

# A Novel Information Measure for Predictive Learning in a Social System Setting

Paolo Di Prodi<sup>1,\*</sup>, Bernd Porr<sup>2</sup>, and Florentin Wörgötter<sup>3</sup>

<sup>1</sup> University of Glasgow  
epokh@elec.gla.ac.uk

<sup>2</sup> University of Glasgow  
b.porr@elec.gla.ac.uk

<sup>3</sup> BCCN Göttingen, Germany  
worgott@bccn-goettingen.de

**Abstract.** We introduce a new theoretical framework, based on Shannon's communication theory and on Ashby's law of requisite variety, suitable for artificial agents using predictive learning. The framework quantifies the performance constraints of a predictive adaptive controller as a function of its learning stage. In addition, we formulate a practical measure, based on information flow, that can be applied to adaptive controllers which use hebbian learning, input correlation learning (ICO/ISO) and temporal difference learning. The framework is also useful in quantifying the social division of tasks in a social group of honest, cooperative food foraging, communicating agents.

Simulations are in accordance with Luhmann, who suggested that adaptive agents self-organise by reducing the amount of sensory information or, equivalently, reducing the complexity of the perceived environment from the agents perspective.

## 1 Introduction

Information measures are usually defined for input/output systems where they determine the quality of the transmission. Behaving agents, however, act as closed loop systems in which there is no clearly defined difference between input and output. What matters most for the organism is to compensate for disturbances introduced by the environment in the perception action loop. If there is no disturbance, the organism cannot differentiate between themselves and the environment. Consequently, the concept of information in these systems needs to be revised [5].

A method for defining closed loop information has been proposed by Ashby - the so called *requisite variety* [1]. The measure is based on the premise that closed loop systems aim to maintain a desired state. The goal of a feedback loop is then to minimise the deviation from the desired state i.e. the number of bits required to successfully compensate a disturbance acting on the forward

---

\* Webpage: <http://isg.elec.gla.ac.uk>.

loop. In this way, the method quantifies the variety, or bits, originating from the disturbance. For example, if the disturbance has a variety of 10 bits and survival requires a desired state of 2 bits, then the reaction to that disturbance must provide a variety of 8 bits. Ashby then proved that error controlled closed loop systems (like PID controllers [21]) cannot achieve perfect regulation. More recently, Touchette et al. [24] in Theorem 10 proved that the entropy reduction achieved by a closed loop system is bounded by the entropy reduction achieved by the open loop control plus the mutual information gathered by the estimation of the state. However the advent of predictive controllers, such as Q-learning [22], that predict future states requires an extension of the information theory for predictive learning.

In this paper we present an extension to the law of requisite variety, called *the predictive requisite variety*, that quantifies the theoretical limits of control (as well as providing a performance index) for predictive adaptive controllers. We argue that a predictive adaptive controller acts as a reactive system before learning and as an open loop forward system after learning. A reactive system comprises an error controlled closed loop and is non optimal because it only reacts after a deviation from its desired state has happened. The environment contains usually predictive signals which can help the agent to react before the error is presented [16]. Thus, bio inspired controllers can be provided with a predictive signal (like vision) and a reflexive signal (like touch). Learning then has the task of avoiding the trigger of the reflexive reaction - thus creating an open loop forward controller which discards the information of the reflexive signal.

Learning is then quantified by the increase in the information flow of the predictive loop and by a corresponding decrease in the information flow of the closed loop. Information flow, or transfer entropy, is not a new idea (see for example [3,23]) but it has never been applied to predictive agents in order to assess their learning performance. The analysis of a predictive agent with a single behaviour, say for example obstacle avoidance, can be done calculating the information flow of the sensory-motor loop.

Analysis becomes more complicated when an agent is provided with a set of competitive behaviours in a social scenario where agents use predictive learning-see, for example, ISO[4] or ICO[17,4] - and are therefore learning from each other. The task of the social system in this analysis is cooperative food foraging in which every agent has 3 adaptive behaviours which are: avoidance for obstacles, attraction to food disks and attraction to others with food. Agents communicate honestly, always signalling to others when they find food. When the social system is adapting, it self-organises into 2 sub-systems each described by a dominant behaviour: seekers have a dominant attraction for food disks, parasites have a dominant attraction to others with food. The information flow explains how the social system divides itself into sub-systems by looking at the information processing of every agent. Luhmann [13] proposed that differentiation of social systems is caused by a decrease in information processing of each subsystem and this is consistent with our information flow measurements.

