

Empowerment and State-dependent Noise - An Intrinsic Motivation for Avoiding Unpredictable Agents

Christoph Salge¹, Cornelius Glackin¹ and Daniel Polani¹

¹University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK
c.salge@herts.ac.uk

Abstract

Empowerment is a recently introduced intrinsic motivation algorithm based on the embodiment of an agent and the dynamics of the world the agent is situated in. Computed as the channel capacity from an agent's actuators to an agent's sensors, it offers a quantitative measure of how much an agent is in control of the world it can perceive. In this paper, we expand the approximation of empowerment as a Gaussian linear channel to compute empowerment based on the covariance matrix between actuators and sensors, incorporating state-dependent noise. This allows for the first time the study of continuous systems with several agents. We found that if the behaviour of another agent cannot be predicted accurately, then interacting with that agent will decrease the empowerment of the original agent. This leads to behaviour realizing collision avoidance with other agents, purely from maximising an agent's empowerment.

Introduction

One important and unique aspect of living organisms is how they generate their behaviour. Sims (1994) demonstrated that simple motivations can be enough to generate complex behaviour that evokes a resemblance of life. Ultimately all organisms are subject to evolution and their behaviour is a product or by-product of a process directed by reproductive fitness and survival. However, from a cognitive perspective, it seems difficult for an agent to always relate behaviour back to survival. From an evolutionary perspective, it is also questionable how the sparse sampling of random behaviours could lead to good solutions. Nature solves this problem with the development of behavioural proxies or motivations (Scott-Phillips et al., 2011), such as the ability to perceive and avoid pain, which produces behaviour considered beneficial for survival. In artificial life the corresponding research aims to identify, quantify and replicate these motivations.

Significant research interest has been directed at methods known as "intrinsic motivations", methodologies to generate behaviours for agents without the requirement of an externally specified reward or utility structure; importantly, they emerge exclusively from the agent-environment dynamics. Here, instead of a specific goal, the generation of behaviour

depends on an internal motivation. Most of them focus on learning and exploration, and try to quantify an organism's urge to understand its environment (Schmidhuber, 1991; Der et al., 1999; Steels, 2004; Prokopenko et al., 2006; Ay et al., 2008).

In this paper we focus on one of these methods, which is based on *empowerment* (Klyubin et al., 2008). Empowerment provides a "universal utility", i.e. a utility landscape over the state space of an agent which is defined purely by the agent-world dynamics. In contrast to other methods it does not focus on learning or exploration, but identifies preferable states in a known local environment. Empowerment considers the probabilistic map from a sequence of the agent's actions to a world state resulting from these actions as a channel; empowerment is then formally defined as the Shannon (1948) channel capacity of this channel. Essentially, empowerment is an information-theoretic generalization of the control-theoretic concept of controllability (Touchette and Lloyd, 2000).

The basic motivation behind the empowerment concept is that it is preferable to be in a state where the agent's actions have the largest influence on the perceivable world, or *Umwelt* (von Uexküll, 1909), of the agent. From an empowerment perspective, the ideal state to be in is one that offers a high number of choices that all lead to different outcomes that can be causally (and predictably) distinguished (i.e. controlled) by the agent. States to avoid are those where noise interferes with the influence of agent actions on its resulting state (lack of controllability), and those where the agent can only reach a low number of possible resulting states through its actions (lack of reachability). If we are dealing with a deterministic system, where each action leads to one specific outcome, then the criterion reduces to pure reachability. An example for the latter special case is given by Klyubin et al. (2005a) where he demonstrates the relationship between average distance and empowerment in a grid-world maze.

In this paper we continue prior work (Salge et al., 2012) which provides a fast approximation of empowerment using a Gaussian linear channel. Here, we expand this method to

allow empowerment computation based on the covariance matrix between actuators and sensors, which, importantly, allows for the incorporating of state-dependent noise. This provides us with an appropriate and fast empowerment algorithm for the study of continuous systems with several agents, which was not possible before. We will outline the modification for the approximation method, and then use it to examine the empowerment in a simple, continuous multi-agent system.

Related Work

While many forms of adaptation and learning require some external goal-orientated supervision, critique, or perspective, it is now well understood, that a focus on embodiment (Pfeifer et al., 2007) provides a vehicle for self-determination which does not necessitate such external goals. Based on this, recent efforts have been made to understand agent control and agent motivation in the framework of the perception-action loop (Lungarella et al., 2005; Bertschinger et al., 2008), see also Fig. 1.

