Two Agents Acting as One

Malte Harder[1], Daniel Polani[1] and Chrystopher L. Nehaniv[1]

[1]Adaptive Systems Research Group, University of Hertfordshire, Hatfield, UK
{m.harder,d.polani,c.l.nehaniv}@herts.ac.uk

Abstract

We consider two agents, each equipped with a controller. When they achieve a joint goal configuration, their coordination can be measured informationally. We show that the amount of *coordination* that two agents need to configure in a certain way depends on the amount of information they obtain from their environment. Furthermore the environment imposes a coordination pressure on the agents that depends on the size of the environment. In a second scenario we introduce a shared centralized controller which leads to a synchronisation of the agents' actions for suboptimal policies. However, in the optimal case this *intrinsic coordination* vanishes and the shared centralized controller can be split into two individual controllers.

Introduction

When one considers biology, many phenomena require that subentities perform actions in a coordinated way. This phenomenon is so prevalent that it requires pivotal treatment. It is seen in swarms, morphogenesis as well as in the actions of different parts of a single organism. We wish to study some principles behind this central phenomenon in an Artificial Life setting. In the sense of a 'life that could have been' (Langton, 1997) we are interested in what minimal assumptions have to be made to investigate coordination and autonomy within a collective of two agents. For this purpose we use the framework of information theory. We do not assume a particular metabolism and intrinsic dynamics but have the choice of certain limitations on information processing. This makes it possible to develop necessary and sufficient conditions for life-like scenarios and to find invariants for Artificial Life in any type of environment.

Nonetheless a physically consistent model can be plugged into the framework. Furthermore, studying coordination in a scenario that approximates nature has many applications: In ethology the understanding of collective tasks like foraging, flocking or group decision-making is active research (Deneubourg and Goss, 1989; Couzin et al., 2005; Nabet et al., 2009). Social interactions and coordination in robotics have been first studied by Walter (1950) and these issues in natural and artificial agents have received more attention lately (Dautenhahn, 1995, 1999; Ikegami and Iizuka,

2007; Di Paolo et al., 2008), for a review see (Goldstone and Janssen, 2005). Furthermore agent based and cellular models of morphogenesis have been studied with respect to coordination: Deneubourg et al. (1991) investigated the dynamics of ant-like agents that were not able to communicate directly but could pick up and drop objects of different types, leading to coordinated behaviour, called stigmergy, among the agents and clustering of objects of the same type. In an effort to understand morphogenesis of a certain slime mold, coordination between cells was modelled on a sub-cellular level, resulting in a simulation of the self-organised migration of the mold via an emergent level of photo- and thermotaxis (Marée and Hogeweg, 2001).

Stigmergy and local observation are common ways to model agent communication to get coordinated behaviour (Beckers et al., 1994; Castelfranchi, 2006). In both cases the communication is 'routed' through the environment, in the case of stigmergy in a very explicit way by altering the environment. In these models communication is spatially bound and limited by the amount of information that can be 'stored' in the environment.

When we talk about information, we specifically mean Shannon information (Shannon, 1948). The theory that comes with it allows to compare and quantify relations between random variables which can be used to model causal relationships in Bayesian graphs. Information theory gives a universal language to quantify conditions and invariants for a large class of models in very general way. Furthermore, this allows to compare quantities of models that are otherwise not directly comparable.

To study agent coordination from an information-theoretic perspective towards a predictive and quantitative theory of agent interactions we will look at embodied agents in a grid-world that is underlain by certain 'physical laws', like movement and blocking by other agents. To isolate the influences that a constraint of the agent's information processing capabilities has on the agents' coordination, we will neither impose an environmental constraint on the communication between them, nor a constraint on their sensors. The agents will have a shared controller, but we will limit their

information processing capabilities. Using the information-theoretic quantification of coordination, we will investigate how much they need to coordinate to achieve a given goal in the grid-world and compare this to the coordination in the case where the agents have independent controllers. Obviously the size of the environment has an impact on the amount of coordination as in large grid worlds with few agents there is a smaller chance of collision and less necessity to deal with this situation in an optimal way. For a shared controller we will investigate when the actions are coordinated in a way such that it is not possible to split the controller into two independent controllers which we interpret as both agents 'acting as one'.

Information theory has been successfully employed to models of embodied agents in a growing body of scientific literature starting from Ashby (1956). The idea that information is a main resource for organisms, but at the same time costly to process, is reflected in the evolution of sensors (Nehaniv et al., 2007) and affects the way information theoretic models of agents are investigated (Polani et al., 2007). Lately this idea received increased attention due to new techniques (Touchette and Lloyd, 2000; Klyubin et al., 2004a, 2007; Ay et al., 2008) and there are now broad applications of information theory to Artificial Life related fields (Linsker, 1988; Shalizi and Crutchfield, 2002). Recent results showed that information theoretic learning principles can lead to higher coordination between linked agents (Zahedi et al., 2009) though a different notion of coordination than in this paper is used. In the context of the Information Bottleneck (Tishby et al., 1999) the concept of *relevant information* was introduced by (Polani et al., 2001) and later extended to the perception-action loop (Polani et al., 2006). Here it will be set in relation to an information theoretic quantification of coordination as the mutual information between actions. Sperati et al. (2008) already used the mutual information between actions as a measure of coordination to evolve maximally coordinated agents.

