Information-theoretic Cost of Decision-making in Joint Action

Dari Trendafilov[1], Daniel Polani[2] and Alois Ferscha[1]
[1]*Institute of Pervasive Computing, Johannes Kepler University, Linz, Austria*
[2]*Adaptive Systems Research Group, University of Hertfordshire, Hatfield, U.K.*

Keywords: Collective Behavior, Information Dynamics, Complex Systems, Information Theory.

Abstract: We investigate the information processing cost relative to utility, associated with joint action in dyadic decision-making. Our approach, built on the Relevant Information formalism, combines Shannon's Information Theory and Markov Decision Processes for modelling dyadic interaction, where two agents with independent controllers move an object together with fully redundant control in a grid world. Results show that increasing collaboration relaxes the pressure on required information intake and vice versa, antagonistic behavior takes a higher toll on information bandwidth. In this trade-off the particular embodiment of the environment plays a key role, demonstrated in simulations with informationally parsimonious optimal controllers.

1 INTRODUCTION

Individual agents with limited information-processing capabilities can successfully coordinate their behavior and make well-informed collective decisions. Examples include collective behaviors, such as coordinated motion of birds and fishes (Couzin et al., 2005), bees and ants (Franks et al., 2002), or coordination of individual cells (Pezzulo and Levin, 2015), which are complex collective phenomena studied from various perspectives, e.g., population models (Couzin et al., 2005; Marshall et al., 2009), game-theoretic (Challet and Zhang, 1997), or multi-agent simulations (Goldstone and Janssen, 2005).

When agents interact socially, sometimes they act in a complete agreement towards a shared goal, and other times their goals may diverge or be completely incompatible with each other. Over time, the interaction dynamics could alternate between cooperative and antagonistic behavior and the flow of information could reveal a given player's contribution for the emergence of a particular strategy. In certain cases, limitations in the perception–action loop or in the decision-making capabilities of agents could make them behave irrationally while coordinating their joint activities (e.g. autistic behavior).

In our study, we explore the information-processing burden imposed on a completely rational agent by its irrational (or adverse) partner. We examine the trade-off between the level of compliance (i.e. cooperative coordination) within a dyad and the information processing cost incurred by the cooperative agent while aiming to achieve a predefined goal (i.e. utility level). Our framework, built on Shannon's information theory (Shannon, 1949), imposes certain information processing constraints, while not assuming any intrinsic dynamics nor a particular metabolism in providing necessary and sufficient environmental conditions and invariants. This approach allows to characterize causal relationships in Bayesian networks. For this purpose, information theory provides a universal language for quantifying conditions and invariants for a large class of models in a generic and principled way. Furthermore, it allows comparison 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 look at embodied agent dyads interacting in a grid-world. The agents have independent controllers and a fully redundant set of actions for controlling the same object in the grid. Using information-theoretic tools, we explore the information processing constraints of an agent in achieving a specific goal, depending on the level of cooperation by its partner.

2 BACKGROUND

Initially, social interaction and coordination have been studied by (Walter, 1950) in natural and artificial agents, and more recently by (Ikegami and Iizuka, 2007; Paolo et al., 2008; Goldstone and Janssen, 2005). The study of coordination has applications

Trendafilov, D., Polani, D. and Ferscha, A.
Information-theoretic Cost of Decision-making in Joint Action.
DOI: 10.5220/0010252303040311
In *Proceedings of the 13th International Conference on Agents and Artificial Intelligence (ICAART 2021) - Volume 1*, pages 304-311
ISBN: 978-989-758-484-8

in ethology, where it helps to understand collective tasks like foraging, flocking or group decision-making (Couzin et al., 2005; Nabet et al., 2009). Stigmergy in coordinated behavior of ant-like agents was investigated by (Meyer and Wilson, 1991) in a study based on cellular models of morphogenesis. Stigmergy and local observation are common ways for modelling agent communication and coordinated behavior (Castelfranchi, 2006). In both cases, communication is 'channeled' 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.