For example, Homeokinesis (Der et al., 1999) is a predictive methodology which adapts its perception-action loop on the fly, and drives an embodied agent to exploit its own embodiment to generate movement. It is related to other intrinsic motivation methods, such as “artificial curiosity” (Schmidhuber, 1991), or the “autotelic principle” (Steels, 2004), where the agent becomes self motivated when the challenges it faces are adequate for its skill level, inspired by the concept of “flow” (Csikszentmihalyi, 2000) for humans.

More recent developments use the framework of predictive information to produce intrinsically motivated robot control, generating behaviour only by looking at how specific actuator inputs change the agents sensor state (Prokopenko et al., 2006; Bialek et al., 2001; Ay et al., 2008). This is similar to the idea of empowerment, which is also fully defined by the channel between an agent’s actuators and sensors. One key difference though, is that empowerment provides a utility landscape by assigning a value to each state of the environment, where previous approaches are focussed on producing specific actions. Of special interest to the topic of this paper is also the work by Capdepuy et al. (2012), where he studies how empowerment is limited when several agents are using a joint channel, and what restriction this applies to agent coordination.

Formalism

Given a perception-action-loop, as seen in Fig. 1, Klyubin et al. (2005a) defined empowerment as the channel capacity (Shannon, 1948) between an agent’s actuators A and sensors S , with the world being in state $r \in R$. Each state $r \in R$ has its own empowerment value, which only depends on the

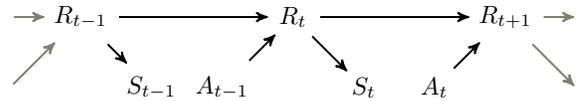


Figure 1: The perception-action-loop, unrolled in time t , visualised as a Bayesian network. The random variable S is the sensor of an agent; A is the actuator of an agent, and R represents the rest of the system.

channel between A and S in that state.

$$\mathfrak{E}(r_t) := C(p(s_{t+1}|a_t, r_t)) = \max_{p(a_t)} I(S_{t+1}; A_t | r_t). \quad (1)$$

Similarly, a sequence of n actions can be considered (called n -step empowerment), where the action-sequences are treated as a vector of random variables. The sensor state $s \in S$ is then usually further in the future, and can also be a vector of random variables.

Continuous Empowerment

Empowerment is defined for both discrete and continuous variables, but while it is possible to determine the channel capacity for the discrete case (for example by using the Blahut-Arimoto Algorithm (Blahut, 1972)), this is not generally possible for the continuous case. Jung et al. (2011) introduces a technique called Monte-Carlo Integration to approximate empowerment, but this method is very computationally expensive.

A faster method (described in detail by Salge et al. (2012)) to approximate the empowerment of a continuous channel is to treat it as a linear channel with added independent and identically distributed (i.i.d.) Gaussian noise.

$$S = TA + Z, \quad (2)$$

where S is an m -dimensional, continuous random variable, A is an n -dimensional, continuous random variable, T is a linear mapping, represented by a $m \times n$ matrix, and $Z_i \sim \mathcal{N}(0, N_i)$, with $i = 1, \dots, m$, is another multi-dimensional, continuous, i.i.d. random variable, modelling isotropic noise in the sensor dimensions.

Assuming that there is a power constraint $E(A^2) \leq P$ (without it the channel capacity would be arbitrarily large), this can be solved (Telatar, 1999) by applying a Singular Value Decomposition (SVD) to the transformation matrix T . The resulting singular values σ_i are then used to compute the channel capacity via the water-filling algorithm, as if this was a parallel Gaussian channel (Cover and Thomas, 1991). The channel capacity is then

$$C = \max_{P_i} \sum_i \frac{1}{2} \log(1 + \sigma_i P_i), \quad (3)$$

where P_i is average power used in the i -th channel, following the constraint that $\sum_i P_i \leq P$. As the channel capacity achieving distribution is a Gaussian distribution, this means the optimal input distribution is a Gaussian with a variance of P_i for each channel.

State-Dependent Noise

Salge et al. (2012) use simplifications that are only possible because the model's noise Z is assumed to be i.i.d distributed with a fixed variance. This forces the previous algorithm to assume the same level of noise for every state of the environment, and also makes it unable to model coloured, i.e. covariate, noise.

