INFORMATIONAL PRINCIPLES OF PERCEPTION-ACTION LOOPS AND COLLECTIVE BEHAVIOURS

by

PHILIPPE CAPDEPUY

A thesis submitted in partial fulfilment of the
requirements of the University of Hertfordshire
for the degree of

DOCTOR OF PHILOSOPHY

Adaptive Systems Research Group, School of Computer Sciences
Faculty of Engineering and Information Sciences
University of Hertfordshire

MARCH 2010

Abstract

Living beings, robotic and software artefacts can all be seen as agents acting and perceiving within an environment. When observed under that perspective, a new concept is accessible: information in the sense of Shannon. It has long been known that information and control are interrelated concepts. However it is only recently that this perspective has been better understood and used in order to study cognition.

In this thesis, we build upon such an information-theoretic perspective and add some biologically motivated assumptions. They introduce various constraints on the capture, the processing, or the storage of information by an agent. Using such constraints it is possible to understand some limits on the control abilities of agents, and to derive algorithms that optimize these abilities.

More specifically this thesis uses the recently introduced concept of *empowerment*, i.e. the ability to act upon the environment and perceive back the changes through the sensors. Maximizing this quantity leads to a wide range of cognitively interesting properties. This work studies some of these properties. One of them, the ability to capture information that is relevant for the perception-action loop of the agent, is deeply investigated and algorithms for exploiting this ability are presented.

The second part of the thesis deals with the use of the information-theoretic framework when multiple agents are interacting with each other. Empowerment maximization in this context leads to two phenomena: the generation of complex structures, and the emergence of synchronised and potentially cooperative interactions. In this thesis, the first phenomenon is empirically investigated through various spatial scenarios in order to understand the kind of structures that are generated and under which conditions they appear. Connections are made between the second phenomenon and the concept of the multiple-access channel. Using recent developments of this information-theoretic model, it is possible to precisely study the kind of interactions that can occur, and the situations that lead to synchronised or cooperative behaviour.

The general aim of this work is to give a comprehensive picture of the information-theoretic framework for studying the perception-action loop, bringing both single and multi-agents aspects together. The concepts presented in this thesis allows one to study some fundamental aspects of cognition, to engineer self-motivated robotic systems, or to drive self-organization in multi-agents systems.

Acknowledgements

I would like to thank my principal supervisor Dr. Daniel Polani and my secondary supervisor Prof. Chrystopher L. Nehaniv for their useful guidance and their support. All the exchanges we had during these years have strongly shaped the content of this thesis.

I am also grateful to my two examiners Dr. Hugo Touchette and Prof. Neil Davey for all the feedback they provided me with. If you find this thesis clear, you should also thank them, otherwise you can blame me.

Many thanks also go to all the staff of the university and especially to Lorraine Nicholls and her team for their kindness and their always efficient support.

Of course, going through this would not have been possible without the support of the melting pot of wonderful people I met during my time in UK: Arnaud, Moritz, Paul, Sven, Ben, Johannes, Nicolas, Antoine, Ester, Tom, Holly, Josh, Frank, Christoph, Sander, Marc just to name a few. Lots of thanks to them for making this experience a very enjoyable one.

I would also like to thank all my friends in France for their support, and especially my family without whom none of this would have been possible.

Finally, many thanks go to Cécile, for her love and patience during all these years.

PHILIPPE CAPDEPUY

Contents

Chapter 1

Introduction

During the course of evolution, living beings have reached staggering levels of complexity. Ranging from unicellular organisms to human societies through multicellular organizations. This complexity appears at two different but connected levels: the behavioural and the organizational level.

Behavioural complexity appears in the very rich and diverse strategies exhibited by living beings. These strategies are implemented using highly complex cognitive abilities such as learning, memorisation, anticipation and decision-making. Instead of thinking of cognition as being a specificity of higher animals, we will prefer here to consider it in its most general sense, attributing it to systems as simple as bacteria, although obviously some of the mechanisms covered by the general meaning of cognition appear only at higher levels of complexity. Indeed, as Maturana and Varela (Maturana, Varela and Beer 1980) have pointed out, it is quite possible that life and cognition are two faces of the same process.

Organizational complexity can be first seen in the intricate web of interconnected biochemical reactions that constitutes living beings such as bacteria. This complexity appears in the spatio-temporal arrangements of the different components that realise the global organizational pattern that we identify as a living entity. Evolution has also provoked the emergence of several levels of spatio-temporal organization involving more than one individual: proto-multicellularity like the aggregative behaviour of slime moulds, multicellularity during the Cambrian explosion, proto-social behaviour in ant colonies, social organizations of primate groups.

These two dimensions of complexity are interrelated, indeed cognitive abilities make possible the emergence of organization at the upper-level, and in return the collective level favours the emergence or improvement of specific cognitive abilities at the expense of some others.

The evolutionary framework allows us to explain a posteriori why we do observe some behaviours. The answer it brings is that such behaviours lead to a higher reproductive success for the individuals that exhibit them, therefore the less successful variants have been eliminated and the ones that are left are the ones we can observe. The main mechanisms underlying evolution are random variation, inheritance, and selection. These mechanisms are effective in explaining how behavioural and organizational complexity are generated over several generations. It also gives an explanation to why we observe the emergence of complex cognitive abilities such as learning that allows individuals to adapt during their lifetime, simply by observing that adaptivity is generally a successful trait for survival (and hence reproduction). However it does not account for the mechanisms by which such lifetime adaptations are driven. Indeed it only applies to populations of individuals and the time-scale is over several generations.