The study of collective behavior has traditionally relied on a variety of different methodological tools, ranging from more theoretical methods such as population or game-theoretic models to empirical ones like Monte Carlo or multi-agent simulations. More recent approaches use information theory as a methodological framework to study the flow of information and the statistical properties of collectives of interacting agents. Several general purpose tools provide efficient information-theoretic analysis, including classical information-theoretic measures, measures of information dynamics and information-based methods to study the statistical behavior of collective systems (Moore et al., 2018; Lizier et al., 2014; Lindner et al., 2011). A study by (Valentini et al., 2018), using simulated agents, applies transfer entropy (Schreiber, 2000) to measure the flow of information generated in collective decision-making mechanisms. The application of complexity measures, and information theory metrics in general, to support the automatic design and analysis of collaborative dynamics, have recently attracted the interest of multi-agent and robotics communities, owing to their capability to capture relevant features of robot behaviors while abstracting from implementation details (Roli et al., 2019). In a study of a homeokinetic dyad (Martius et al., 2008) suggest the fundamental role of the maximal integration of the environment into the sensorimotor loop. Building on the free energy principle of the brain (Friston, 2010), in conjunction with the theory of sense-making, (Davis et al., 2017) proposed a method to quantify interaction dynamics of creative collaboration, e.g. rhythm of interaction or turn-taking style, in order to characterize collaborations over time. Multi-objective reinforcement learning has been applied with respect to fairness (Matsui, 2019) and dyadic trust models have been investigated by (Callebert et al., 2016).

Information theory has been successfully employed to models of embodied agents in a growing body of scientific literature starting from (Ashby, 1956). It provides a quantitative framework for assessing how information is exchanged and processed in collectives and has been explored in studies of natural (Zenil et al., 2015; Meyer, 2017; Butail et al., 2016; Mwaffo et al., 2017) and artificial systems (Sperati et al., 2011; Boedecker et al., 2011; Walker et al., 2012; Kim et al., 2015). The premise is that information is a key resource for organisms, which is costly to process and affects the way information-theoretic models of agents are investigated (Polani et al., 2007). This idea has received an increased attention due to new techniques by (Touchette and Lloyd, 2000; Klyubin et al., 2007), and there are now broad applications of information theory (Shalizi and Crutchfield, 2002). The concept of relevant information, introduced by (Polani et al., 2001; Polani et al., 2006), builds on the Information Bottleneck principle (Tishby et al., 1999), extended to the perception–action loop. Building on the relevant information formalism, (Harder et al., 2010) introduced a measure of individuality (or autonomy) in a collaborative task under information processing constraints.

3 INFORMATION THEORY

Shannon's Information Theory defines entropy by

$$H(X) = -\sum_x p(x)\log p(x)$$

where X denotes a discrete random variable taking values from $\mathcal{X}$, and $p(x)$ the probability of X taking on the value $x \in \mathcal{X}$. Entropy can be interpreted as the average amount of information gained when a variable's value is revealed. When multiple random variables are correlated, then knowing the value of one reduces the uncertainty about the other. The average uncertainty about Y left after revealing the value of X is quantified by the conditional entropy

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

The average reduction in uncertainty, interpreted as the amount of information that knowing X gives about Y, is defined as the mutual information between the two variables

$$I(X;Y) = H(Y) - H(Y|X).$$

In this study we apply and adapt the Relevant Information method originally introduced in (Polani et al., 2006), to joint action. It provides a measure depending only on the stochastic model and independent of the topology of the environment. The measure quantifies the minimal amount of information an agent needs to process in order to achieve a certain level

of utility as specified by a reward function. It reflects an information parsimony principle, suggesting that processing information has a metabolic cost (Polani et al., 2007) and complies with findings that certain neurons work at informational 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. An algorithm, proposed in (Polani et al., 2006), computes the information-utility trade-off using a utility in terms of a reward structure.

4 SCENARIO

In our experimental scenario two agents perform the task of achieving a particular goal configuration together (under information processing constraints) using controls from the same set of actions and alternating their moves on even (agent A) and odd (agent B) steps. We can compute the optimal policy for each agent using the Relevant Information method, assuming we have a prediction for the policy of the other agent. Our goal is to investigate how the behavior of agent B influences the optimal policy and the information constraints of agent A, and how that affects the joint action.