To address this problem we used the covariance matrix between actuators and sensors to capture the relationship between them, as well as the current noise level, and then reduced this problem to a parallel Gaussian Channel with i.i.d noise with the same capacity. First we chose n actuator variables A_1, \dots, A_n , and m sensor variables S_1, \dots, S_m . Now we determine the covariance matrix K between all these values. In our example, this is done by computing the pairwise covariance between sampled values for each of these variables. Alternatively, one could use the covariance function $k(\cdot, \cdot)$ of a Gaussian Process that models the system to obtain the covariance matrix (not done in this experiment).

$$K = \begin{pmatrix} k(a_1, a_1) & \dots & k(a_1, a_n) & k(a_1, s_1) & \dots & k(a_1, s_m) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ k(a_n, a_1) & \dots & k(a_n, a_n) & k(a_n, s_1) & \dots & k(a_n, s_m) \\ k(s_1, a_1) & \dots & k(s_1, a_n) & k(s_1, s_1) & \dots & k(s_1, s_m) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ k(s_m, a_1) & \dots & k(s_m, a_n) & k(s_m, s_1) & \dots & k(s_m, s_m) \end{pmatrix} \quad (4)$$

$$= \begin{pmatrix} K_{a,a} & K_{a,s} \\ K_{s,a} & K_{s,s} \end{pmatrix}. \quad (5)$$

Now, if the variable $A = A_1, \dots, A_n$ assumes a concrete value $a = a_1, \dots, a_n$, then this results in a specific, multivariate Gaussian distribution for $S = S_1, \dots, S_m$, with

$$S = \mathcal{N}(\mu_s, K_s). \quad (6)$$

Note that $K_s \neq K_{s,s}$. K_s can be computed (Rasmussen and Williams, 2006) as:

$$K_s = K_{a,a} - (K_{s,a} K_{s,s}^{-1} K_{a,s}). \quad (7)$$

Assuming that the mean of actuator distributions of A is zero¹ we can also determine the mean for S given as a specific value of a as

$$\mu_s = K_{s,a} K_{s,s}^{-1} a. \quad (8)$$

We see from Equ. 7 that the covariance only depends on the original covariance matrix, and not on the actual value

¹If the mean of a distribution is not zero it can be shifted without affecting the mutual information

of a . Also, from Equ. 8 we see that the new mean of the distributions is a linear transformation of a , with the matrix $K_{s,a} K_{s,s}^{-1} = T'$. So, a variation of a affects the mean of the resulting distribution of S , but not its covariance.

As a result, the relationship between S and A , as modelled by the covariance matrix, can be expressed as a linear, multiple input, multiple output channel with added coloured noise as

$$S = T'A + Z', \quad (9)$$

with $Z' \sim \mathcal{N}(0, K_s)$. Note, that there is no approximation in this step, the linear channel fully captures the dynamics of the system that are still present in the covariance matrix.

This can be further reduced to a channel with i.i.d. noise. For this, note that rotation, translation and scaling operators do not affect the mutual information $I(S; A)$. We start by expressing Z' as

$$Z' = U\sqrt{\Sigma}ZV^T, \quad (10)$$

where $Z \sim \mathcal{N}(0, I)$ is isotropic noise with a variance of 1, and $U\sqrt{\Sigma}V^T = K_s$ is the SVD of K_s . U and V^T are orthogonal matrices, and Σ contains the singular values. The square roots of the singular values scale the isotropic noise to the right variance; the noise is then rotated to resemble the original coloured noise. Note, that all singular values have to be strictly larger than zero, otherwise there would be a channel in the system without noise, which would lead to infinite channel capacity. Thus, we can consider $\sqrt{\Sigma}^{-1}$, a diagonal matrix with entries which are the inverse of the singular values in $\sqrt{\Sigma}$. This allows us to reformulate:

$$S = T'A + U\sqrt{\Sigma}ZV^T \quad (11)$$

$$U^T S V = U^T T' A V + \sqrt{\Sigma} Z \quad (12)$$

$$\sqrt{\Sigma}^{-1} U^T S V = \sqrt{\Sigma}^{-1} U^T T' A V + Z \quad (13)$$

$$\sqrt{\Sigma}^{-1} U^T S = \sqrt{\Sigma}^{-1} U^T T' A + Z V^T \quad (14)$$

$$\sqrt{\Sigma}^{-1} U^T S = \sqrt{\Sigma}^{-1} U^T T' A + Z \quad (15)$$

The last step follows from the fact that the rotation of isotropic Gaussian noise is isotropic Gaussian noise. This reduces the whole problem to a MIMO channel with isotropic noise with the same channel capacity. We simply define the transformation matrix T used in $S = TA + Z$ as

$$T = \sqrt{\Sigma}^{-1} U^T T', \quad (16)$$

and apply the solution outlined for the simpler channel. This reduction allows the fast approximation of empowerment based on the covariance matrix between actuators and sensors, which can be either obtained via sampling of the environment, or by relying on a Gaussian Process Learner. This allows us to model the actual noise present in different states of the environment, which is then represented in the modified T .

Model

We now apply this new method to a simple, continuous multi-agent system. The other agents are introduced to provide a changing level of noise in the environment. The covariance based empowerment approximation allows us to study how an agent would deal with different levels of noise.