The paper is divided in sections covering the following topics: regulation and entropy (as defined originally by Ashby), a new information measure for predictive learning, a simulation model with social adaptive agents, results, and a discussion.

## 2 Ashby’s Law of Requisite Variety

First, we review Ashby Law of Requisite Variety for the forward (see Fig.1(B)) and closed loop controller (see Fig.1(A)). Fig.1 uses the same notation introduced by Ashby:

- D= finite state machine whose states are the disturbances from the environment
- E= finite state machine whose states are the essential variables partitioned in  $E = \eta \cup \bar{\eta}$ , where  $\eta$  is a partition of desired states or goals of the organism and its complementary partition  $\bar{\eta}$  represents the non-desired states.
- R= finite state machine whose states are the available regulations/actions that the organism can perform
- T= finite state machine whose states are the set of possible states of the environment

In this work we consider deterministic finite state machines but the analysis can also be extended to Markov processes [6]. It is very important for our analysis to understand that only the forward controller can achieve perfect regulation whereas the closed loop controller cannot because the reflex always comes too late. Gatsby [1] stated that a good controller  $R$  blocks the flow of variety<sup>1</sup> from disturbances  $D$  to essential variables  $E$ : if R is a regulator, the insertion of R between D and E decreases the variety that is transmitted from D to E. An organism can be described by a body  $R$  with goals to be achieved  $\eta$  and an environment  $T$  which forms a closed loop between actions and sensors. As an analogy, the organism is a perfect regulator if is able to keep the essential variables E within a desired sub-set  $\eta$  in spite of the disturbances D -thus having a null entropy for E,  $H(E) = 0$ .

If no regulator  $R$  is provided (see Fig.1(C)), the disturbance D tends to drive  $E_0$  outside a set of desired states  $\eta$  by means of the environment  $T$ , .Thus, in the worse case, the disturbance completely controls the status of the organism:

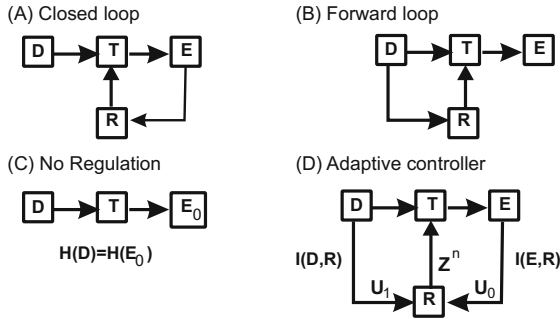
$$H(D) = H(E_0) \tag{1}$$

The regulator  $R$  can be connected in a feed-forward configuration as in Fig.1(B) or in a closed loop configuration as in Fig.1(A). The performance of the forward regulator is measured by the maximum entropy reduction  $\Delta H_{forward}^{max}$  which is the difference between the entropy of the essential variable  $H(E_0)$  before regulation and after regulation  $H(E)$ .

$$\Delta H_{forward}^{max} = H(E_0) - \min H(E) \tag{2}$$

---

<sup>1</sup> Ashby defines variety precisely as the number of different states a variable can take and is equivalent to the Shannon’s entropy  $H$  measured in bits.



**Fig. 1.** (A) The organism with a closed loop controller. (B) The same organism with an forward controller. (C) The organism before regulation. (D) An adaptive controller is a mix of forward and closed loop control. Every block is a finite state machine whose inputs are indicated by incoming arrows and outputs are indicated by outgoing arrows.

The maximum entropy reduction in the forward condition  $\Delta H_{forward}^{max}$  can be calculated by using the Law of Requisite Variety:

$$H(E) \geq H(D) + H(R|D) - H(R) \tag{3}$$

where  $H(R|D)$  is the regulator noise<sup>2</sup>. Thus:

$$\Delta H_{forward}^{max} = H(R) - H(R|D) \tag{4}$$

because combining Eq.2 and Eq.3 gives:

$$\Delta H_{forward}^{max} = H(E_0) - H(D) - H(R|D) + H(R) \tag{5}$$

Considering the initial condition in Eq.1 we obtain Eq.4:

$$\Delta H_{forward}^{max} = H(D) - H(D) - H(R|D) + H(R) = H(R) - H(R|D) \tag{6}$$

The quantity  $\Delta H_{forward}^{max}$  in Eq.4 tells us that better performance can be achieved by either increasing the regulation entropy  $H(R)$  or by decreasing the controller noise  $H(R|D)$ .