The experimental setup consists of two agents jointly moving an object in a 2-D grid-world by using four actions (up, down, left, right). The goal is to move the object from one corner of the grid to the opposite corner along the diagonal. The state of the environment is denoted with the random variable S. The scenario uses deterministic state transition model $p(s_{t+1}|a_t, s_t) \in \{0, 1\}$, which reflects the movement of the object in the grid-world constrained by the walls. In every step the agents get a reward 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 is executed. The reward function of agent A is -1 for all states except the goal state where it is 0. Thus, a policy that maximises the expected utility of agent A is one that takes the shortest path to the goal configuration. Note that due to the particular action set there are many different shortest paths to the goal, which however differ in informational cost as will be demonstrated later.

The underlying Markov Decision Process (MDP) defines a reinforcement learning problem in which a state value function $V^{\pi}(s)$ specifies 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 following the policy π, defined as

$$U^{\pi}(s,a) = \sum_{s'} p(s'|s,a)[r(s,a,s') + \gamma V^{\pi}(s')]. \quad (1)$$

The discount factor $\gamma \in [0,1]$ controls the trade-off between long-term and short-term rewards – if γ is large, future rewards weigh more in the expected sum, and if it is small immediate rewards have higher relative weight.

Finding an expected future reward maximising policy is equivalent to solving the following optimization problem:

$$\pi^* = \arg\max_{\pi} V^{\pi}(s),$$

where the optimal policy π^* is not necessarily unique. Moreover, any convex combination $\alpha\pi_1^* + (1-\alpha)\pi_2^*$ ($\alpha \in (0,1)$) of two optimal policies π_1^* and π_2^* results in a new optimal policy. Having multiple optimal policies poses the question of whether one is preferred over the others. One natural criterion is the cognitive processing associated with executing a specific policy, i.e. the cost of making a decision. The standard value iteration algorithm finds an optimal policy by computing first the optimal value function V^*:

$$V^*(s) = \max_{a} p(s'|s,a)[r(s,a,s') + \gamma V^*(s')] \quad (2)$$

starting from an arbitrary V and iterating Equation 2 until convergence (Bellman, 1957). From dynamic programming this iteration is known to converge to a global optimal solution, providing a numerical method for its computation (Sutton et al., 1999).

If the agents have independent controllers and their actions are interleaving, that is, if agent A makes a move on even and agent B on odd steps, the problem boils down to optimizing one agent's policy with respect to the other agent's policy. In the optimisation we use a static prediction for agent B's policy, which we scale in different trials from fully cooperative to completely antagonistic. Assuming a particular strategy of agent B, agent A optimizes its own action policy in order to achieve the goal under informational constraints.

The scenario involves two dependent MDPs that are not deterministic anymore and whose transition probabilities depend on a prediction of the other agent's action. In every iteration we use as a predictor the policy of the other agent. In a scenario where agents do not know anything about each other it is reasonable to set the predictor to a uniform distribution. In this study, we investigate how the behavior of one agent adapts to various strategies of the other agent.

5 OPTIMAL POLICIES

The relevant information represents the minimal level of information (mutual information between states and actions over all optimal policies) required for achieving a certain level of performance as characterised by the expected utility $E[U(S,A)]$, which is defined as:

$$E_\pi[U(S,A)] = \sum_{s,a} \pi(a|s)p(s)U(s,a).$$

The problem of informational parsimony can be formulated as search for a value-optimal strategy $\pi^*(a|s)$, which at the same time minimises the required relevant information $I(S;A)$, i.e. is also information-optimal. This double optimisation is transformed into an unconstrained minimisation problem via Lagrange multipliers:

$$\min_{\pi(a|s)} (I(S;A) - \beta E[U(S,A)]).$$

The β multiplier implicitly enforces the maximisation of $U(S;A)$. For $\beta \to \infty$ the optimisation restricts the policy to a value-optimal one. A reduction of β, however, makes the minimisation less sensitive to a drop in utility. This minimax optimisation problem corresponds to trading in utility for a reduction of relevant information.

To compute the policies corresponding to completely rational and completely irrational behavior we executed the Relevant Information algorithm with a single agent and two extreme β multipliers, $\beta \to -\infty$ and $\beta \to \infty$, as suggested in (Ortega and Braun, 2013), the former providing the antagonistic policy and the latter the cooperative one. In both cases the reward function was the same: 0 for the goal state and -1 elsewhere. The task is episodic and the agents act only until they reach the goal configuration.