Adaptation during lifetime is therefore a quite mysterious mechanism from the perspective of evolution. In higher animals typical aspects of adaptation can be found in the numerous mechanisms associated with learning, e.g. reward-based conditioning and latent learning. From an engineering perspective we are also interested in such mechanisms. Indeed one wants to build robots that are able to adapt to their environment during their 'lifetime'. Another example is adaptation to perturbations of their embodiment, for example to the failure of some of their mechanical parts. Living beings have mastered this skill. An animal that loses a leg will adapt its movement strategy to this new constraint and will still be able to perform this task. Neuroimaging has shown that people with a disability such as blindness reallocate computational resources (i.e. parts of their brain) in order to compensate with other senses.

Evolution also has trouble dealing with organizations of individuals. Several theories are competing to explain how selfish behaviour (maximum reproductive success) can lead to seemingly altruistic behaviour, e.g. kin selection and group selection. Also, even if such theories can explain why some individuals will cooperate using an external viewpoint, it does not give any hint about what kind of mechanisms will implement this cooperation. Indeed the only interaction that the theory of evolution considers between individuals is the relative reproductive success. For instance, the presence of predators reduces the reproductive success of prey, and the presence of prey improves the reproductive success of predators. Such mechanisms can lead to complex population dynamics, as shown by Lotka-Volterra equations, but it is a narrow view of the rich range of interactions that individuals can experience. Understanding them requires an internal viewpoint that evolution does not provide.

Considering the two main shortcomings of evolutionary theory mentioned above, namely the lack of explanatory power over lifetime adaptation, and for the emergence of collective organization, one can wonder whether it is possible to find principles that would fill this gap. It is this question that this thesis tries to answer. Taking the perspective of individuals as entities that perceive and act over their environment, I study some fundamental aspects and limitations of the possible behaviours they can implement. The major dimension of this investigation is based on information in the sense of Shannon (1948). Removing the teleological aspects of behaviour implied by evolution, i.e. maximization of reproductive success, information is seen as a 'semantic-free' currency that any embodied agent (individual) has to deal with, whatever task it wants to accomplish.

Building on previous work from Klyubin (2007) and (Klyubin, Polani and Nehaniv 2007, Klyubin, Polani and Nehaniv 2008), and mainly the concept of empowerment, I describe the relationships between information theoretic quantities and the perception-action loop of embodied agents. Maximization of such quantities leads to important qualitative behaviour observed in living beings, such as knowledge acquisition and environment modification. Moreover, by looking at these quantities for several agents interacting in the same environment, I show that the emergence of collective organizations can be a consequence of these maximization principles, mainly due to the fact that information can exhibit synergistic phenomena, instantiating the famous saying that 'the whole is more than the sum of its parts'.

1.1 Contributions of the Thesis

The work presented in this thesis brings the following contributions by order of importance:

Extending the perception-action loop framework to multiple agents: A minimal model of the perception-action loop of two agents interacting in a common environment is presented. This model is then connected to network information theory and the formalism of the *multiple-access channel*, allowing us to use recently introduced algorithms that can compute the capacity of such channels (Watanabe and Kamoi 2002, Rezaeian and Grant 2004), and therefore study the empowerment of multiple agents in such situations. The use of minimal models is a first step that allows us to identify general properties of multiple-access channels that can be used to understand more complex situations.

Identifying drives towards organization and coordination: It is shown that empowerment maximization in multi-agent systems can lead to competitive or collaborative situations. Global organization emerges as the result of such behaviours in a spatial environment. The multiple-access channel model allows us to study the effect of coordination between agents. It is shown that such coordination can increase their empowerment for a class of channels. Empowerment maximization in various scenarios induces coordinated behaviour.

Identifying theoretical bounds on empowerment gain: An agent that has access to information about the state of the environment can increase its empowerment from two perspectives: (i) better distinguishing the effect of its actions, and (ii) picking actions according to the current state of the environment. These two aspects are distinguished, and some theoretical bounds on their respective empowerment gain are identified.

Introducing new algorithms for context extraction: Information about the state of the environment is only available to the agent through its sensorimotor history. Extracting this information has been done in (Klyubin 2007) using evolutionary algorithms. This technique is improved upon by using one of the theoretical bounds previously identified. Two iterative algorithms inspired from the information bottleneck principle (Tishby, Pereira and Bialek 1999, Slonim 2002) are proposed that efficiently perform the same task.

Extending the perception-action loop framework to include feedback: The causal Bayesian model of the perception-action loop and the associated quantities introduced in (Klyubin 2007) are modified in order to include feedback mechanisms using the formulation of *directed information* (Massey 1990) and *feedback capacity* (Yang, Kavcic and Tatikonda 2005, Tatikonda and Mitter 2009).

Introducing heuristics for on-board model-acquisition: An accurate probabilistic model of the perception-action loop is needed to make use of the information theoretic tools presented in this thesis. A first heuristic is presented that allows an agent to acquire such a model with a minimum number of samples and that adapts to changes in the environment or embodiment. A second set of heuristics is introduced that help identify causal relationships that may span over long time delays, reducing the computational cost by not having to process all the intermediate time-steps.