We will now show that a closed loop controller cannot achieve perfect regulation ( $H(E) = 0$ ) as it requires a deviation from the desired state  $\eta$  to work  $H(E) > 0$ . Thus, the disturbance transmits all its entropy to the essential variable  $H(D) = H(E)$  and no entropy reduction can be achieved:

$$\Delta H_{close}^{max} = 0 \tag{7}$$

If for  $H(E) = 0$  then  $R$  blocks the information flow in the channel  $D \rightarrow E$  and thus no information is transmitted to  $R$  for the regulation task: the regulator  $R$  is asserting a perfect control on  $E$  without knowing the status. In the next section we extend the law of requisite variety for adaptive controllers.

<sup>2</sup> If the controller is not noisy  $H(R|D) = 0$ .

### 3 Law of Adaptive Requisite Variety

An adaptive controller (see Fig.1(D)) is a mix of a forward [8] and closed loop controllers [21] because  $R$  has now 2 inputs:  $D$  and  $E$ . We can think of  $D$  as a predictor of the deviation of  $E$ , because  $D$  transfers its entropy to  $E$  by means of the environment  $T$ .

In order to explain the new law, we introduce the mutual information  $I(E, R)$  for the closed loop channel  $E \rightarrow R$  with the corresponding channel capacity  $C_{E,R}$ :

$$I(E, R) = H(E) + H(R) - H(E, R) \tag{8}$$

$$C_{E,R} = \max_{p(E)} I(E, R) \tag{9}$$

the mutual information  $I(D, R)$  for the forward channel  $D \rightarrow R$  with the corresponding channel capacity  $C_{D,R}$ :

$$I(D, R) = H(D) + H(R) - H(D, R) \tag{10}$$

$$C_{D,R} = \max_{p(D)} I(D, R) \tag{11}$$

The channel capacity of the regulator channel  $D \rightarrow T$  is then  $C_{R,T}$ .

The adaptive controller (denoted *ada*) begins as a closed loop controller with  $\Delta H_{ada}^{max}(before) = H_{close}^{max}$  (see Eq.7) as it mainly uses the  $E \rightarrow R$  reflex channel and blocks the  $D \rightarrow R$  predictor channel whose mutual information is very low. In summary:

$$0 < I(E, R) \leq C_{E,R} \tag{12}$$

$$I(D, R) \simeq 0 \tag{13}$$

$$\Delta H_{ada}^{max}(before) = 0 \tag{14}$$

The adaptive controller achieves perfect regulation (see Eq.4) when

$$\Delta H_{ada}^{max}(after) = H_{forward}^{max} \tag{15}$$

because it blocks the  $E \rightarrow R$  reflex channel and opens the  $D \rightarrow R$  predictor channel. To summarise:

$$0 < I(D, R) \leq C_{D,R} \tag{16}$$

$$I(E, R) \simeq 0 \tag{17}$$

$$\Delta H_{ada}^{max}(after) = H(R) - H(R|D) \tag{18}$$

If we assume realistically that the regulator has a common channel capacity  $C_{E,R} = C_{D,R} = C_{R,T}$ , the constraint for learning becomes:

$$I(E, R) + I(D, R) \leq C_{R,T} \tag{19}$$

thus an adaptive controller can achieve optimal regulation  $\Delta H_{ada}^{max}(after)$  when is able to compensate the mutual information of the closed loop  $I(E, R)$  with

the mutual information of the forward controller  $I(D, R)$ . An imperfect regulator will likely work in the sub-optimal regime  $I(D, R) < I(E, R)$ . So to quantify the performance of an adaptive predictive controller we have to compute the mutual informations  $I(D, R)$  and  $I(E, R)$ . This is however not always possible because it is hard to identify the reflex channel and the predictor channel. Therefore in the next section we use an approximation of these 2 quantities using the concept of information flow.