When agents socially interact, or coordinate in an environment they sometimes seem to act as a single entity (e.g. bee hives, ant colonies, multicellular organism), at the same time they are individuals acting at a 'lower' level. In our experiment we will study under which constraints the agents can still be considered as autonomous with respect to the other agents and whether acting as a single entity helps to perform better to achieve a given configuration. Therefore we will introduce a measure of *intrinsic coordination* between two agents which vanishes if both agents have an independent controller and attains its maximum if the action of one agent is fully determined by the action of the other. We will then analyse how much intrinsic coordination is actually needed when acting optimally under an information processing constraint.

Information Theory

Information Theory was introduced by Shannon (1948). We will give a brief introduction: In information theory, *entropy* is given by $H(X) = -\sum_x p(x) \log p(x)$ 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

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

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. Again, this can be conditioned on a third random variable Z which gives the *conditional mutual information* $I(X;Y|Z) = H(Y|Z) - H(Y|X,Z)$. For a detailed account on information theory, see Cover and Thomas (2006).

Coordination

We propose measures of coordination that are independent of the topology of the environment and only depend on distributions of states and actions. Let S denote the random variable of the world states and A the random variable representing its actions where the actions only depend on the current state of the environment.

An important quantity in this context is Relevant Information: it is the minimal amount of information an agent needs to process to perform optimal actions (Polani et al., 2006), denoted by

$$I(S; A^*) = \min_{p(a|s):p(a|s)p(s)>0 \Rightarrow a \text{ optimal for } s} I(S; A).$$

This minimises the mutual information between states and actions but still requires that in each state with positive probability the optimal action is taken. Relevant information reflects, as mentioned in the introduction, the information parsimony principle that processing information has a metabolic cost (Polani et al., 2007) and complies with findings that certain neurons work at information limits, minimising the bandwidth to just maintain their function (Laughlin, 2001).

In theory the relevant information can be much lower than the bandwidth of the sensor, that is, different sensory inputs lead to the same distribution of actions. Moreover, one can ask the converse question: how well can a policy perform if $I(S; A)$ is limited? To do this a utility in terms of a reward structure will be used and the trade-off will be calculated with an algorithm introduced by (Polani et al., 2006).
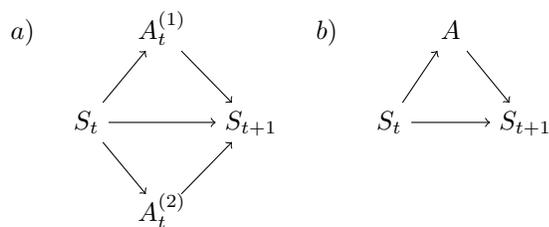
Figure 1: Bayesian network of the perception-action loop for a) independent actions b) joint actions. Here $A^{(1)}$ and $A^{(2)}$ denote the random variable of the action of each agent, A denotes the random variable of the joint action $(a^{(1)}, a^{(2)})$ and t is the time index. In both cases the actions are fully determined by the current state of the environment.

Suppose now there are two agents; the *coordination* is then defined as the mutual information between their actions $I(A^{(1)}; A^{(2)})$ where $A^{(1)}$ is the random variable representing the actions of the first agent and $A^{(2)}$ the random variable representing the actions of the second agent. In the case of independently embodied agents, that is, if $p(a^{(1)}, a^{(2)}|s) = p(a^{(1)}|s)p(a^{(2)}|s)$ the coordination is limited by the relevant information of each agent

$$I(A^{(1)}; A^{(2)}) \leq \min\{I(S; A^{(1)}), I(S; A^{(2)})\}.$$

This follows easily from the data processing inequality (Cover and Thomas, 2006, p. 34). If the agents however have a joint policy $p(a^{(1)}, a^{(2)}|s)$ the coordination is only limited by the entropy of the actions. See Figure 1 for the perception-action loop of the whole system in the case of a) independent controllers and b) one shared controller.

For such an agent pair that has one shared controller it is interesting to see whether there is any intrinsic coordination or whether the controller could be split into two independent controllers. We define *intrinsic coordination* as the conditional mutual information $I(A^{(1)}; A^{(2)}|S)$ which vanishes if $p(a^{(1)}, a^{(2)}|s) = p(a^{(1)}|s)p(a^{(2)}|s)$, that is, the agents come to independent decisions given the state of the environment. By definition intrinsic coordination can be higher or lower than the coordination. In the case that the actions are independent of the state, that is, $H(A^{(1)}|S) = H(A^{(1)})$ and $H(A^{(2)}|S) = H(A^{(2)})$, coordination equals intrinsic coordination, however, the converse is not always the case.