1.2 Outline of the Thesis

This thesis is structured as follows. Chapter 2 describes the background of this work and how this thesis is connected to other research. Chapter 3 presents the formalism behind this work, i.e. the formulation of the perception-action loop of embodied agents as a causal Bayesian network. The original formalism from (Klyubin 2007) is presented along with the various quantities that are investigated in this thesis. Recent advances in information theory concerning feedback channels are introduced and integrated into the model in order to give a consistent perspective of the quantities of interest.

In Chapter 4 the concept of *empowerment* is introduced and its maximization is described through two main mechanisms: modification of the environment, and context extraction. The first mechanism is illustrated through different examples in order to give an intuitive understanding of empowerment maximization to the reader. The context extraction mechanism is studied in greater detail by looking at some of its theoretical limits. This allows the derivation of efficient algorithms for empowerment-optimal acquisition of context. Heuristics are then introduced that allow an agent to efficiently capture the relevant statistics of the perception-action loop in changing environments and when causal relationships span over several time-steps.

The next two chapters present the multiple-agent perspective of the perception-action loop. In Chapter 5, we use a simplified version of collective behaviour where many agents are interacting in an asynchronous way. This simplification allows to get rid of the interference phenomenon of agents performing actions at the same time. It is shown that empowerment maximization in such systems can create competitive or collaborative situation. We then focus on the spatial organization of multiple agents and we show through various scenarios that empowerment maximization at both the local or global level leads to complex organizational patterns. Chapter 6 studies the interference phenomenon using a simplified model of two interacting agents. This allows us to connect this phenomenon with the information-theoretic model of the multiple-access channel. Using this connection, we empirically study a range of such channels in order to identify the dynamic of the informational quantities of interest, and impact of coordinated behaviour on these quantities is studied. This allows us to identify situations in which empowerment maximization leads to spontaneous coordination between agents. Different scenarios are then studied using game-theoretic tools and evolutionary algorithms.

Chapter 7 sums up and integrates the different perspectives presented for single-agent and multiple-agents empowerment maximization, and provides various directions for future research.

Chapter 2

Background

2.1 Embodiment and Cognition

Over the last three decades, artificial intelligence has gone through an important paradigm shift. Before this change, intelligence and cognition were perceived in a human-centric way. Considering reason as the main mechanism behind intelligence, artificial models were originally based on logic and rule systems. This approach is highly successful in dealing with specific problems where the context can be made fully explicit. It led to the development of expert systems able to make deductions and decisions in their field of knowledge. Another success of this approach is artificial chess players.

However it was soon realised that these methods were not able to deal with problems that cannot be defined in an explicit and rational way. A typical example of such a problem arises in computer vision. Recognising an object such as a chair from a digital picture is indeed very difficult because there are multiple possible designs for it, and several parameters such as lighting and orientation can dramatically alter its appearance. Researchers then realised that such problems are actually commonplace, and that all living beings are able to deal with them in an efficient way, even for relatively simple organisms. This led to the understanding that reason and logic are only a very small part of what intelligence and cognition encompass.

Hence the focus switched from human-centred intelligence to the kind of abilities found in the animal kingdom and other apparently simpler species. Understanding how living beings make sense of their environment requires to take into account what they can do with their environment. This is one of the early insights brought by (Uexküll 1934). For instance, recognising a chair for a human, whatever the design of this chair may be, requires to identify an object on which we can sit and lean. This idea was revived later in (Gibson 1979) with the concept of *affordances*. These describe all the actions that can be performed on an object, in relation to a particular agent.

From these two perspectives, researchers realised that a key aspects of intelligence is the *embodiment*, i.e. the sensor and motor capacities of an agent in a given environment. Indeed all living beings are embodied, their spectrum of potential actions is limited by their own physical properties along with how these physical properties can interact with their environment. This perspective led to an important paradigm shift in artificial intelligence, which has been called embodied and situated cognition (Varela, Thompson and Rosch 1993, Almeida e Costa and Rocha 2005). In this new paradigm, cognition is not based on abstract knowledge and concepts but on situated knowledge that is inherently related to the embodiment and the actual environment that the agent is operating in. In the recent years, the fields of artificial intelligence and artificial life have been strongly influenced by this paradigm. This line of thought is best illustrated by (Brooks 1990) where many seemingly 'life-like' behaviours are exhibited by simple robots which use almost direct mappings from sensors to actuators without any 'symbolic' computation.

2.2 The Perception-Action Loop

In the embodied cognition paradigm a key concept is the perception-action loop of an agent. It is defined as the interplay between actions that the agent performs on the environment and the resulting perceptions it gets through its sensors which can potentially be fed back into the decision process in order to decide on the next action. This process constantly occurring over time can be seen as a loop between perception and action which is mediated by the environment for the external part of the loop, and by the controller of the agent for the internal part.

This perspective is strongly reminiscent of control theory. Indeed the embodied agent acts as a controller that performs actions to put its environment in a specific state, and that may use feedback from its sensors to improve its control. The concept of homeostasis (Cannon 1939), which has been identified as a crucial aspect of the activity of living beings, is indeed a control task where the goal is to keep the essential variables into a defined range. Homeostasis is the ability of a system to actively control the value of some internal variables (e.g. level of sugar in blood, temperature, etc...) despite the environmental perturbations that the system encounters.