### 4 Information Flow for Adaptive Predictive Controllers

Looking at Fig.1(D), we can estimate  $I(E, R)$  by computing the information flow of the reflex-output channel  $Z^n \rightarrow U_0$  and  $I(D, R)$  by computing the information flow of the predictive-output channel  $Z^n \rightarrow U_1$ . We denoted them as:

$$MI_{U_0}^n = I(Z^n, U_0) \leftrightarrow I(E, R) \tag{20}$$

$$MI_{U_1}^n = I(Z^n, U_1) \leftrightarrow I(D, R) \tag{21}$$

where  $U_0$  is the reflex input,  $U_1$  is the predictor input and  $Z^n$  the extended output:

$$Z^n = [z(k)z(k + 1) \dots z(k + n - 1)] \tag{22}$$

which contains  $n$  outputs of the agent and  $U$  the random variable describing the temporal signal  $u(k + n)$  which is the input of the agent resulting from previous actions(for more details see [11,14]). Fig.2(A) shows an organism composed of 3 ICO [17] controllers and the corresponding information flow measures for every controller. Each ICO controller takes 2 continuous inputs  $U_0, U_1$  and one continuous output  $Z_n$ . ICO correlates the predictive signal  $u_1$ <sup>3</sup> with the derivative of the reflexive signal  $u_0$  according to the formula:

$$\frac{d\omega_1}{dt} = \mu \cdot u_1 \cdot \frac{du_0}{dt} \tag{23}$$

where  $\omega_1$  is the gain of the predictive signal  $u_1$  and  $\mu$  is the learning speed (see Fig.2(C)). Since the ICO controller works in continuous mode, the input and output signals must be discretized in order to compute the information flow and channel capacity (see Simulation Details). The two measures  $MI_{U_0}^n, MI_{U_1}^n$  are used to compute the channel capacities  $C_{E,R}$  and  $C_{D,R}$ :

$$\zeta^n(Z^n \rightarrow U_0) = \max_{p(Z^n)} MI_{U_0}^n \leftrightarrow C_{E,R} \tag{24}$$

$$\zeta^n(Z^n \rightarrow U_1) = \max_{p(Z^n)} MI_{U_1}^n \leftrightarrow C_{D,R} \tag{25}$$

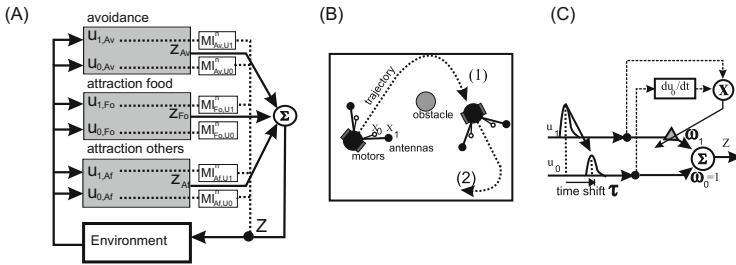
In the simulations in the next section, we will estimate the mentioned quantities for individual agents of a social group.

---

<sup>3</sup>  $u_1$  and  $u_0$  indicates temporal signals  $u_1(t)$  and  $u_0(t)$ .

## 5 Methods

The previous measures are applied to a social system where all agents learn continuously from each other and from the environment. This scenario is very interesting because the social system is able to self-organise by forming 2 sub-systems with task division. The social system described in [15] is composed of  $N$  identical agents and  $M$  food disks randomly placed in a square world for every simulation. Food disks contain a certain amount of food that is depleted when an agent finds it. The task is cooperative food foraging. The simulated agent is shown in Fig.2(B) and has also been used by [12]: it is a Braitenberg [7] vehicle with 2 lateral wheels and 2 antennas. By default the agent drives straight forward, with speed  $v = 1$  units per time step. It has 2 sensor-pairs, near contact antennas and far contact antennas. Every agent has a MISO (multiple inputs single output) controller and a variable of 1 bit for the food status. The agent has competitive 3 tasks: avoid obstacles (empty food disks and other agents without food), find food from the disks, find foods from other agents with food. The MISO is composed of 3 parallel ICO controllers (see Fig.2(A)) which are provided with a reflex input error  $u_0$ , a predictive signal error  $u_1$ , a learnt weight  $\omega_1$  and an output  $z$ . The outputs of the 3 ICO controllers are summed to  $z = z_{Av} + z_{Fo} + z_{Af}$ <sup>4</sup> which gives the steering angle:  $z = 0$  the robot goes straight forward at speed  $v$ ,  $z > 0$  the robot rotates clockwise,  $z < 0$  the robot rotates anti-clockwise. Every