Experimental Setup

We want to study how much (intrinsic) coordination the agents have when they follow an optimal policy to achieve a particular goal configuration (under information processing constraints). Furthermore the amount of coordination will be compared to the coordination in the case where the agents have independent controllers.

The setup consists of two agents, determined by a joint state $s = (s^{(1)}, s^{(2)}) \in \mathcal{S}$ in the state space $\mathcal{S} = \mathcal{W} \times \mathcal{W} - \Delta$ where $\mathcal{W}$ is a $n \times m$ grid-world and $\Delta = \{(w, w)|w \in \mathcal{W}\}$ the diagonal. Hence only one agent is allowed to occupy a particular grid cell per time step. As before, the random variable representing the state of the environment is denoted by S. The goal is given by two particular adjacent cells in the centre of the grid-world and it is not relevant which agent occupies which goal cell, hence there are two goal states in the state space $\mathcal{S}$.

Each agent has five possible actions $\{N, S, W, E, H\}$, go to one of the four neighbouring cells or halt. The actions are denoted by the random variables $A^{(1)}, A^{(2)}$, and their joint action $a = (a^{(1)}, a^{(2)})$ by the random variable A. The distribution of the actions only depends on the location of the two agents. In this scenario the transitions to the next step are deterministic $p(s_{t+1}|a_t, s_t) \in \{0, 1\}$ and reflect the movement of the two agent in the grid-world, blocked by the walls and blocking each other symmetrically (see Figure 2). The agents are blocked if they try to move to the same field or if one agent moves to a field where the other agent stays.

For every step the agents get a reward that is determined by a reward function $r(s_{t+1}, a_t, s_t)$ which depends on the current state, the action taken and the state of the world after the action was executed. A negative reward is given unless both agents occupy a goal cell in which case no reward or penalty is given. Thus, a policy that maximises the expected reward over the lifetime of the agent is one that takes the shortest way to the goal configuration. This defines a Markov Decision Process (MDP), for which reinforcement learning can be used to find such a policy. Given the MDP we can define a *state value function* $V^\pi(s)$ that gives the expected future reward at some state s following the policy π and a *utility function* $U^\pi(s, a)$ that gives the expected reward incorporating the action chosen at state s and then following
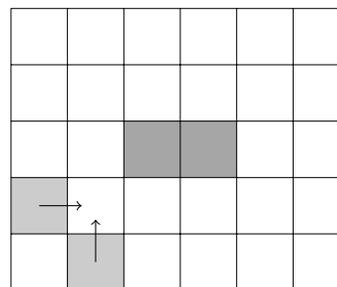
Figure 2: In this 6×5 grid-world, the two dark-grey rectangles show the goal configuration, the light-grey rectangles show a configuration where the agents block each other if they move in the directions of the arrows. This causes that the agents stay at their current position.

the policy π:

$$V^\pi(s) = \sum_a \pi(a|s) \sum_{s'} p(s'|a,s) \left(r(s',a,s) + V^\pi(s') \right),$$

$$U^\pi(s,a) = \sum_{s'} p(s'|a,s) \left(r(s',a,s) + V^\pi(s) \right).$$

The definition of the state value function is recursive and the correct value function is a fixed point of this equation. Iterating the recursive definition of the value function converges to the correct value function for a given policy. If the policy is updated to be greedy with respect to the current utility in every step, the iteration, called optimistic policy iteration, results in an optimal policy for the MDP (Sutton et al., 1999).

If the agents' actions are independent, that is, if $p(a^{(1)}, a^{(2)}|s) = p(a^{(1)}|s)p(a^{(2)}|s)$, the problem breaks down to two dependent MDPs that are not deterministic anymore but whose transition probabilities depend on a prediction of the other agent's action $\tilde{p}(a^{(i)}|s)$. For instance when agent i expects j to act according to $\tilde{p}(a^{(i)}|s)$, then the predictor for the transition of i is:

$$\tilde{p}(s_{t+1}|a_t^{(i)}, s_t) = \sum_{a^{(j)}} \tilde{p}(a^{(j)}|s_t) p(s_{t+1}|a^{(i)}, a_t^{(j)}, s_t),$$

where $i, j \in \{1, 2\}$ and $i \neq j$. In this paper we will update the predictor in every iteration to be the same as the policy of the other agent: $\tilde{p}(a^{(i)}|s) = \pi(a^{(i)}|s)$. That means the agents can do the best possible prediction of the action of the other agent in every step.