The concept of homeostasis has been studied in greater detail by (Ashby 1956) and led to the law of requisite variety. This law states that 'only variety can destroy variety'[1]. This can be interpreted in the following way: if a control mechanism tries to achieve stability, then the number of states of the control mechanism has to be greater than or equal to the number of states of the system that it controls. This has strong connections with the information-theoretic perspective of control, and indeed it has been extended later by (Touchette and Lloyd 2000, Touchette and Lloyd 2004). In their work, information bounds on control are identified for both the open-loop and closed-loop situations

2.3 Information Theory and Embodied Cognition

Information theory (Shannon 1958) was initially developed as a theory of communication with practical applications in telecommunication systems. It is a mathematically well founded theory that quantifies statistical relationships between probabilistically related events such as the input and output of a communication channel.

The potential use of information theory in the context of embodied cognition had been pointed by Gibson (1979). However, he was quite pessimistic about the effectiveness of such an approach. His main argument against it is that the environment and the agent are not 'speaking' to each other, in the sense that the environment is not sending semantically loaded messages to the agent, i.e. it is not trying to communicate with it.

It is only later, using the formalism of causal Bayesian networks, that the information-theoretic study of embodied agents was made possible. It then became clear that, even though the agent and the environment cannot really be seen as 'sender' and 'receiver', communication channels and flows of information between the agent and the environment could be identified and quantitatively treated. Also, even if one agrees with Gibson's perspective, it is still possible to consider the agent as 'speaking' with itself over time, while the environment is just the medium of this communication. Such a perspective was developed by Klyubin (2007) and is central to this thesis. It will be described in more

[1] This is not always true. In some cases, for example dissipative systems, variety can spontaneously be reduced. A more precise formulation of this law is 'the larger the variety of actions available to a control system, the larger the variety of perturbations it is able to compensate' (Heylighen 1992).

details in Section 2.6.

Even though, until Klyubin's work, the perception-action loop of embodied agents was not treated in an integrated way using information theory, some parts of it were. Indeed many researchers studied perception and the processing of sensors as informational mechanisms. Some of the first insights were made by (Attneave 1954, Barlow 1959). They hypothesised that sensory information represents huge volumes but contains a lot of redundancy. The goal of information-processing systems would then be to reduce this redundancy, according to an *economy* principle.

Other researchers also realised that the statistical regularities of sensor information were captured and internalised into the brain (Shepard 1984). These regularities are hypothesised to be used for both compression of information in sensory pathways and recoding of this information for suitable processing. Furthermore, the existence of *information bottlenecks* (Atick 1992) in the sensory pathways implies that information must be compressed. For instance, Atick (1992) states that the perceptual capacity of the visual pathway in humans is around $40 - 50$ bits/s, whereas photo-receptors of the eye collect around 5×10^6 bits/s. This implies a dramatic reduction of the collected information in this bottleneck. Information-theoretic methods were developed later in order to characterise such bottlenecks (Tishby et al. 1999, Slonim 2002) and find optimal information-preserving compression.

The redundancy reduction hypothesis was later revised. Because highly compressed representations are unlikely to be suitable for neuronal processing, it was suggested that redundancy could be used by the brain in order to increase its robustness to noise and to improve its anticipation abilities (Barlow 2001). His work also suggests that the brain 're-encodes' sensory signals in ways that makes their exploitation easier, for example by factorizing them into independent signals.

Information-theoretic approaches have also been used in the context of artificial neural networks. One of the pioneering work has been that of (Linsker 1988). He suggested that information transmission between successive layers in a neural network should maximize the amount of preserved information. This principle has been called *infomax*. It has been used successfully in generating artificial neural networks that are organized similarly to the early stages of visual perception in living beings. However, his results are based on various assumptions about the structure of the receptive field and the neural network.

The understanding of neuronal computation from the perspective of information theory is also a very active field of research. It has been found for instance that information maximization principles are equivalent in some cases to well-known learning rules such as synaptic time-dependent plasticity (Chechik 2003).

Using information theory to study embodied agents was also suggested by Nehaniv (1999). This perspective was picked up later on in (Olsson, Nehaniv and Polani 2005, Olsson 2006) and (Lungarella and Sporns 2006). By measuring various information quantities generated by agents engaging into sensorimotor interactions with their environment, it was shown that structured interaction resulted in the emergence of information flows between different parts of the system. These flows are a result of both the embodiment and the behaviour of the robot, but they can also be shaped through learning. Later on, maximization of such measures was used in order to directly generate sensorimotor coordination (Sporns and Lungarella 2006).

2.4 Intrinsic Motivation

Ethological and psychological studies have shown that not all actions performed by living beings are externally motivated, i.e. motivated by a potential reward or an external drive. Some behaviours are performed 'for their own sake', i.e. as a result of internal drives not directly related to any external reward. This has led to the formulation of several theories and mechanisms to account for this phenomenon. These have been labelled differently although they all convey the same general idea, i.e. living beings behave according to the satisfaction of internal drives. Examples of these are intrinsic motivation (Deci and Ryan 1985), adaptive curiosity (Schmidhuber 2006, Oudeyer and Kaplan 2004), the autotelic principle (Steels 2004). In this work the term *intrinsic motivation* will be used to refer to these mechanisms in a general sense.

From an evolutionary perspective, being intrinsically motivated can be a source of increased fitness, and therefore the associated mechanisms have a good chance of being selected. Intrinsically motivated agents are able to explore and learn by themselves some aspects of their environment which may prove useful in terms of survival. However evolution does not explain what kind of mechanisms give rise to this property. Homeostasis (Cannon 1939) has been identified as one such mechanism. To the roboticist trying to engineer autonomous robots, intrinsic motivation is also a very interesting property. It allows the robot to explore and learn by itself instead of having to be explicitly taught all the things he might need to know.

