008, p. 321. [Online]. Available: <http://panmental.de/papers/FlagPottsGRNALife11.pdf>
- [17] R. Balzer, "An 8-state minimal time solution to the firing squad synchronization problem," *Information and Control*, vol. 10, no. 1, pp. 22–42, 1967. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0019995867900320>
- [18] A. S. Klyubin, D. Polani, and C. L. Nehaniv, "Tracking Information Flow through the Environment: Simple Cases of Stigmergy," in *Artificial Life IX*, J. Pollack, M. Bedau, P. Husbands, T. Ikegami, and R. A. Watson, Eds. Cambridge, MA: MIT Press, 2004, pp. 563–568.
- [19] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, vol. 27, no. October, pp. 379–423, Aug. 1948.
- [20] H. Touchette and S. Lloyd, "Information-theoretic approach to the study of control systems," *Physica A*, vol. 331, p. 140, 2004. [Online]. Available: [doi:10.1016/j.physa.2003.09.007](https://doi.org/10.1016/j.physa.2003.09.007)
- [21] R. Linsker, "Self-Organization in a Perceptual Network," *Computer*, vol. 21, no. 3, pp. 105–117, 1988.
- [22] K. Zahedi, N. Ay, and R. Der, "Higher coordination with less control - A result of information maximisation in the sensori-motor loop," *Adaptive Behavior*, vol. 18, no. 3-4, p. 338, 2010. [Online]. Available: <http://arxiv.org/abs/0910.2039>
- [23] A. S. Klyubin, D. Polani, and C. Nehaniv, "Keep your options open: an information-based driving principle for sensorimotor systems," *PLoS ONE*, vol. 3, no. 12, p. 4018, Jan. 2008. [Online]. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2607028>
- [24] M. Möller and D. Polani, "Common Concepts in Agent Groups, Symmetries and Conformity in a Simple Environment," in *Artificial Life XI*, S. Bullock, J. Noble, R. Watson, and M. A. Bedau, Eds. MIT Press, Cambridge, MA, 2008, pp. 420–427.
- [25] T. M. Cover and J. A. Thomas, *Elements of Information Theory 2nd Edition*, ser. Wiley Series in Telecommunications and Signal Processing. Wiley-Interscience, Jul. 2006. [Online]. Available: <http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0471241954>