Given a scenario where agents do not know anything about each other, it is possible to set the predictor to a uniform distribution. But we want to study how the performance of a split controller compares to the shared controller and will use the policy of the other agent to make the best prediction about the action of the other agent as possible.

The performance of a policy π is measured by the expected utility over all state action pairs, denoted $\mathbb{E}[U^\pi(S, A)]$. To compare both cases a different reward is used in each case: For the shared controller a reward of -2 is given whenever the agents do not enter a goal state. For the independent controllers a reward of -1 is given to each of the agents if it does not enter a goal state, so in each case the summed reward per step is -2 if the goal is not reached. Using the current policy as the predictor $\tilde{p}$ gives another advantage: For the joint policy $\pi(a^{(1)}, a^{(2)}|s) = \pi(a^{(1)}|s)\pi(a^{(2)}|s)$, now the following holds

$$\mathbb{E}[U^\pi(S, A)] = \mathbb{E}[U^{\pi^1}(S, A^{(1)})] + \mathbb{E}[U^{\pi^2}(S, A^{(2)})],$$

where U^π is the utility consistent with the joint policy and U^{π^1}, U^{π^2} are the utilities consistent with the policies $\pi(a^{(1)}|s), \pi(a^{(2)}|s)$. Thus we have a common scale for the expected utilities.

Algorithm

As introduced before, the relevant information is the mutual information between sensor and actions, minimised over all optimal policies. Minimising mutual information under the constraint of a distortion measure can be done using the Blahut-Arimoto algorithm (Blahut, 1972). To obtain a policy that is optimal and minimising, Polani et al. (2006) used a Blahut-Arimoto iteration with the utility $U^\pi(s, a)$ as a distortion measure. The Blahut-Arimoto iteration is given by

$$\pi_{k+1}(a|s) = \frac{p_k(a)}{Z_k(s, \beta)} \exp(\beta U^\pi(s, a)),$$

$$p_{k+1}(a) = \sum_s p_k(s)\pi_k(a|s),$$

where k denotes the iteration step, $Z_k(s, \beta)$ is a normalisation term and $\beta > 0$ a trade-of between optimality and relevant information. Now the iteration is alternated with an update of the state probabilities and a value iteration to get a consistent utility U^{π_k}.

The agents act only until they reach the goal configuration, the task is episodic. The probability to be in state s after t steps is given by

$$p(s|t) = \frac{1}{|\mathcal{S}|} \sum_{s'} P^t(s, s')$$

where P is the state transition probability matrix and a uniform distribution for $t = 0$ is assumed. Let s^{g_1}, s^{g_2} denote the two goal states. Now the probability that the agent is in state s and it has not reached the goal, denoted as living, is

$$p(s|\text{living}) = \lim_{T \to \infty} \frac{\sum_{t=0}^T \delta(s)p(s|t)}{\sum_{t=0}^T 1 - p(s^{g_1}|t) - p(s^{g_2}|t)},$$

where δ is zero if s is a goal state and one otherwise. Now we set $p(s) = p(s|\text{living})$. Updating the state probabilities is important as a correct state distribution is essential for good convergence of the algorithm.

For the whole iteration the iterations steps are then done in the following order

$$\pi_k \to p_k(s) \to V^{\pi_k} \to U^{\pi_k} \to \pi_{k+1}.$$

The algorithm then minimises the functional

$$\mathcal{L}[p(a|s)] = I(S; A) - \beta\, \mathbb{E}[U^\pi(S, A)].$$

As an optimal policy maximises the expected utility, the Lagrange multiplier β determines a trade-of between an optimal policy and limited relevant information. Iterating the algorithm for small β results in optimal policies given a limitation on the relevant information, which is of particular interest as many real world agents especially in collectives have very limited information processing capabilities. For

$\beta \to \infty$ the resulting policy is optimal and at the same time minimises the mutual information $I(S; A)$.

Recent work shows that extending relevant information to multiple steps, results in a similar algorithm that unifies the value iteration and the Blahut-Arimoto iteration and gives a new framework for minimising information quantities in Bayesian graphs under optimality constraints (Tishby and Polani, 2010). A proof of convergence for these algorithms is work in progress.

Having two agents with independent actions will change the algorithm. The iteration is now alternated between the two agents. For each agent a value iteration and a Blahut-Arimoto iteration is done using the current policy of the other agent as a predictor in the utility update. This gives the following scheme of iterations:

$$\pi_k^1, \pi_k^2 \to p_k(s) \quad \to \quad V^{\pi_k^1} \to U^{\pi_k^1} \to \pi_{k+1}^1 \to ...$$
$$... \quad \to \quad V^{\pi_k^2} \to U^{\pi_k^2} \to \pi_{k+1}^2.$$

First, we have the two policies for each agent from which the common environmental state distribution is calculated. This is followed by a value iteration step for the first policy and a Blahut-Arimoto update that gives the new policy for the first agent. Using this policy as a predictor the value iteration step for agent two is done, again followed by a Blahut-Arimoto step.