One mechanism that has been proposed is the maximization of learning progress (Kaplan and Oudeyer 2004). The idea is that an agent will favour the execution of actions that will maximize the efficiency of a predictor. The predictor learns from previous samples of action-perception couples and its derivative over time is used as the measure of learning progress. Another mechanism called homeokinesis was introduced by (Der, Steinmetz and Pasemann 1999) as a dynamic version of homeostasis. Its main principle is that the agent tries to generate dynamic behaviour which can be predicted by an adaptive model of the agent's interactions with the environment. An information-theoretic version of this principle was later developed by (Ay, Bertschinger, Der, Güttler and Olbrich 2008) and is referred to as *predictive information*. It is defined as the mutual information between sensor states in the past and sensor states in the future. Intrinsically motivated agents based on predictive information try to maximize this quantity. The outcome is two-fold, on the first hand the agent has to engage in exploratory behaviour in order to generate diversity in its sensor states; on the other hand it has to behave in a regular and structured way in order to preserve correlations between past and future.

Research in intrinsically motivated agents is a very active field. All the candidate functions for driving it are quite similar in spirit, but differ strongly in their details. Also it is very difficult to objectively compare the behaviours generated by these different principles. One of the difficulties is that each of them has been implemented on specific robots or virtual agents. A comparative study using similar embodiments is still lacking at this point.

2.5 Self-Organization in Collective Systems

It has been known for a long time that multi-agent systems, or collective systems, have the ability to self-organize in complex organizational patterns. A striking example in nature is the behaviour of ant colonies. From very simple individuals that can only perceive a small portion of the global environment one can obtain, using only basic behaviour rules, quite complex global behaviour which shows ability to solve difficult problems. For example ant colonies as a whole are able to explore the environment in order to find food sources and identify the shortest routes to exploit these sources (see (Bonabeau, Dorigo and Theraulaz 1999)). One of the mechanisms they rely upon is *stigmergy*, i.e. the ability to indirectly influence and coordinate behaviours by leaving traces of actions in the environment, using pheromones for example. Similar principles have been adapted to develop network routing algorithms, or more generally as abstract search principles which are referred to as *Ant Colony Optimization*.

Even though some specific tasks have been well understood from a multi-agent perspective, it is not yet clear if there are more general principles that could help us understand the self-organization of collective systems. Following a line similar to intrinsic motivation, some researchers are looking for generic candidate principles that could drive the local behaviour of individuals in order to generate some global organization. The main interest in such a system is that if one is trying to engineer a local behaviour that is able to globally solve a specific task, then such principles could be used in order to smooth the fitness landscape of the search space. Put another way one could use such a function as a heuristic to search the space of possible behaviours. This avenue is investigated in (Prokopenko, Gerasimov and Tanev 2006). Their goal was to obtain a robotic snake made up of several connected modules (basically segments of the snake) that is able to move at a satisfying speed on a wide range of possible terrains. They show that instead of using direct velocity-based fitness, the search could be made more efficient by taking into account information-theoretic measures expressing coordination between the different modules.

A similar approach has been pursued by Sperati, Trianni and Nolfi (2008). In their work the behaviour of individual robots is evolved according to a fitness function that incorporates information-theoretic measures of the correlation between the robots' actions. This principle allowed them to efficiently search the space of possible behaviours to achieve globally structured behaviour such as moving forward as a group. In some cases the robots evolved to use explicit communication in order to achieve the global coordination.

Simple scenarios involving stigmergy have also been studied from an information-theoretic perspective in (Klyubin, Polani and Nehaniv 2004). In their work, stigmergy is understood as the offloading of information in the environment and later acquisition of this information. They showed that agents could use the environment either as an external memory for a single agent or as a medium for communication with another agent, and that these can be quantitatively measured using Shannon information.

2.6 Prior Work of Klyubin

The research that led to this thesis is heavily based on the work of Klyubin (2007). Even though most of the aspects that are relevant to this thesis will be described in the next chapters, a quick overview of his work is presented here.

Based on the causal Bayesian graph formalism, Klyubin introduced a model of the perception-action loop of embodied agents unrolled over time. This model consists of various random variables which represent the sensors, the actuators, the state of the environment and the state of the agent's memory at different time-steps. The causal Bayesian graph describes the causal probabilistic relationships between these variables. On top of this formalism, and using Pearl's concept of *causal effect* (Pearl 2000), he introduced a notion of information flow that quantifies the amount of information that is causally transmitted from a set of nodes to another set of nodes in the graph (see Chapter 3 for the formalism). The concept of information flow is studied in great depth in (Ay et al. 2008).

In the first part of his thesis, he studies the maximization of information flow from the starting position of an agent moving into a grid to its memory. By doing so what is obtained is a mechanism that captures information about the starting position from sensorimotor experience. When this information is mapped to the spatial position of the agent, one can see that a representation of space emerges, distinguishing between different parts of the grid. A similar approach is used to capture information about time in the experiment (which goes on for 15 time-steps). Again the maximization principle allows the emergence of a representation of time. Later on he uses informational principles to factorise the obtained representations, i.e. to split the memory variable into different variables that are as independent from each other as possible. The resulting mappings show that the representation of space can be factorised into different coordinate systems, referring for instance to the upper-left versus lower-right and upper-right versus lower-left parts of the space. Similarly, the factorisation of time leads to two variables that specify whether the experiment is at an odd or even time-step, and how close is the experiment to its beginning or its end. More details about these results can be found in (Klyubin et al. 2007).