World Model

The model is a continuous, flat, two-dimensional world, populated by circular agents. Each agent has a radius $l = 0.15$ meters². Each agent is defined by:

- a position, stored in a vector q , which contains a real-valued x and y coordinate,
- a speed \dot{q} , expressing the change in x and y per second,
- a direction d the agent is currently facing, measured in radians, where 0 means the agent is heading north.

Actuation

At the beginning of each time step, the agent has the choice to turn within its turning radius, which is 45 degrees, or one-eighth of a circle. The agent chooses a real value between -1.0 (turning 45 degrees counter-clockwise) and 1.0 (turning 45 degrees clockwise); 0.0 means the agent maintains its current heading. The agent turns instantaneously and then, for the duration of the time step, the agent continuously accelerates at 0.03 m/s². In our model the agent will accelerate at full power, the only choice is the direction of acceleration.

Simulation

The only other acceleration force that applies to agents arises from collisions with other agents. Whenever the distance between two agents becomes less than the sum of their radii, a collision occurs. This is modelled as an elastic collision, so the agents can come closer than this distance, but will be subject to linearly increasing acceleration away from the center of mass of the other agent. The acceleration from the collision for the first agent can be computed as

$$\ddot{q}_c = \max(0, |l_i + l_j| - |q_1 - q_2|) \cdot (q_1 - q_2) \cdot c, \quad (17)$$

where c is a constant that determines how hard the elastic collision is. For lower c , colliding agents move further into each other before they bounce apart. Furthermore, to keep the velocity of the agents limited, there is a constant amount of friction applied to the agents. At each time step agents lose 5% of their velocity.

The progress of the model through time is simulated by breaking each time step into 20 pieces of equal length, and for each of those an appropriate fraction of the acceleration

²For ease of notation the unit length will be called meter, and the length of a time step will be one second

of the agent is added to its speed, and then the speed is added to the agent's current position. This is equivalent to explicit Euler integration.

Note that this model allows slip, i.e. an agent can head in one direction (where it is also accelerating to), while moving in a different direction. Turning does not change the current inertial movement.

Experiments

Hypothesis

Preliminary observations of the agent's behaviour indicated that an increase in the chance of a future collision with other agents is accompanied by a reduction in the current empowerment value. Therefore, our hypothesis for this simulation is that since the behaviour of other agents cannot be predicted, they will act as a source of noise in the environment, and colliding with them would be detrimental for an agent's empowerment.

Different Scenarios

To test this hypothesis, and evaluate it systematically, we set up three different scenarios. In each scenario there are two agents. For the first agent we measure the empowerment and collision chance at different starting coordinates, located between -1 and 1, both for the x and y coordinate. The first agent is always heading north-east, and starts with a speed of 0.03 m/s in that direction. The second agent is always located in position (0.5,0.5) and is heading south-west with a speed of 0.03 m/s. The three scenarios differ in the behaviour of the second agent:

Unpredictable Agent: The second agent chooses actions uniformly random at the beginning of each time step, turning within its possible turn radius. The choice of the second agent cannot be predicted by the first.

Predictable Agent: The second agent always chooses to maintain its current direction, i.e. it moves in a straight line. This is known to the first agent, and incorporated into its model.

Immovable Agent: The second agent is anchored to its position, essentially constituting a fixed obstacle. It still reflects other agents colliding with it.

Note that the term agent here is used loosely as a "catch-all" term for other objects in the environment, which could be agents, movable objects or just fixed obstacles.

Measurements

We computed the 4-step empowerment for Agent 1 for the three scenarios, for different starting positions. So, the actuation variables a_1, \dots, a_4 denote what action Agent 1 chooses at the beginning of the first, second, third and fourth