For most samples the algorithm converged very fast, but for certain values of β this is not the case, however, these values can be detected by taking a fine distribution of samples for β.

Results

Iterations were performed with different environment sizes ($6 \times 7, 6 \times 5, 4 \times 5$, 4×3, 4×2 and $n \times 1$ with $n = 5, 6, 7, 8$). Samples were taken for different values of β ranging from 0.05 to 10.0 with steps ranging from 0.005 to 0.1, greater worlds required a larger step size due to computational limitations. Each value β leads to a policy and a state distribution, the performance of the policy can be plotted against the mutual information between actions and states (see Figure 3). At the upper limit of $\beta = 10.0$ the trade-of was already completely in favour of an optimal policy. For each sample the iteration was stopped when $\sum_s |V_{k+1}^\pi(s) - V_k^\pi(s)| < 10^{-6}$. In all runs the setup with a shared controller/policy outperforms the case where the actions are independent (see Figure 3). However the optimal ($\beta \to \infty$) shared controller shows almost no intrinsic coordination, that is $I(A^{(1)}; A^{(2)}|S)$ vanishes. Here the agents perform equally well with a shared controller as with independent controllers (see Figure 3 and 4). This suggests that in the optimal limit intrinsic coordination does not help to perform better. Similarly Zahedi et al. (2009) showed that

for linked robots, those performed better that had split controllers for their motors, although this was in the context of maximising predictive information.

In the suboptimal region, especially small values of β, the shared controller performs better with the same amount of relevant information. In this region the coordination behaves differently depending on the kind of controller. With independent controllers the coordination tends to zero, as less relevant information is processed (see Figure 5). While this was expected due to coordination limited by relevant information, the coordination is not even close to the possible limit. The shared controller shows the opposite behaviour: the coordination increases as less relevant information is processed. This is also valid for the intrinsic coordination, which vanishes in the optimal limit (see Figure 4).

The maximum of coordination of the shared controller depends closely on the size and geometry of the world (see Figure 6). The spikes in the graph are due to convergence problems for certain values of β. For larger worlds the coordination still increases for $\beta \to 0$, but by a significantly smaller amount: In a 6×7 grid world the difference be-
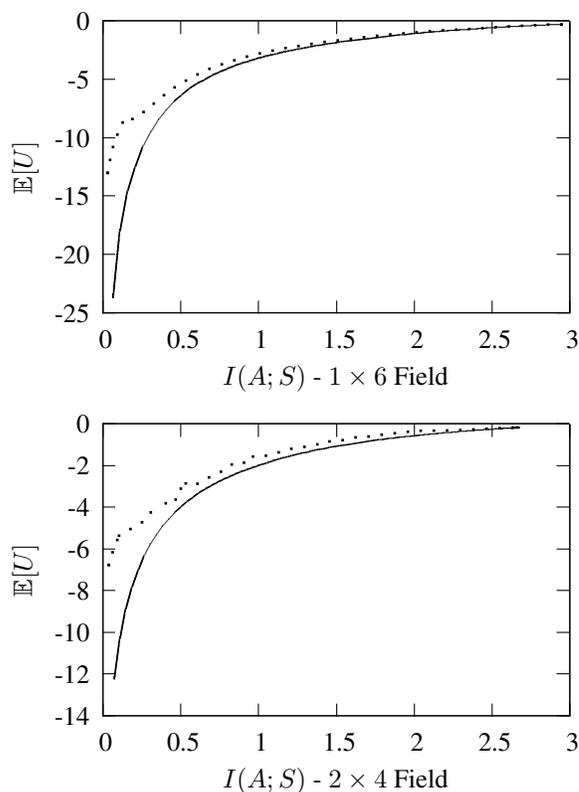
Figure 3: Performance of agents, dotted line – shared controller, solid line – individual controllers with summed expectation of utility per agent and relevant information for the joint distribution of $(a^{(1)}, a^{(2)})$. Both graphs show the same features but the scales differ.
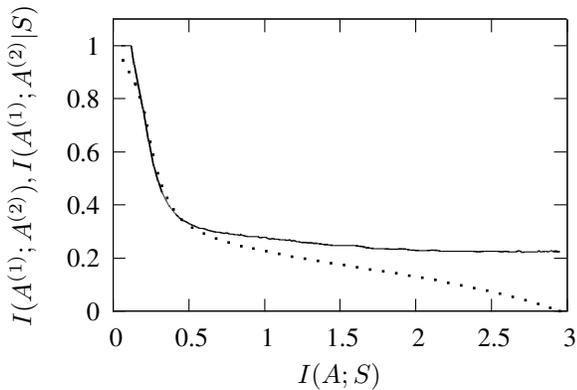
Figure 4: Coordination of agents with shared controller on a 6×1 field, dotted line – intrinsic coordination, solid line – coordination.