The second part of his thesis introduces the concept of *empowerment.* It will be described in great details in the next chapter, but a simple definition is that it is a quantitative measure of the ability of an agent to impact the environment with its actions and perceive the resulting changes through its sensors. Several properties of empowerment are introduced. The first one is its use as a state-space utility. By studying different scenarios, it is shown that empowerment can associate a value with different states of the environment. In one such scenario, an agent moving in a maze, empowerment is able to identify situations where the agent can potentially reach a maximum number of different states in a minimum number of steps. In a pole-balancing scenario, maximum empowerment identifies 'homeostatic-like' states, basically having the pole in the upright position, because this situation allows the agent to easily reach any other position of the pole. Detailed descriptions of these experiments and their results are presented in (Klyubin et al. 2008)

A second use of empowerment and its maximization is the evolution of sensors and actuators. If one assumes that an agent has constraints on the amount of information its sensors can capture or its actuators can send, then maximization of empowerment under such constraints allows the agent to evolve sensors and actuators that maximize the capture of information that is relevant to the use of its actuators. A simple way to understand this is the example of the cave-fish (Jeffery 2009). In an environment which is mostly dark, having eyes does not help the fish because there is no information to extract from them. Because of the metabolic cost that eyes imply, evolutionary pressure led to a fish that has no eyes but uses other kind of sensors.

The third use of empowerment presented in Klyubin's thesis is the extraction of relevant contexts. Basically the idea is to use the past sensorimotor experience in order to identify states of the environment that increase the control the agent has. This idea is illustrated by an experiment with an AIBO robot. The robot, sitting in front on an empty space, is able to perceive the distance of nearby objects in front of him. The only actions it can use are moving its head up or down. Two different conditions are experimented, either the robot has nothing in front of him or it has a book. An automaton that processes the sensorimotor data is evolved in order to maximize empowerment. The state of the automaton is then used as a context for the robot, i.e. a state value that it is aware of. It is shown that the evolved automaton is able to capture the presence or absence of the book (roughly 1 bit of information) because this information is relevant to its control abilities.

2.6.1 Relation of the Thesis with Prior Work

This thesis extends the work previously described in the following directions:

- The perception-action loop formalism is updated in order to incorporate feedback mechanisms using the concept of *directed information* (e.g. (Massey 1990, Yang et al. 2005, Tatikonda and Mitter 2009)), allowing the formulation of empowerment using channel capacities with or without feedback.
- A theoretical treatment of the impact of contexts on empowerment is presented, allowing the identification of various bounds on empowerment gain through the use of contexts.
- An improved version of the evolution of context-automaton is proposed using the newly identified bounds.
- Iterative algorithms for context-extraction are presented, that use a bottleneck-like approach. Simulations using these algorithms provide further indications on the limits of empowerment gain with context.
- Heuristics are proposed that allow an agent to efficiently acquire a model of its perception-action loop in changing environments and when time-extended causal relationships are involved.

Moreover, by applying the perception-action loop formalism to multiple agents, this thesis opens avenues that were unexplored in the previous work. Namely:

- The use of empowerment maximization for multiple agents as a mechanism for inducing competitive or collaborative situations and for generating collective organizations.
- The connection with the multiple-access channel model, revealing a new phenomenon of interferences between the agents actuation channels.
- The impact of coordination between agents on these interferences, and its impact on empowerment.
- The ability of empowerment maximization to generate coordinated behaviour between interacting agents.

Chapter 3

Information in the Perception-Action Loop

This chapter presents the core of the formalism that is used throughout the thesis. Some parts of this material has been previously described in the work of Klyubin (Klyubin 2007) but some aspects will differ. A notational glossary can be found in appendix A. Also, because information-theory is the foundation of this formalism, a quick introduction can be found in appendix B. For a complete treatment of information theory the reader is referred to (Cover and Thomas 2006).

The content of this chapter is divided into three parts. The first section presents the global idea behind the perception-action loop of embodied agents and how it can be formalised using causal Bayesian networks.

The second section introduces the concept of *information flow* (Ay and Polani 2008) which is at the core of the quantities used in this thesis. The different communication channels that constitute the perception-action loop of an embodied agent are described and the notion of context is introduced. The quantities of interest are illustrated using a simple example.

The last section deals with perception-action loops spanning over multiple timesteps. The notions of temporal horizons and feedback are introduced. The concept of directed information is presented and connected to the existing perception-action loop framework, allowing us to formulate feedback capacities for time-extended perception action loops.

3.1 Global Picture

In order to model the perception-action loop of an agent, we use the framework of *causal Bayesian networks* along with concepts of causality and intervention defined by (Pearl 2000). In this framework, all the relevant variables are modelled as random variables, and the causality constraints are defined by conditional probability distributions. In the case of the perception-action loop we define the following variables:

- the sensor of the agent S which takes values $s \in \mathcal{S}$,
- the actuator of the agent A which takes values $a \in \mathcal{A}$,
- the rest of the environment R which takes values $r \in \mathcal{R}$,
- and in some cases the memory of the agent M which takes values $m \in \mathcal{M}$.