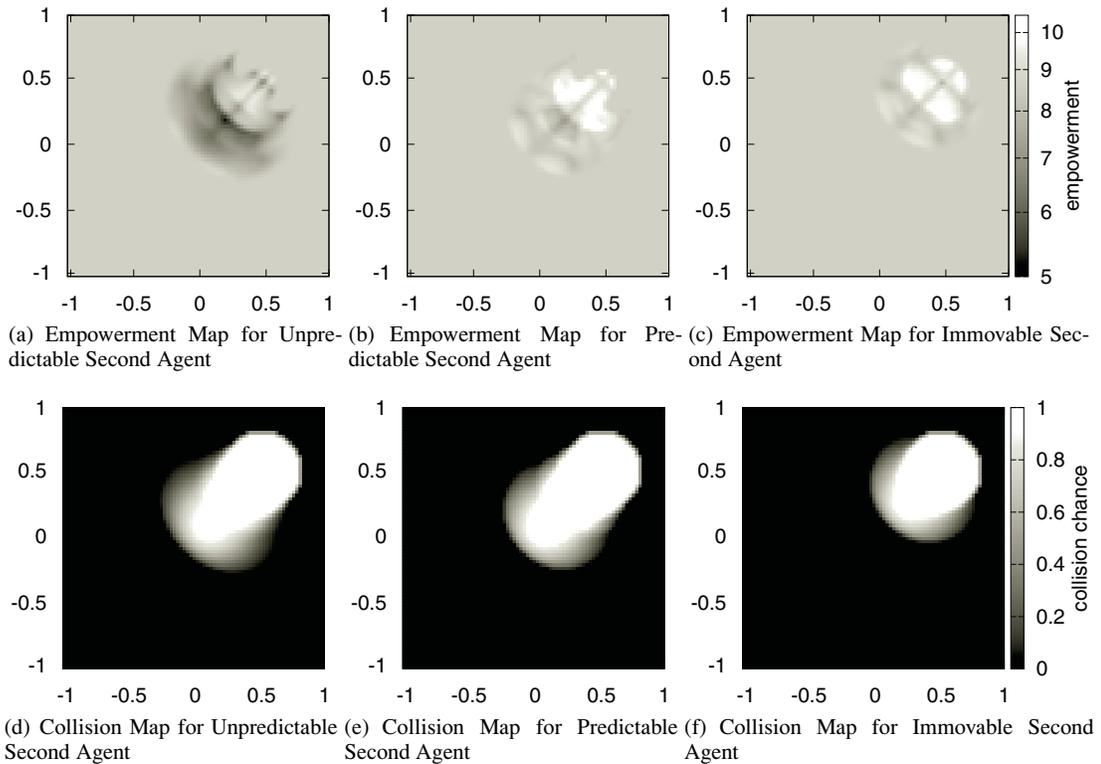


Figure 2: Plots of the empowerment and collision probability for different starting positions of the first agent. The Figs. 2(d)-2(f) show the fraction of action sequences that lead to a collision between the agents. Figs. 2(a)-2(c) show the empowerment of the first agent. The second agent is always located at position (0.5,0.5). Initial heading for the first agent is north-east, for the second agent it is south-west.

time step, respectively. The sensor input considered for empowerment were the values of x, y, \dot{x}, \dot{y} after the fourth time step, so the speed and position of the agent after the actuation sequence has been executed.

For each starting position for Agent 1, we used the actual simulation model to create an amount of samples, consisting of actuator variables and resulting sensor values. We used regular sampling, so that each time step the agent could only choose the 5 values of $\{-1.0, -0.5, 0.0, 0.5, 1.0\}$, leading to $5^4 = 625$ possible action sequences in four steps. Each of these sequences was then simulated 10 times, leading to 6250 samples overall.

In the “unpredictable agent” scenario the action of the other agent was chosen uniformly random for those simulations, whereas in the predictable and unmovable agent scenario the simulation “knew” what the other agent would do, leading to a predictable outcome for each action sequence. So the 10 repeated samplings of the same action sequence only led to different results in the unpredictable agent scenario, since the pseudo-random generator would potentially chose different actions for the second agent. The resulting 8 times 6250 values were then used to pairwise compute a

covariance matrix between all 8 values (4 actuation values, 4 resulting sensor values), which was in turn used to compute the channel capacity from the actuation variables to the sensor values. This allowed us to compute empowerment of Agent 1 for different starting positions. We also recorded, for each starting position, what percentage of the sampled action sequences would lead to a collision with the second agent within the first four time steps.

Results

The results demonstrate that colliding with the unpredictable agent leads to a substantial loss in empowerment, compared with the other scenarios. Consider first the collision maps depicted in Figs. 2(d)-2(f), which show what fraction of the action sequences in a given starting position leads to a collision between the agents. This segments the space of starting positions for agent one into three areas.

The area with zero collision probability are all locations where there is no chance for the agents to interact. These areas are thus of little interest for our central hypothesis. The empowerment landscape in these areas is constant, as expected for an unstructured environment. We will consider

this constant value to be the baseline value of empowerment for comparison.