To obtain a policy that is optimal and informationally parsimonious (Polani et al., 2006) extended the standard value iteration method to accommodate for the double optimisation by adding an interleaved Blahut-Arimoto iteration (Blahut, 1972) with the utility as a distortion measure for computing the strategy π' as follows:

$$p_k(a) = \sum_s p(s)\pi_k(a|s)$$

$$\pi_{k+1}(a|s) = \frac{1}{\zeta} p_k(a) \exp[\beta U^\pi(s,a)] \qquad (3)$$

where k denotes the iteration step, ζ is a normalising partitioning function and β a parameter trading off utility and relevant information.

The policy update iteration is alternated with a value iteration to get a consistent utility. This leads to the following iteration:

$$\pi \xrightarrow{(2)} V^\pi \xrightarrow{(1)} U^\pi \xrightarrow{(3)} \pi',$$

which generates a sequence of consistent policy–utility (π–U) pairs for a specific β, by interlacing the updates corresponding to Equations 1 and 3, and performing these iterations until convergence of both policy and utility.

This approach allows the computation of optimal strategies while trading off utility and relevant information. However, it does not address the cost involved with obtaining information, but focuses only on how much (relevant) information needs to be acquired (and processed) in order to achieve a certain level of utility, ignoring the possible acquisition cost. It specifies the amount of information an agent takes in and processes on average per time-step in the action selection process. This amount depends on the policy – different policies require different informational bandwidth, i.e. processing capacity. It is hypothesised that the required capacity is correlated to the metabolic cost of information processing and constitutes a quantitative measure of cognitive burden. The parsimony pressure tries to minimise this quantity while the efficiency pressure drives towards higher performance. Therefore for $\beta \to \infty$ the resulting optimal policy maximizes the expected utility and at the same time minimizes the mutual information $I(S;A)$.

Two extensions of relevant information to multiple steps provide similar algorithms unifying Value Iteration with Blahut-Arimoto for minimizing information quantities in Bayesian graphs under optimality constraints (Tishby and Polani, 2011).

To adapt the relevant information algorithm for the scenario involving two agents with independent controllers we introduce the following modification. As the agents alternate in taking actions, the policy optimization of agent A needs to take into account the policy of agent B, which we replace by a static prediction. On every iteration we combine the static prediction of agent B's policy with the current policy of agent A:

$$\tilde{\pi} \sim \pi_A + \pi_B$$

and use the resulting policy $\tilde{\pi}$ in the utility update, which provides the following modified iteration:

$$\pi_A \to V^{\tilde{\pi}} \to U^{\tilde{\pi}} \to \pi'_A.$$

This iteration optimizes the policy of agent A while taking into consideration the implications of the cooperative or antagonistic behavior of agent B in the shared environment.

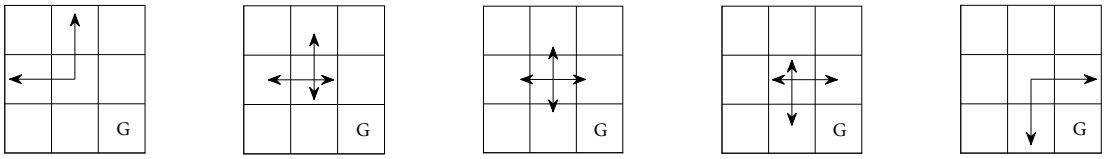

Figure 1: Various uniform strategies of agent B ranging from completely antagonistic to fully cooperative (left to right). G denotes the goal state and the arrow length represents the action probability, which is 0.5 at the extremes (left/right) and 0.25 in the middle.

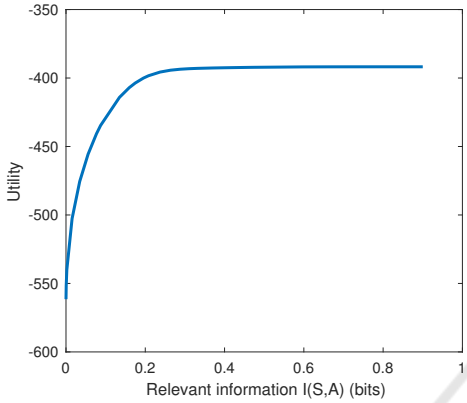

Figure 2: Utility vs. Relevant Information trade-off curve of agent A corresponding to the uniform policy of agent B. For different policies of agent B the curve has similar shape with specific offsets in utility levels.