In order to obtain a causal graph, the perception-action loop is unrolled over time, meaning that all these variables have an associated time-step t. The causality dependences can be described in the following way: the environment is in state R_t, this causes the sensor of the agent to have value S_t. The actuator value A_t is a consequence of the sensor reading. The state of the environment R_t combined with the action performed by the agent A_t leads to a new state of the environment R_{t+1}. This new state is then read by the sensor, leading to S_{t+1}, and so on. The resulting causal Bayesian graph is depicted on Fig. 3.1 and the corresponding joint probability distribution is defined as:

$$p(r_{t+1}, a_t, s_t, r_t) = p(r_{t+1}|a_t, r_t)p(a_t|s_t)p(s_t|r_t)p(r_t). \tag{3.1}$$

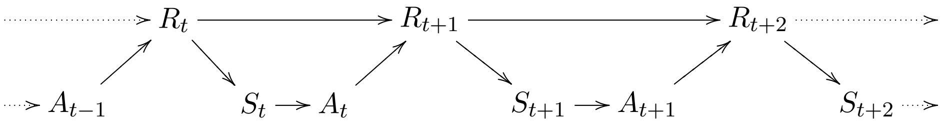

Figure 3.1: Representation of the perception-action loop as a causal Bayesian network unrolled in time. R_t stands for the state of the environment, S_t is the sensor of the agent and A_t its actuator. Empowerment measures the capacity of the actuation channel toward future perceptions. This channel depends on both the sensorimotor apparatus of the agent and on the environment through which the information flows.

It is clear in this model that the environment and the agent's embodiment are fully described by the actuation channel $p(r_{t+1}|a_t, r_t)$ and the sensor channel $p(s_t|r_t)$. On the other hand the agent's controller is described by $p(a_t|s_t)$.

It has to be understood that the actual implementation of the controller is fully abstracted in this model. It could be anything, from a neural network to a lookup table, or any other controller system that one could think of, as long as it can be described in a probabilistic fashion. This abstraction allows to study absolute limits of agents that fit into this model. One could compare it with the abstract Carnot cycle, that does not consider the actual implementation of the mechanism that the machine performs, but considers only the external constraints that apply to it.

It is also possible to add memory to the agent. There are different ways of doing it, one of the most complete is to add a variable M_t on which the action A_t can depend, and the next memory state M_{t+1} being conditioned on the previous one M_t, the last action A_t and the last perception S_t. In that case, the resulting causal Bayesian graph is depicted on Fig. 3.2.

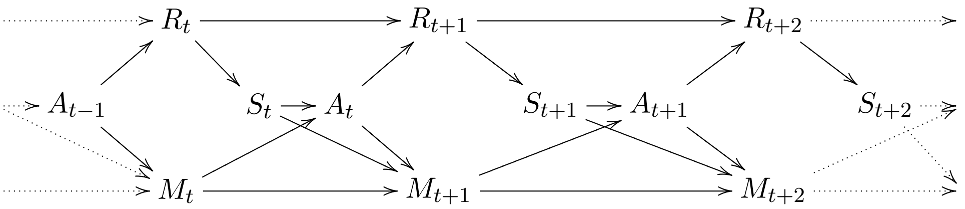

Figure 3.2: Perception-action loop of an agent that has a memory accounted for by M_t. It is a function of its previous state M_{t-1}, the last action A_{t-1} and sensor state S_{t-1}.

Again, the environment and agent's embodiment are fully described by the same conditional distributions of the actuation channel $p(r_{t+1}|a_t, r_t)$ and the sensor channel $p(s_t|r_t)$. However, the aspects of the model which are internal to the agent are described by the memory-dependent action policy $p(a_t|m_t, s_t)$ and the memory mechanism $p(m_{t+1}|a_t, s_t, m_t)$.

3.2 Information Flows

This section describes the concept of information flow (Ay and Polani 2008). It has to be mentioned that this is a rather technical section that is not needed to understand the remaining part of the thesis. Its main goal is to understand the reasoning behind the approach taken in (Klyubin 2007) on which this thesis is based. Indeed, the concept of empowerment has been introduced on the basis of information flows. However, one of the contributions of this thesis is to redefine the concept of empowerment using only channel capacities, and especially the recently introduced feedback channel capacity (Tatikonda and Mitter 2009).

When one wants to uncover the underlying causal links of a system, there are in general two approaches. The first one is to observe the system in the course of its action. By doing so one can identify correlations between different parts of the system. In the context of Bayesian graphs, this can be measured using the mutual information between different random variables of the graph.

However, simply observing the system is generally not enough to recover the actual causal structure. Typically, one is not able to distinguish between $X \to Y$ and $Y \to X$ by observation alone, because mutual information is a symmetric quantity. Note that, in some situations, complexity considerations can also be used to infer causality without intervening on the system (Sun, Janzing and Schölkopf 2008).

In order to disambiguate between different causal structures one has to intervene on the system. This can be thought of as an experimenter setting some parameters of the system to specific values and observing how the system responds to these values. The experimenter is generally considered as a 'free' source of information that is able to 'disconnect' some parameters of the system from their normal causes (or alter the causal links) and inject information into them.

A useful analogy for information flows presented in (Klyubin 2007) is the radioisotope tracer used for medical imaging. Doctors locally intervene on the circulatory system of the patient by injecting information (the presence of the tracer at a specific place) and observing how far this information flows, revealing the overall structure of the circulatory system.

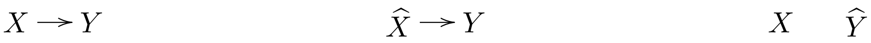

Figure 3.3: Simple causal Bayesian network (left). Intervention on node X (middle). Intervention on node Y (right).