Figure 5: Solid line – coordination of agents with individual controllers on an 6×1 field, dotted line – limit given by each controllers relevant information.

tween the coordination for small and large values of β is only ≈ 0.05 bit whereas in a 4×5 world the difference is ≈ 1.54 bit. For very narrow worlds (size $n \times 1$) the coordination even reached its maximum $\max H(A^{(1)}) = \max H(A^{(2)}) = 1$ bit. It may seem unintuitive that this can happen while the relevant information is positive, as it means that one action fully determines the other and each of the two possible actions is chosen with probability $\frac{1}{2}$. However the coordination takes the expectation over all states: the actions can be totally synchronised, that is, $H(A_1|A_2) = 0$ while $H(A_1|S)$ is not maximal. Thus the distribution of the possible two synchronous actions is not uniform, but this effect can vanish when the expectation over all states is taken, which can also be seen by that fact that the intrinsic coordination does not equal the coordination and therefore the actions cannot be independent of the states.

The distribution of the states is not uniform and S has rather low entropy as the cells that are closer to the goal are visited more often by the agents. To ensure that the observed behaviour of coordination is prevalent over the whole state space and not just appearing close to the goal the resulting policies were also analysed assuming a uniform distribution of S, which resulted in insignificant differences.

Discussion

We introduced intrinsic coordination as a measure how much different agents' actions are correlated given the state of the environment. The setting we investigated is a grid world with two agents and a goal to configure in a certain way. As both agents have the same possible two goal states, they have to cooperate to reach the goal in an optimal way. The actions only depend on the current location of the agent (the agents are memoryless) thus the joint intent to move to the goal states is explicitly encoded in the controllers. Using an alternated fixed point iteration method we computed optimal policies for the agents under information processing constraints.
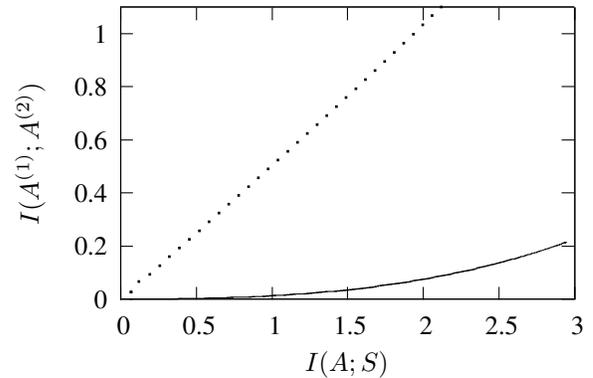
The results show that agents use intrinsic coordination to overcome limitations of their environment. This coordination is not needed in the optimal case where every agent can get all the relevant information from the environment that it needs to choose an optimal action. Though plausible, this is not entirely obvious a priori. One could think of various scenarios where the controllers are stochastic and the precise knowledge of the others agent action would lead to a better performance.

Now, large agent collectives will usually perform suboptimal policies as each agents' abilities will be limited: In real environments, the size of the agent and its supply of energy are just some limiting factors to information processing capabilities. Furthermore having many agents acting in the environment leads to spatial limitations that were here matched by the situation of narrow grid-worlds. In these cases intrinsic coordination performs better than just prediction of the other agents' behaviour: The shared controller cannot be split into two independent controllers, this is what we understand as 'acting as one'. The intrinsic coordination gives a measure of how strong this behaviour is. In the case of the 6×1 world and a small β the actions of the agents are always in the opposite direction, but with a small bias whether the agents move towards each other or away from each other. Despite being a feature of the controller the synchronisation does not depend on the state and there is no information needed to decide whether to act synchronised or not. The agents perform even better with this strategy. This could be interpreted as a kind of morphological computation (Pfeifer and Bongard, 2006) where the synchronisation is a feature of the embodiment of the agents used to perform better in reaching the goal configuration. Due to the symmetry of the present environment and the embodiment of the agents there is also a symmetry in the shared controller. However, intrinsic coordination does not specifically depend on symmetries
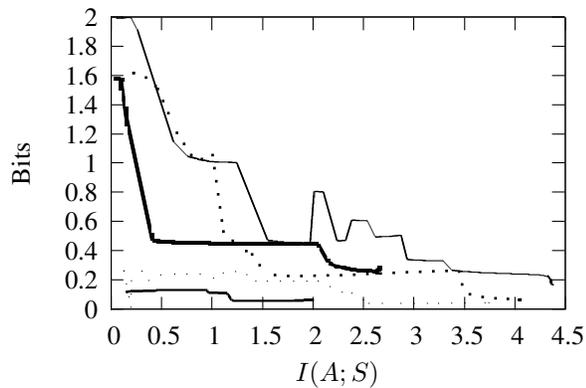
Figure 6: Coordination of agents with shared controllers on a,medium thick line – 6×7, thin dotted line – 6×5, thick dots 4×5, thick line 4×3, thin line – 4×2 field.

and can occur in any scenario within this formalism.