6 INFORMATION-UTILITY TRADE-OFF

We computed the optimal policies of agent A for a range of antagonistic and cooperative behaviors of agent B (see Figure 1) and for multiple levels of the trade-off factor β. The level of collaboration of agent B (denoted with δ) varies in the range of (0,1) – where 0 corresponds to fully antagonistic and 1 to fully cooperative behavior (see Figure 1. Factor of 0.5 corresponds to a uniformly distributed policy over the action space (see Figure 1). As expected, the resulting utility vs. relevant information trade-off curves have similar shapes, with specific offsets in utility levels, increasing with the level of cooperation of agent B. For brevity, Figure 2 presents only the trade-off curve of agent A corresponding to the uniform policy of agent B (see Figure 1).

A theoretic upper limit on the amount of relevant information is given by the cardinality of the action space, which in this case is $\log|\mathcal{A}| = 2$ bits. However, the required information is below 1 bit on average (see Figure 2), since partitioning the state space by only two actions (right and down) provides an optimal solution. As β tends to 0 so is the amount of relevant information, since in every state of the grid a uniform action policy over the two actions (right and down) provides a solution, however with a lower utility.

Figure 5 reveals the change in the optimal policy of agent A corresponding to a uniform policy of agent B for two levels of β, low and high. The size of the arrows reflects the probability of dominant actions. For low β the (soft) policy is close to a uniform distribution over two actions (right and down), while for high β most states have deterministic action probabilities. A characteristic policy transition can be observed in the upper-left corner of the grid in the form of rectangle.

The simulated trajectories generated when applying the optimal policies of agent A for two levels of β – high (see Figure 3) and low (see Figure 4) – are demonstrated for five levels of collaboration. Figure 3 reveals how increasing cooperation (left to right) influences the optimal path of agent A. For antagonistic behaviors of agent B, agent A prefers to stick to the wall, which protects the object from moving back and decrease utility. As cooperation of agents increases, agent A gradually moves away from the wall towards the diagonal, which represents the optimal informationally parsimonious policy. In this particular scenario the collaboration rate plays the role of a trade-off factor between following the wall and the diagonal. Initially, the agent feels more confident by the wall before switching to the diagonal, and this period of initial uncertainty gets shorter as the agent can increasingly rely on its partner. Figure 4 reveals that the trends for lower β are similar with the typical blur, which relaxes the required level of relevant information. Interestingly, even with a lower pressure on the utility level, agent A still tends to stick to the wall, which shows how the embodiment in this scenario helps the agent tackle excess uncertainty and stabilize performance.

Figure 6 demonstrates similar shapes of the utility as a function of the collaboration rate for two levels of β (high and low). The normalized distance between the two utility curves reveals that increased collaboration rates emphasize the utility gap between different levels of β. This suggests that decreasing the collaboration level has a damping effect on the influence β has on utility, i.e., for antagonistic policies utility lev-

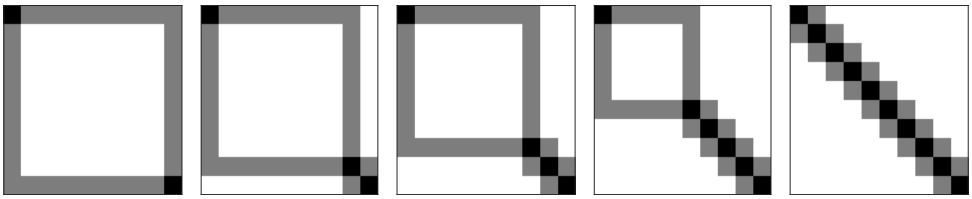

Figure 3: Simulated trajectories of agent A's optimal policies for five collaboration rates increasing from left to right (high β). Darker color denotes higher recurrence. For antagonistic behaviors of agent B agent A initially sticks to the wall to mitigate risk, before gradually moving towards the diagonal as cooperation increases.