The second area are those starting locations where the agents always collide. This is mainly a circle of diameter 0.3 m around $(0.5,0.5)$, where the agents already start in collision, and a connected area where the agents start separated but are moving towards each other. Towards the center of the circle the agents overlap the most, and there are several areas of higher empowerment. This results from the specific collision mechanics in our simulation. As we model near-physical elastic collision, agents who overlap can be considered as “storing potential energy”, to apply a physical analogy. This high potential energy allows faster acceleration, which allows the agent to reach a greater variety of locations. The analogy here would be riding a bike, both on a flat surface, and starting on top of a hill. The extra speed gained riding downhill allows the rider to reach a greater variety of locations. Similarly, Agent 1 can control, in part, where this extra acceleration moves it, resulting in greater empowerment for Agent 1. We can also see that this effect is greater in Fig. 2(b) and Fig. 2(c), since the agent here can fully predict where this extra acceleration will lead it. In Fig. 2(a) this effect is less powerful, as the unpredictable movement of the second agent makes it harder to predict where the first agent will end up, thereby lowering the empowerment of Agent 1.

The most interesting area for our hypothesis are those locations where it is uncertain whether the agents collide or not. These areas are located further to the south-west of the second agent. So a first agent starting here is moving towards the second agent, but is far enough away that some actions might lead to an avoidance of collision. Now, the outcome of different action sequences not only depends on the first agent’s actions, but also on what the second agent does. If the second agent is unpredictable, i.e. moves at random, then the simulation done by the first agent will result in different outcomes for the same actions, introducing noise to the channel between actuators and sensors. This causes a measurable increase in uncertainty. The effect of this can be seen in Fig. 2(d). If we now compare this with the predictable second agent scenario, we see much less empowerment reduction with a predictable agent. While the second agent might still block access to some locations, it does not introduce noise into the outcome. Thereby, a collision with the second agent does not reduce the empowerment as much, as seen in Fig.2(b). Similarly for the immovable agent, the empowerment here is only slightly reduced when an agent is on a collision course. So, the main cause for the drop of empowerment in our model is not the *collision* with a second agent, but the *collision* with an *unpredictable* agent.

Empowerment Control

The difference between a predictable and unpredictable agent becomes even clearer if we look at the resulting agent

control. We implemented a greedy empowerment maximisation control. From five candidate actions $(-1,-0.5,0,0.5,1)$ it picks the action that *leads to* the state with the largest empowerment value. For this, 4-step empowerment is calculated for all five states resulting from the candidate actions.

Fig.3 shows the resulting trajectories of the first agent for different starting positions. Once a collision with the second agent occurs, the line becomes dashed. In both figures the first agent selects actions that maximise its empowerment for the next step; the only difference is the behaviour of the second agent. Both second agents start heading towards the first agent, but the second agent in Fig.3(a) just moves straight, while the second agent in Fig.3(b) moves at random. So in Fig.3(b) the simulations of the first agent to determine the empowerment of possible future states cannot accurately predict the second agent. This means that the possibility to interact with the second agent becomes a source of noise, and empowerment maximisation avoids actions that lead to trajectories where the possibility of interaction with the second agent might arise. As a result, only three of the resulting trajectories collide with the second unpredictable agent. In the other case, shown in Fig.3(a), empowerment sees no problem with colliding with the second agent, as it does not introduce noise into its action-perception channel and therefore it permits a lot of trajectories to end up in collisions.

Discussion

The specific model we are considering here results in a collision avoidance behaviour regarding the second agent, if that agent is unpredictable. While it could be argued that other agents tend to be hard to predict, and therefore the assumption that other agents introduce noise is reasonable, we emphasize that the aim here was *not* to specifically produce obstacle avoidance. Also, the specific behaviour of empowerment depends on how the environment is modelled. If collisions result in loss of velocity, or even loss of actuation possibilities (like broken motors), then empowerment would avoid any collision. Even if this was not the case, if obstacle avoidance was desired, one could ‘spike’ the system dynamics as to make the agent believe that it would break down when it collides, which would induce strict obstacle avoidance. This would correspond to ‘programming’ or, rather, ‘nudging’ the empowerment-based behaviour engine towards desirable behaviours. Importantly, if there were an explicit goal of obstacle avoidance we would not contest that explicit obstacle avoidance algorithms would be superior. The advantage of an empowerment controlled agent lies in its universality. The same algorithm that avoids collision also balances an inverted pendulum (Jung et al., 2011), finds central positions in a maze (Klyubin et al., 2005b), can be used to adapt sensors (Klyubin et al., 2005a), and leads to the formation of patterns in multi-agent simulations (Capdepuuy et al., 2007).