In the setup the intrinsic communication is not limited: the two agents share a common 'brain'. But often coordination is only 'routed' through the environment: In the case of stigmergy the environment takes the role of the communication channel (Klyubin et al., 2004b). Other ways of communication that have low interference with the environment like sound, dissolving molecules or radio signals qualify more to be modelled as intrinsic coordination, although their limited channel capacities must be considered. In our experiment intrinsic coordination was not modelled using directed communication and the agents came to a instantaneous joint decision. What we have not done here, but to what the formalism could be changed, is a dependence of $A^{(2)}$ on $A^{(1)}$, which would model connected controllers where the first agent can express an intent to which the second can react. This would be a more restrictive model than the shared controller. Moreover this framework can be further elaborated to take issues of time shifts and turn taking during the decision process into account. Examples where collectives of cells use molecular signalling, with almost no interference, to activate a certain behaviour in the whole collective (Marée and Hogeweg, 2001) could then be modelled as intrinsic coordination. One can argue that the molecular signalling should be modelled with each cell having an independent controller and a sensor for these molecules, but a model allowing intrinsic communication could lead to a simpler description and therefore be more preferable.

Furthermore it is not necessarily obvious whether a particular collective of agents is just a collection of individuals or acts as one individual. If there is a simpler model allowing intrinsic coordination does that automatically mean that it acts as a single entity? Ant colonies are sometimes called super-organisms (Theraulaz and Bonabeau, 1999) and were recently found to fulfil certain laws that apply for animals (Hou et al., 2010), melting the boundary between the individual and the collective. If two agents have the possibility of maximal intrinsic coordination they can hardly be viewed as individual agents as their actions are completely synchronised. Thus having non-maximal intrinsic coordination gives each agent a certain degree of freedom to decide for an action solely on its own perception of the environment. This means that a collective with a shared centralized controller still can undertake actions that conflict each other, especially in the suboptimal case, but intrinsic coordination can be used to avoid this to a certain degree. In the spirit of defining autonomy for a system in an information theoretic way (Bertschinger et al., 2008), intrinsic coordination could function as another measure of individuality or autonomy with respect to other agents.

References

Ashby, W. R. (1956). *An Introduction to Cybernetics*. Chapman & Hall Ltd.

Ay, N., Bertschinger, N., Der, R., and Güttler, F. (2008). Predictive information and explorative behavior of autonomous robots. *The European Physical Journal B - Condensed Matter and Complex Systems*, 63(3):329–339.

Beckers, R., Holland, O., and Deneubourg, J. (1994). From local actions to global tasks: Stigmergy and collective robotics. In *Artificial Life IV*, pages 181–189. MIT Press.

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

Blahut, R. (1972). Computation of Channel Capacity and Rate Distortion Functions. *IEEE Transactions on Information Theory*, 18(4):460–473.

Castelfranchi, C. (2006). Silent agents: From observation to tacit communication. *Advances in Artificial Intelligence-IBERAMIA-SBIA*, 4140:98–107.

Couzin, I. D., Krause, J., Franks, N. R., and Levin, S. a. (2005). Effective leadership and decision-making in animal groups on the move. *Nature*, 433(7025):513–6.

Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory 2nd Edition*. Wiley Series in Telecommunications and Signal Processing. Wiley-Interscience.

Dautenhahn, K. (1995). Getting to know each other–artificial social intelligence for autonomous robots. *Robotics and Autonomous Systems*, 16:333–356.

Dautenhahn, K. (1999). *Embodiment and interaction in socially intelligent life-like agents*, pages 102–142. Springer Lecture Notes in Artificial Intelligence.

Deneubourg, J. and Goss, S. (1989). Collective patterns and decision-making. *Ethology, Ecology and Evolution 1*, pages 295–311.

Deneubourg, J., Goss, S., and Franks, N. (1991). The dynamics of collective sorting: Robot-like ants and ant-like robots. *Proceedings of the first international conference on simulation of adaptive behavior*.

Di Paolo, E., Rohde, M., and Iizuka, H. (2008). Sensitivity to social contingency or stability of interaction? Modelling the dynamics of perceptual crossing. *New Ideas in Psychology*, 26(2):278–294.

Goldstone, R. and Janssen, M. (2005). Computational models of collective behavior. *Trends in Cognitive Sciences*, 9(9):424–430.

Hou, C., Kaspari, M., Vander Zanden, H. B., and Gillooly, J. F. (2010). Energetic basis of colonial living in social insects. *Proceedings of the National Academy of Sciences*, 107(8):3634–8.