**Fig. 2.** (A) MISO controller composed of 3 stacked ICO controllers for avoidance, food attraction and attraction to others. The output of every controller is summed to  $z$ . For every controller/behaviour the pair of mutual information is computed between the output and the input  $MI_{U_0}^i, MI_{U_1}^i$ . (B) Agent with short antennas (reflexive inputs,  $x_0$ ) and long antennas (predictive inputs,  $x_1$ ). The agent is learning to avoid obstacles. The motor reaction will reduce the intensity of the painful reflex  $x_0$  as well as delay its occurrence. (C) Schematic diagram of the input correlation learning rule and the signal structure [17]. The  $u_0$  and  $u_1$  are, respectively, the difference between the filtered values of the left and right antennas of the agent. During learning the  $u_0$  peak will be shifted in time and reduced in amplitude as the agent learn successfully by increasing the predictor gain  $\omega_1$ .

<sup>4</sup> Av stands for obstacle avoidance, Fo for food attraction and Af for attraction to others with food.

simulation is run for  $0 \leq k \leq 6 \cdot 10^5$  time steps and is divided in 3 stages. At every stage, each agent produces 6 input time series and 1 output time series  $z(k)$  which means that we can calculate the information flow for every pair of reflex-output and predictor-output:  $MI_{U_0}^n, MI_{U_1}^n$ . For a single simulation:

1. for  $0 \leq k_1 \leq 2 \cdot 10^5$  all agents are reactive ( $\mu = 0$ ). For each agent  $i = 1, \dots, N$  we have 3 pairs of information flow:
  - (a) avoidance:  $MI_{Av,U_1}^n, MI_{Av,U_0}^n$
  - (b) food attraction:  $MI_{Fo,U_1}^n, MI_{Fo,U_0}^n$
  - (c) others attraction:  $MI_{Af,U_1}^n, MI_{Af,U_0}^n$
2. for  $2 \cdot 10^5 < k \leq 4 \cdot 10^5$ : every agent is learning  $\mu = 1.0$  and the weight for every ICO controller  $\omega_{1,Av}, \omega_{1,Fo}, \omega_{1,Af}$  is increasing.
3. for  $4 \cdot 10^5 < k_3 \leq 6 \cdot 10^5$ : every agent stop learning  $\mu = 0.0$  and is using the last weight set at  $k = 4 \cdot 10^5$ . For each agent we compute again the 3 pairs of the  $MI^n$ .

The channel capacities for every agent are computed by providing each isolated output  $z = z_{Av}, z = z_{Fo}, z = z_{Af}$  with a source of independent randomness during a simulation of  $2 \cdot 10^5$  time steps for every case. Then we apply the Blahut-Arimoto algorithm [20,18] with a bound error of  $\varepsilon = 10^{-11}$  and 5000 maximum iterations to estimate the channel capacity for every agent in the reflex-output loop  $\zeta^n(Z_k^n \rightarrow U_0)$ . There is no difference between  $\zeta^n(Z_k^n \rightarrow U_0)$  of every agent so we define  $\zeta_{all}^n$ . To compute the capacity for the predictor-output loop  $\zeta^n(Z_k^n \rightarrow U_1)$  we use the same approach but preset the weights of every agent to an arbitrary high value to simulate perfect learning:

$$\omega_{1,Av} = 10.0, \omega_{1,Fo} = 10.0, \omega_{1,Af} = 10.0 \quad (26)$$

and we obtain the same results

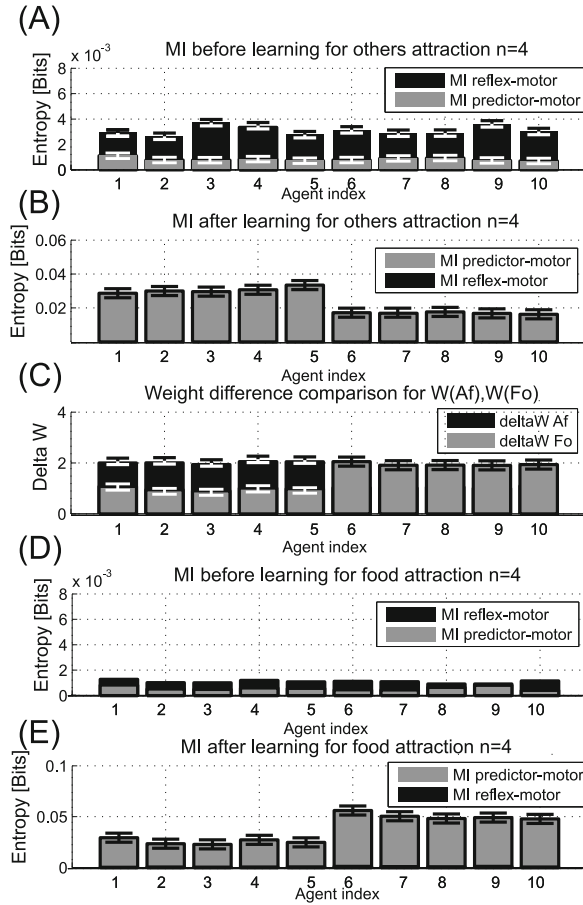
$$\zeta^n(Z_k^n \rightarrow U_1) = \zeta^n(Z_k^n \rightarrow U_0) = 2.0 \quad (27)$$

for  $n \geq 2$  as anticipated in Eq.24,25.

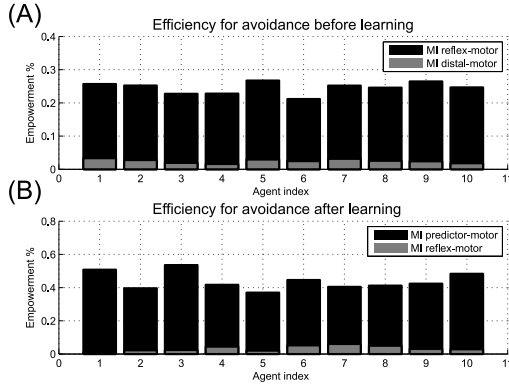
## 6 Results

The results of this sections are based on a simulation with  $N = 10$  agents and  $M = 5$  food disks. All agents start with the same weights for every ICO controller  $\omega_{1,Av} = 0.1, \omega_{1,Fo} = 0.1, \omega_{1,Af} = 0.1$ . In stage 3 there are 5 agents with  $\omega_{1,Af} < \omega_{1,Fo}$  and 5 agents with  $\omega_{1,Af} > \omega_{1,Fo}$ . The first group is identified by a strong attractive behaviour for the food disks (seekers), whereas the second group is identified by a strong attractive behaviour for others agent with food (parasites).

We estimate the  $MI^4$  in stage 1 and stage 3 for every agent by using the corrected standard deviation formula [19]. Before learning (Fig.3 (A),(D)) the



**Fig. 3.** (A) Information flow before learning for attraction to others  $MI_{Af,U1}^4$  (grey bars),  $MI_{Af,U0}^4$  (black bars) expressed in bits. (B) Information flow after learning for attraction to others in bits. (C) Weight difference for every agent:  $\Delta W_{Af} = \omega_{1,Af} - 0.1$ ,  $\Delta W_{Fo} = \omega_{1,Fo} - 0.1$  (D) Information flow before learning for attraction for food  $MI_{Fo,U1}^4$  (grey bars),  $MI_{Fo,U0}^4$  (black bars) in bits. (E) Information flow after learning for attraction for food in bits. Error bars are centered on the average for 100 simulations. The error width is equal to the maximum-minimum interval of the computed measures over 100 simulations.