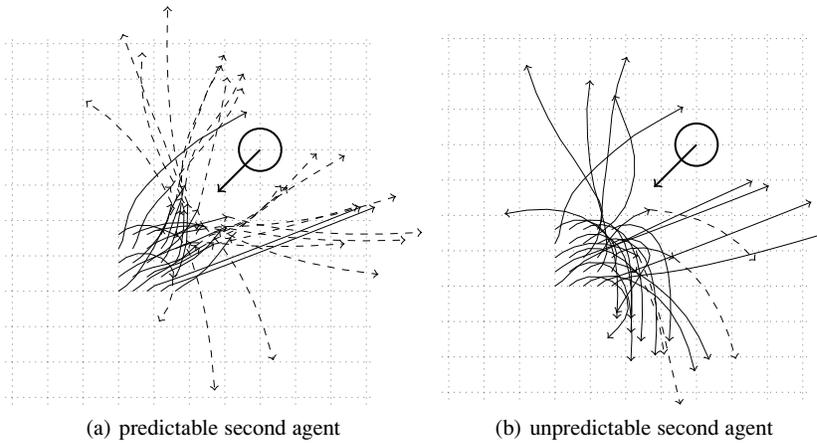


Figure 3: A comparison plot for the behaviour of an empowerment maximising agent. The arrows trace the trajectories of the empowerment controlled agent; each arrow is a different simulation, starting from a slightly different point. The initial heading of the empowerment agent is north-east. The circle indicates the starting position of the other agent, its initial heading is south-west. The lines become dashed if a collision between the agents occurs. Fig. 3(a) has a predictable second agent, which just moves straight ahead. Therefore, empowerment sees no need to avoid, and most trajectories lead to collisions. Fig. 3(b) has a second agent that chooses random actions. It is therefore a source of noise and the empowerment driven agent avoids colliding with it in most cases.

The point of using the intrinsic, empowerment-based behaviour is that it is more generic and grounded in the agent-system dynamics, and incorporates implicit difficulties that the agent may encounter. Thus, instead of imposing explicit conditions on when to activate a sub-behaviour, such as obstacle avoidance, one could incorporate desired hard behaviours, where required, into “surrogate” modifications of the physics of the system and let the empowerment-based behaviour engine generate the behaviours based on these modifications, whilst leaving the natural dynamics of the system unchanged for all other situations. Such an approach may be able to provide the agent with a more flexible repertoire of options, whilst respecting required hard constraints. It would also take a step towards “implicit programming”.

In terms of applications, it would also be interesting to see how an empowerment-based system would deal with a navigation task in a crowded environment, such as walking down Oxford Street at prime shopping time. In general, one should avoid colliding with people, but one could speculate that understanding how another person is going to move would allow an agent to operate closer to that person, with less chance of collision, and therefore less loss of empowerment. This would require the introduction of separate models for different agents, which would then allow an agent to model how predictable another agent is, and consequently adjust its behaviour towards different agents. Note that in this hypothetical example empowerment is clearly computed based on the agent model of the world. It does not matter how predetermined another agent’s behaviours are, but how well this can be predicted by the internal model.

More generally, we see that empowerment depends on the agent’s internal model of the world. Reducing the uncertainty in one’s internal model increases empowerment, which then raises the question, how suited is empowerment for exploration? We speculate that this depends on the horizon of the empowerment optimization. In the short term, interacting with another unpredictable agent will be detrimental, and avoiding it will preserve an agent’s empowerment. However, in the long term interacting with another agent might increase the predictability of said agent (by virtue of learning a better model of the agent), and this will improve the empowerment of the first agent on subsequent later interaction. This also indicates a more general distinction between different sources of noise in the environment; those that are unpredictable at first, but can be learnt, and those that are actually random.

Future Work

In terms of robot control the development of both continuous empowerment (Jung et al., 2011) and fast continuous empowerment (Salge et al., 2012) was crucial to applying empowerment to real life systems in real time. The addition of state dependent noise now brings back the aspect of controllability to empowerment, and opens the possibilities for robotic control applications.

Imagine a robot that follows a human around in order to assist it. An empowerment map of the environment could provide the robot with an additional, supporting fitness function. Primarily, the robot would be interested in keeping its distance from the human. The reachability aspect of empow-

erment would keep it from getting stuck (actions would all lead to the same outcome), or ending up in a dead end. Controllability would keep the robot from getting too close to any human agent, considering they would be hard to predict. This could offer some additional incentives to the robot, once the primary objective is reached. So, instead of waiting, the robot could manoeuvre into a better position, where it could quickly get to a lot of other places, or where it would be less in danger of crossing paths with an unpredictable human agent.

Conclusion

We demonstrated how state dependent, coloured noise can be integrated into the fast quasi-linear Gaussian approximation of empowerment. This allows this faster empowerment approximation to regain the state-dependent noise-sensitivity of the original formalism. The extension allows us to demonstrate with our examples how empowerment is negatively influenced by interacting with a local source of noise. We see that a greedy empowerment-maximising agent tries to avoid such interaction; in our specific case the collision with another agent. The same principle would apply to other forms of noise, i.e. other aspects of the environment that either cannot be, or have not been properly modelled by the agent.