Ikegami, T. and Iizuka, H. (2007). Turn-taking interaction as a co-operative and co-creative process. *Infant Behavior and Development*, (30):278–288.

Klyubin, A. S., Polani, D., and Nehaniv, C. L. (2004a). Organization of the Information Flow in the Perception-Action Loop of Evolved Agents. In Zebulum, R. S., Gwaltney, D., Hornby, G., Keymeulen, D., Lohn, J., and Stoica, A., editors, *Proceedings of 2004 NASA/DoD Conference on Evolvable Hardware*, pages 177–180. IEEE Computer Society.

Klyubin, A. S., Polani, D., and Nehaniv, C. L. (2004b). Tracking Information Flow through the Environment: Simple Cases of Stigmergy. In Pollack, J., Bedau, M., Husbands, P., Ikegami, T., and Watson, R. A., editors, *Artificial Life IX: Proceedings of the Ninth International Conference on the Simulation and Synthesis of Living Systems*, pages 563–568, Cambridge, MA. MIT Press.

Klyubin, A. S., Polani, D., and Nehaniv, C. L. (2007). Representations of Space and Time in the Maximization of Information Flow in the Perception-Action Loop. *Neural Computation*, 19(9):2387–2432.

Langton, C. (1997). *Artificial life: An overview*. The MIT Press.

Laughlin, S. B. (2001). Energy as a constraint on the coding and processing of sensory information. *Current Opinion in Neurobiology*, 11(4):475–480.

Linsker, R. (1988). Self-Organization in a Perceptual Network. *Computer*, 21(3):105–117.

Marée, A. F. and Hogeweg, P. (2001). How amoeboids self-organize into a fruiting body: multicellular coordination in *Dictyostelium discoideum*. *Proceedings of the National Academy of Sciences of the United States of America*, 98(7):3879–83.

Nabet, B., Leonard, N., Couzin, I., and Levin, S. (2009). Dynamics of Decision Making in Animal Group Motion. *Journal of Nonlinear Science*, 19(4):399–435.

Nehaniv, C., Polani, D., and Olsson, L. (2007). *Information-Theoretic Modeling of Sensory Ecology: Channels of Organism-Specific Meaningful Information*, pages 241–282. MIT Press.

Pfeifer, R. and Bongard, J. (2006). *How the Body Shapes the Way We Think: A New View of Intelligence*. MIT Press, Cambridge, MA.

Polani, D., Martinetz, T., and Kim, J. T. (2001). An Information-Theoretic Approach for the Quantification of Relevance. In *ECAL '01: Proceedings of the 6th European Conference on Advances in Artificial Life*, pages 704–713, London, UK. Springer-Verlag.

Polani, D., Nehaniv, C. L., Martinetz, T., and Kim, J. T. (2006). Relevant Information in Optimized Persistence vs. Progeny Strategies. In *Artificial Life X : Proceedings of the Tenth International Conference on the Simulation and Synthesis of Living Systems*, pages 337–343. The MIT Press (Bradford Books).

Polani, D., Sporns, O., and Lungarella, M. (2007). How Information and Embodiment Shape Intelligent Information Processing. In *50 Years of Artificial Intelligence*, pages 99–111.

Shalizi, C. R. and Crutchfield, J. P. (2002). Information Bottlenecks, Causal States, and Statistical Relevance Bases: How to Represent Relevant Information in Memoryless Transduction. *Advances in Complex Systems*, 5:91.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423.

Sperati, V., Trianni, V., and Nolfi, S. (2008). Evolving coordinated group behaviours through maximisation of mean mutual information. *Swarm Intelligence*, 2(2-4):73–95.

Sutton, R., Precup, D., and Singh, S. (1999). A Framework for Temporal Abstraction in Reinforcement Learning. *Artificial intelligence*, 112(1-2):181–211.

Theraulaz, G. and Bonabeau, E. (1999). A brief history of stigmergy. In *Artificial Life V*, pages 97–116. MIT Press.

Tishby, N., Pereira, F. C., and Bialek, W. (1999). The information bottleneck method. In *The 37th annual Allerton Conference on Communication, Control, and Computing*, pages 368–377.

Tishby, N. and Polani, D. (2010). Information Theory of Decisions and Actions. In Cutsuridis, V., Hussain, A., and Taylor, J., editors, *Perception-Reason-Action Cycle: Models, Algorithms and Systems*. Springer. In Press.

Touchette, H. and Lloyd, S. (2000). Information-theoretic limits of control. *Physical Review Letters*, 84(6):1156–1159.

Walter, W. (1950). An imitation of life. *Scientific American*, (May):42–45.

Zahedi, K., Ay, N., and Der, R. (2009). Higher coordination with less control - A result of information maximisation in the sensori-motor loop. *CoRR*, abs/0910.2.