**Fig. 4.** (A) Efficiency for every agent of the reflex-output and predictive-output loop in terms of capacity before learning (stage 1):  $MI_{Av,U0}^4/\zeta_{all}^4\%$  (dark bars),  $MI_{Av,U1}^4/\zeta_{all}^4\%$  (grey bars). (B) Efficiency after learning (stage 3).

reflex-output loop predominates over the predictor-output loop for both the food attraction behaviour and the others attraction behaviour:

$$MI_{Af,U1}^4 < MI_{Af,U0}^4 \simeq 0.0025 \tag{28}$$

$$MI_{Fo,U1}^4 < MI_{Fo,U0}^4 \simeq 0.001. \tag{29}$$

After learning (stage 3). The configuration is reverted and the predictor-output loop dominates the reflex-output loop for both behaviours as in Fig.3(B),(E):

$$MI_{Af,U0}^4 \ll MI_{Af,U1}^4 \tag{30}$$

$$MI_{Fo,U0}^4 \ll MI_{Fo,U1}^4 \tag{31}$$

This result matches our expectations in terms of the increase of  $I(D, R)$  and decrease of  $I(E, R)$ . If we compare the  $MI_{Af,U1}^4$  in Fig.3(B) to  $MI_{Fo,U1}^4$  in Fig.3(E) we can see that the agents with indices 1,2,3,4,5 (parasites) have a larger weight  $\Delta W_{Af} \simeq 2.0$  (see Fig.3(C)) for the attraction to others and, therefore, a larger information flow  $MI_{Af,U1}^4 > MI_{Fo,U1}^4$ , whereas agents with indices 6,7,8,9,10 (seekers) have a larger weight change  $\Delta W_{Fo} \simeq 2.0$  for the food attraction and so a bigger  $MI_{Fo,U1}^4 > MI_{Af,U1}^4$ .

Thus, the information measure is directly correlated with the weight change and can be used to quantify the learning performance of a single agent before and after learning. However, it can also be used to quantify the dominant behaviour and, consequently, the self-organising properties of social systems.

In Fig.4 we measure the efficiency of the reflex-output and predictive-output loop  $MI_{Av,U1}^4, MI_{Av,U0}^4$  for the avoidance behaviour in relation to the capacity for the agents  $\zeta_{all}^4 = 2.0$ . Fig.4(A) shows that before learning  $MI_{Av,U0}^4$  is using 0.25% of the full channel capacity and Fig.4(B) shows that after learning  $MI_{Av,U1}^4$  is using about 0.45% of the channel capacity. The  $MI$  of order

$n = 1, 2, 3$  does not provide enough discrimination for the previous analysis because the output history of the agent is too short to be correlated with the inputs. The capacity  $\zeta_{all}^n$  takes its maximum of 2 bits when  $n \geq 2$ .

## 7 Discussion

In summary, we have introduced an extension to Ashby's requisite variety theory called the law of adaptive requisite variety, computed the information flow to measure the learning performance for agents with competitive behaviours and found the relation between the efficiency of the information flow  $MI$  and the weight change of the adaptive controller  $\Delta\omega_1$ . We also linked our information approach to the Luhmann theory that sub-systems are formed to reduce the perceived complexity of the environment. In our simulations, after the learning experience 5 agents have a dominant attraction behaviour for food disks (seekers) and 5 have a dominant attraction behaviour for others (parasites). The seekers mainly use the predictive information of the food disks while the parasites mainly use the predictive information of the others who possess food. Thus, we conclude that predictive learning in a social context leads to the formation of subsystems. This can be demonstrated with the help of our approach. While Polani [11,9] and Lungarella [16,10] used the empowerment measure as a general cost function to optimise the agent's behaviour or evolution, we use it as the upper bound of the MI to measure the efficiency of the sensory-motor loop use. Ay in his work [2] uses an adaptive controller which maximises the excess entropy (the mutual information between past and present) at the input side to achieve a working regime exploratory and sensitive to the environment. We can calculate the MI for this case by considering the reflex as the present input and the predictor as the past history. Our approach is not restricted to MISO controllers. Kulvicius et al. [12] measures the temporal input development, the output and path entropy of the adaptive agents to study the optimality of the antenna ratio for an avoidance task, thus completing the tools required to evaluate a single task controller. Current work is focusing on using a model checking approach to verify the properties of the system in terms of information flow.

## 8 Simulation Details

The world is a toroidal square of  $300 \times 300$  units ( $Um$ ), the agent has a diameter of  $10 Um$ , the reflex antennas have a range of  $40 Um$ , the predictor antennas have a range of  $60 Um$ , every food disk has a diameter of  $20 Um$ , the agent consumes food after 30 time steps. Every food disk starts with 100 food units and, if depleted, is reset after 5 time steps. To compute the entropy, the input space is discretized into 4 equally spaced bins and normalised in the range  $[-1, 1]$  both for the predictor  $U_1$  and the reflex  $U_0$  signal, the output signal  $Z$  is discretized in 8 directions.