Acknowledgements

This research was supported by the European Commission as part of the CORBYS (Cognitive Control Framework for Robotic Systems) project under contract FP7 ICT-270219. The views expressed in this paper are those of the authors, and not necessarily those of the consortium.

References

- Ay, N., Bertschinger, N., Der, R., Güttler, F., and Olbrich, E. (2008). Predictive information and explorative behavior of autonomous robots. *The European Physical Journal B-Condensed Matter and Complex Systems*, 63(3):329–339.
- Bertschinger, N., Olbrich, E., Ay, N., and Jost, J. (2008). Autonomy: An information theoretic perspective. *Biosystems*, 91(2):331–345.
- Bialek, W., Nemenman, I., and Tishby, N. (2001). Predictability, complexity, and learning. *Neural Computation*, 13(11):2409–2463.
- Blahut, R. (1972). Computation of channel capacity and rate-distortion functions. *Information Theory, IEEE Transactions on*, 18(4):460–473.
- Capdepuy, P., Polani, D., and Nehaniv, C. (2007). Maximization of potential information flow as a universal utility for collective behaviour. In *Artificial Life, 2007. ALIFE'07. IEEE Symposium on*, pages 207–213. IEEE.
- Capdepuy, P., Polani, D., and Nehaniv, C. (2012). Perception-action loops of multiple agents: informational aspects and the impact of coordination. *Theory in Biosciences*, pages 1–11.
- Cover, T. M. and Thomas, J. A. (1991). *Elements of Information Theory*. Wiley-Interscience, 99th edition.
- Csikszentmihalyi, M. (2000). *Beyond boredom and anxiety*. Jossey-Bass.
- Der, R., Steinmetz, U., and Pasemann, F. (1999). *Homeokinesis: A new principle to back up evolution with learning*. Max-Planck-Inst. für Mathematik in den Naturwiss.
- Jung, T., Polani, D., and Stone, P. (2011). Empowerment for continuous agent environment systems. *Adaptive Behavior*, 19(1):16.
- Klyubin, A., Polani, D., and Nehaniv, C. (2005a). All else being equal be empowered. *Advances in Artificial Life*, pages 744–753.
- Klyubin, A., Polani, D., and Nehaniv, C. (2005b). Empowerment: A universal agent-centric measure of control. In *Evolutionary Computation, 2005. The 2005 IEEE Congress on*, volume 1, pages 128–135. IEEE.
- Klyubin, A., Polani, D., and Nehaniv, C. (2008). Keep your options open: an information-based driving principle for sensorimotor systems. *PLoS ONE*, 3(12):e4018.
- Lungarella, M., Pegors, T., Bulwinkle, D., and Sporns, O. (2005). Methods for quantifying the informational structure of sensory and motor data. *Neuroinformatics*, 3(3):243–262.
- Pfeifer, R., Bongard, J., and Grand, S. (2007). *How the body shapes the way we think: a new view of intelligence*. The MIT Press.
- Prokopenko, M., Gerasimov, V., and Tanev, I. (2006). Evolving spatiotemporal coordination in a modular robotic system. *From Animals to Animats 9*, pages 558–569.
- Rasmussen, C. and Williams, C. (2006). *Gaussian processes for machine learning*, volume 1. MIT press Cambridge, MA.
- Salge, C., Glackin, C., and Polani, D. (2012). Approximation of empowerment in the continuous domain. *Advances in Complex Systems*, 16(1 & 2):1250079.
- Schmidhuber, J. (1991). Curious model-building control systems. In *Neural Networks, 1991. 1991 IEEE International Joint Conference on*, pages 1458–1463 vol.2.
- Scott-Phillips, T. C., Dickens, T. E., and West, S. A. (2011). Evolutionary theory and the ultimate-proximate distinction in the human behavioral sciences. *Perspectives on Psychological Science*, 6(1):38–47.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell Sys. Tech. Journal*, 27:623–656.
- Sims, K. (1994). Evolving 3d morphology and behavior by competition. *Artificial life*, 1(4):353–372.
- Steels, L. (2004). The autotelic principle. *Embodied Artificial Intelligence*, pages 629–629.
- Telatar, E. (1999). Capacity of multi-antenna gaussian channels. *European transactions on telecommunications*, 10(6):585–595.
- Touchette, H. and Lloyd, S. (2000). Information-theoretic limits of control. *Phys. Rev. Lett.*, 84:1156–1159.
- von Uexküll, J. (1909). *Umwelt und Innenwelt der Tiere*. Springer.