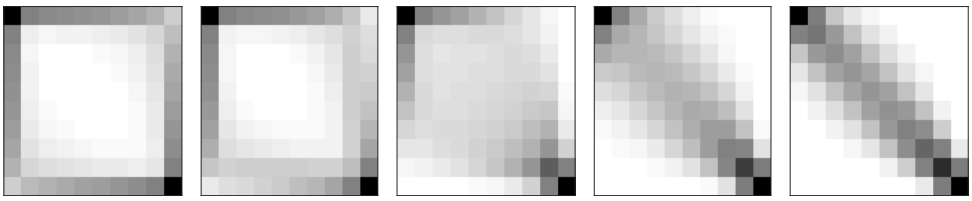

Figure 4: Simulated trajectories of agent A's optimal policies for five collaboration rates increasing from left to right (low β). Darker color denotes higher recurrence. Regardless of lower pressure on utility level agent A sticks to the wall to gain control over uncertainty, revealing the role of embodiment in this scenario.

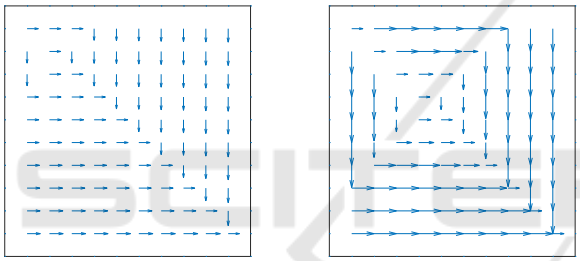

Figure 5: Optimal policies of agent A corresponding to a uniform policy of agent B for low (a) and high (b) β. Arrow size reflects dominant action probability. A policy transition propagates from the upper-left corner along the diagonal.

els are relatively close to each other across different β. This trend can also be observed in Figures 3/a and 4/a, which reveal similar trajectory paths.

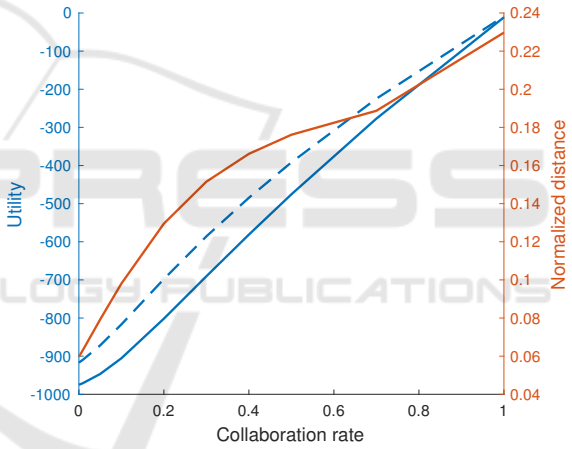

Figure 6: Utility as a function of collaboration rate for high (dashed) and low (solid) levels of β (blue). The normalized distance between these utility curves (red) reveals how higher collaboration emphasize the utility gap in β levels.

7 DISCUSSION

We explored a dyadic collaborative scenario in a 2-D grid-world, where two memoryless agents with independent controllers interact using fully redundant control. We applied a variation of the relevant information method to compute the optimal policies of one of the agents under information processing constraints. Our study investigated how the behavior of one agent influences the optimal strategy of the other in a range of different configurations. The results demonstrate that when the agents cooperate towards a common goal they achieve a higher utility at a lower informational cost. However, when their behavior is antagonistic to each other, the utility level drops significantly and the informational cost increases. One could think of various situations where this scenario could arise, for example, if the controllers are stochastic, or the precise knowledge of the other agent's model is unavailable or compromised, or the other agent is adverse or autistic.

The results reveal that when facing an antagonistic partner, the cooperative agent exploits the particular embodiment of its environment in order to maximize its control abilities. This suggests that the applied method of relevant information is able to identify salient features of the state space and benefit from them in order to provide an optimal informationally

parsimonious strategy in various environmental conditions. This initial study did not reveal further interplay between the informational cost and the rate of collaboration, however, future work could explore this trade-off with more elaborate and focused scenarios. The optimal policies derived in our study operate with relevant information intake in the range of [0,1] bits on average per step, which is considerably lower than the theoretical limit of 2 bits, suggesting that a different embodiment could result in a higher information burden and provide deeper insights.