## References

1. Ashby, W.: *An Introduction to Cybernetics*. Chapman & Hall, Boca Raton (1956)
2. Ay, N., Bertschinger, N., Der, R., Güttler, F., Olbrich, E.: Predictive information and explorative behavior of autonomous robots. *The European Physical Journal B* 63(3), 329–339 (2008)
3. Polani, D., Ay, N.: Information flows in causal networks. *Adv. Compl. Syst.* (2007)
4. Wörgötter, F., Porr, B.: Isotropic sequence order learning in a closed loop behavioural system. In: *Roy. Soc. Phil. Trans. Mathematical*, pp. 2225–2244
5. Wörgötter, F., Porr, B.: Inside embodiment what means embodiment to radical constructivists? *Kybernetes*, 105–117 (2005)
6. Booth, T.L.: *Sequential Machines and Automata Theory*, 1st edn. (1967)
7. Braitenberg, V.: *Vehicles: Experiments in synthetic psychology*. MIT Press, Cambridge (1984)
8. Zhang, J., Bi, D., Wang, G.L.: Novel learning feed-forward controller for accurate robot trajectory tracking. In: Wang, L., Chen, K., S. Ong, Y. (eds.) *ICNC 2005*. LNCS, vol. 3611, pp. 266–269. Springer, Heidelberg (2005)
9. Nehaniv, C., Klyubin, A.S., Polani, D.: Empowerment: A universal agent-centric measure of control. *Proceedings of the IEEE Congress on Evolutionary Computation* 1, 128–135 (2005)
10. Nehaniv, C.L., Klyubin, A.S., Polani, D.: Keep your options open: An information-based driving principle for the sensorimotor systems. In: *PLoSOne*, vol. 3 (2008)
11. Polani, D., Klyubin, A.S., Nehaniv, C.L.: Organization of the information flow in the perception-action loop of evolved agents, pp. 177–180 (June 2004)
12. Kolodziejski, C., Kulvicius, T.: On the analysis of differential hebbian learning in closed-loop behavioral systems. In: *Frontiers in Computational Neuroscience*. Conference Abstract: Bernstein Conference on Computational Neuroscience (2009)
13. Luhmann, N.: *Social Systems* (1996)
14. Bullwinkle, D., Lungarella, M., Pegors, T.: Methods for quantifying the information structure of sensory and motor data. *Neuroinformatics* 3, 243–262 (2005)
15. Di Prodi, P., Porr, B., Wörgötter, F.: Adaptive communication promotes subsystem formation in a multi agent system with limited resources. In: *LAB-RS 2008: Proceedings of the 2008 ECSIS Symposium on Learning and Adaptive Behaviors for Robotic Systems*, pp. 89–96 (2008)
16. Sporns, O., Pfeifer, R., Lungarella, M., Kuniyoshi, Y.: On the information theoretic implications of embodiment - principles and methods. In: *Proc. of the 50th Anniversary Summit of Artificial Intelligence*, pp. 76–86 (2008)
17. Porr, B., Wörgötter, F.: Strongly improved stability and faster convergence of temporal sequence learning by utilising input correlations only. *Neural Computation* 18(6), 1380–1412 (2006)
18. Blahut, R.E.: Computation of channel capacity and rate distortion functions. *IEEE Trans. on Inform. Theory* 18(4), 460–473 (1972)
19. Roulston, M.S.: Estimating the errors on measured entropy and mutual information. *Physica D*, 285–294 (1999)
20. Arimoto, S.: An algorithm for computing the capacity of arbitrary memoryless channels. *IEEE Transactions on Information Theory* 18(1), 14–20 (1972)
21. Stuart, B.: Nicholas minorsky and the automatic steering of ships. *IEEE Control Systems Magazine* 4(4) (1984)
22. Sutton, A.G., Barto, R.S.: *Reinforcement learning: An introduction*. MIT Press, Cambridge (1998)
23. Schreiber, T.: Measuring information transfer. *Phys. Rev. Lett.* 85, 461–464 (2000)
24. Touchette, H., Lloyd, S.: Information-theoretic limits of control. *Phys. Rev. Lett.* 84(6), 1156–1159 (2000)