In this particular scenario, our approach highlights two types of optimal trajectories, one following the walls and another one traversing the diagonal, and trades off these two paths on the base of cooperation level. Furthermore, the results demonstrate that high cooperation rates (δ) emphasize the role of β in reinforcing high performance levels, which suggests a specific trade-off between δ and β and is an interesting topic for future research.

The key benefits of the applied information-theoretic treatment are that it is universal, general and could enable the direct comparison of scenarios with different computational models. The proposed approach reframes the problem into a trade-off between the reward achieved and the informational cost of performing a task by incorporating limits on the information processing capacity, which are fundamental properties of agent–environment systems. The framework can be further extended to tackle issues of time shifts and turn-taking in the decision process.

8 CONCLUSION

This paper presents an application of an information-theoretic framework to a reward driven decision process in the perception–action cycle of a dyad. We demonstrate interesting behaviors generated in robotic agents based on self-organization following the principle of homeokinesis. Driven by this principle through the interaction with its environment, the agent shows preferences for states where its control actions are most effective in avoiding unpredictable and undesired situations. We believe that theories and tools from complex systems and information theory may successfully be applied in the future for facilitating the automated design of robot collectives and for the analysis of their dynamics. The proposed approach could provide a framework supporting the creation of artificial agents, which not only act optimally, but optimize also their computational resources in the decision-making process. Further work is required to evolve this methodology into the continuous action domain and showcase its application in more realistic practical scenarios.

ACKNOWLEDGEMENTS

The authors would like to acknowledge support by H2020-641321 socSMCs FET Proactive and FFG-6112792 Pro^{2}Future projects.

REFERENCES

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

Bellman, R. E. (1957). *Dynamic Programming*. Princeton University Press.

Blahut, R. (1972). Computation of channel capacity and rate distortion functions. *IEEE Transactions on Information Theory*, 18(4):460–473.

Boedecker, J., Obst, O., Lizier, J., Mayer, N., and Asada, M. (2011). Information processing in echo state networks at the edge of chaos. *Theory in biosciences*, 131:205–13.

Butail, S., Mwaffo, V., and Porfiri, M. (2016). Model-free information-theoretic approach to infer leadership in pairs of zebrafish. *Physical Review E*, 93.

Callebert, L., Lourdeaux, D., and Barthès, J.-P. (2016). A trust-based decision-making approach applied to agents in collaborative environments. In *Proceedings of the 8th International Conference on Agents and Artificial Intelligence*, pages 287–295.

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

Challet, D. and Zhang, Y.-C. (1997). Emergence of cooperation and organization in an evolutionary game. *Physica A: Statistical Mechanics and its Applications*, 246(3):407 – 418.

Couzin, I., Krause, J., Franks, N., and Levin, S. (2005). Effective leadership and decision-making in animal groups on the move. *Nature*, 433 (7025):513–516.

Davis, N., Hsiao, C., Singh, K. Y., Lin, B., and Magerko, B. (2017). Quantifying collaboration with a co-creative drawing agent. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 7:1–25.

Franks, N., Pratt, S., Mallon, E., Britton, N., and Sumpter, D. (2002). Information flow, opinion-polling and collective intelligence in house-hunting social insects. *Philosophical Transactions B: Biological Sciences*, 357 (1427):1567–1583.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature reviews. Neuroscience*, 11:127–38.

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

Harder, M., Polani, D., and Nehaniv, C. L. (2010). Two agents acting as one. In *Artificial Life*, pages 599–606.

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

Kim, H., Davies, P., and Walker, S. (2015). New scaling relation for information transfer in biological networks. *Journal of the Royal Society, Interface*, 12(113):20150944.

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.

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

Lindner, M., Vicente, R., Priesemann, V., and Wibral, M. (2011). Trentool: A matlab open source toolbox to analyse information flow in time series data with transfer entropy. *BMC neuroscience*, 12:119.

Lizier, J. T., Prokopenko, M., and Zomaya, A. Y. (2014). *A Framework for the Local Information Dynamics of Distributed Computation in Complex Systems*, pages 115–158.

Marshall, J. A. R., Bogacz, R., Dornhaus, A., Planqué, R., Kovacs, T., and Franks, N. R. (2009). On optimal decision-making in brains and social insect colonies. *Journal of the Royal Society, Interface*, 6(40):1065—1074.

Martius, G., Nolfi, S., and Herrmann, J. (2008). Emergence of interaction among adaptive agents. In *10th International Conference on Simulation of Adaptive Behavior*, pages 457–466.

Matsui, T. (2019). A study of joint policies considering bottlenecks and fairness. In *Proceedings of the 11th International Conference on Agents and Artificial Intelligence*, pages 80–90.

Meyer, B. (2017). Optimal information transfer and stochastic resonance in collective decision making. *Swarm Intelligence*, 11:1–24.

Meyer, J. A. and Wilson, S. W. (1991). *The Dynamics of Collective Sorting Robot - Like Ants And Ant - Like Robots*, pages 356–363.

Moore, D. G., Valentini, G., Walker, S. I., and Levin, M. (2018). Inform: Efficient information-theoretic analysis of collective behaviors. *Frontiers in Robotics and AI*, 5:60.

Mwaffo, V., Butail, S., and Porfiri, M. (2017). Analysis of pairwise interactions in a maximum likelihood sense to identify leaders in a group. *Frontiers in Robotics and AI*, 4:35.

Nabet, B., Leonard, N., Couzin, I., and Levin, S. (2009). Dynamics of decision making in animal group motion. journal of nonlinear science. *Proceedings of the National Academy of Sciences of the United States of America*, 19(4):399–435.

Ortega, P. and Braun, D. (2013). Thermodynamics as a theory of decision-making with information-processing costs. *Royal Society: Mathematical, Physical and Engineering Sciences*, 469(2153).

Paolo, E. D., 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.

Pezzulo, G. and Levin, M. (2015). Re-membering the body: applications of computational neuroscience to the top-down control of regeneration of limbs and other complex organs. *Integr. Biol.*, 7:1487–1517.

Polani, D., Martinetz, T., and Kim, J. T. (2001). An information-theoretic approach for the quantification of relevance. *6th European Conference on Advances in Artificial Life*, pages 704–713.

Polani, D., Nehaniv, C., Martinetz, T., and Kim, J. T. (2006). Relevant information in optimized persistence vs. progeny strategies. In *10th International Conference on the Simulation and Synthesis of Living Systems*, pages 337–343. MIT Press.

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.

Roli, A., Ligot, A., and Birattari, M. (2019). Complexity measures: Open questions and novel opportunities in the automatic design and analysis of robot swarms. *Frontiers in Robotics and AI*, 6:130.

Schreiber, T. (2000). Measuring information transfer. *Phys. Rev. Lett.*, 85(2):461–464.

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. (1949). *The mathematical theory of communication*. The University of Illinois Press, Urbana.

Sperati, V., Trianni, V., and Nolfi, S. (2011). Self-organised path formation in a swarm of robots. *Swarm Intelligence*, 5:97–119.

Sutton, R., Precup, D., and Singh, S. (1999). A framework for temporal abstraction in reinforcement learning. *Artificial intelligence*, 112(1-2):181–211.

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. (2011). Information theory of decisions and actions. In Cutsuridis, V., Hussain, A., and Taylor, J., editors, *Perception-Action Cycle: Models, Architecture and Hardware*, pages 601–636. Springer.

Touchette, H. and Lloyd, S. (2000). Information-theoretic limits of control. *Phys. Rev. Lett.*, 84(6):1156–1159.

Valentini, G., Moore, D. G., Hanson, J. R., Pavlic, T. P., Pratt, S. C., and Walker, S. I. (2018). Transfer of information in collective decisions by artificial agents. *Artificial Life Conference Proceedings*, (30):641–648.

Walker, S., Cisneros, L., and Davies, P. (2012). Evolutionary transitions and top-down causation. *13th International Conference on the Simulation and Synthesis of Living Systems*, pages 283–290.

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

Zenil, H., Marshall, J. A. R., and Tegnér, J. (2015). Approximations of algorithmic and structural complexity validate cognitive-behavioural experimental results. *CoRR*, abs/1509.06338.