In a Bayesian graph, information flows can be measured using the notion of *causal effect* and *intervention* defined by (Pearl 2000). An in-depth treatment of information flows can be found in (Ay and Polani 2008) but the general idea can be expressed quite simply. The Bayesian graph under study is modified through intervention. Intervening on the graphs implies removing some existing causal links and replacing the impacted variables with some other distributions.

In the case of a causal Bayesian graph $X \to Y$ (see Fig. 3.3), where one wants to measure the information flow from X to Y, the intervened node is denoted by $\widehat{X}$, and the conditional probability $p(y|\widehat{x})$ is then considered as the causal channel of interest. The *information flow* from X to Y is defined as:

$$\Phi(X \to Y) = I(\widehat{X}; Y). \tag{3.2}$$

In this simple example, one can easily see that if $\widehat{X}$ has the same marginal as X then information flow and mutual information are equivalent. However, intervening in a similar way on node Y reveals that no information can flow from Y to X (causal dependencies have disappeared on Fig. 3.3).

There are several possibilities for choosing the new distribution $p(\widehat{x})$. One can consider for example using an equidistribution. This is used in (Tononi and Sporns 2003) under the name of *effective information*, however this choice is rather arbitrary.

A more natural choice, which is used in (Ay and Polani 2008), is to define $\widehat{X}$ as having the same distribution as the marginal of the original X, but without the causal dependency. This situation is used in the example described below.

A third option is to consider the *maximum* amount of information that can be transmitted, i.e. the *maximum potential information flow*. This is the approach used throughout this thesis. One of its advantages is that it can be directly expressed as the capacity of the causal channel of interest. The distribution and capacity can easily be computed using the standard Blahut-Arimoto algorithm (Blahut 1972, Arimoto 1972).

It has to be made clear that information flow and mutual information measures, although related, are different from each other. Mutual information is a symmetric measure that quantifies the amount of correlation between two variables. On the other hand, information flow is an asymmetric measure that quantifies the amount of information that is causally transmitted from one variable (or set of variables) to another. In the context of a causal Bayesian graph which incorporates time (e.g. the perception-action loop), it is clear that, although there can be positive mutual information between two time separated variables such as R_t and R_{t+1}, the information flow can only be positive when measured along the causality path, i.e. $\Phi(R_t \to R_{t+1}) \geq 0$ but $\Phi(R_{t+1} \to R_t) = 0$.

For the sake of clarity it is useful here to consider one of the examples provided in (Ay and Polani 2008) for which mutual information and information flow are compared. This example presents the diamond structure implying four binary variables W, X, Y, and Z. The corresponding causal Bayesian graph and its intervened versions are described on Fig. 3.4.

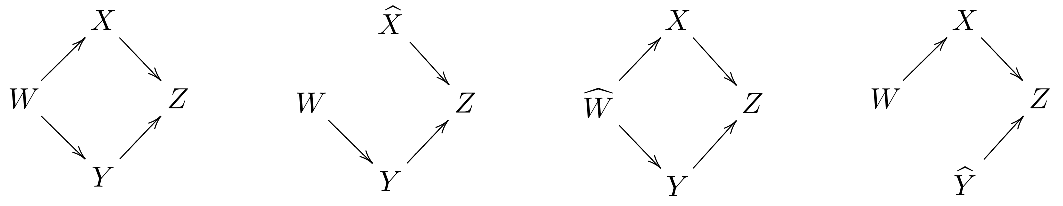

Figure 3.4: Causal Bayesian graphs representing the diamond structure and interventions on X, W and Y. Intervened nodes are replaced by random variables that preserve the marginal distribution of the original node.

The variable W is uniformly distributed over 0 and 1. The content of this variable is then directly copied to both X and Y. The last variable Z is the result of a XOR operation on X and Y. The result is that Z always contains 0 because the two inputs of the XOR are always the same.

The authors intervene on various nodes by removing their causal dependencies while preserving their marginal distribution (see Fig. 3.4). Causation measured using the information flow on intervened Bayesian graphs can then be compared to the correlation measured on the original graph.

- $I(X;Y) = 1$ and $\Phi(X \rightarrow Y) = I(\widehat{X};Y) = 0$: even though X and Y are actually correlated, because of their common source, no information actually flows from X to Y, which is obvious when the structure of the graph is considered.
- $I(X;Y|W) = 0$ and $\Phi(X \rightarrow Y|\widehat{W}) = I(X;Y|\widehat{W}) = 0$: in both cases conditioning on W 'explains away' the correlations between X and Y. Also because W has no causal dependencies (and is replaced by its marginal), the intervention does not change the Bayesian network.
- $I(W;Z|Y) = 0$ and $\Phi(W \rightarrow Z|\widehat{Y}) = I(W;Z|\widehat{Y}) = 1$: because in the original graph Z does not contain information (it is always 0), correlation is necessarily 0. However by intervening on Y and decoupling it from its source Z, the correlation between X and Y is removed and the output of the XOR gate becomes a uniform distribution. In this situation, the amount of information that flows from W to X then Z when the variable Y is known becomes 1 bit. This is invisible when using observational data from the network, but it becomes visible when one intervenes on it.

Even though the original treatment of the perception-action loop in (Klyubin 2007) relies mostly on the information flow formalism which provides strong generality, the specific information flows that this thesis deals with can be expressed as channels and their associated capacities. For example, consider the sensorimotor channel that goes from actions A_t to the next perceptive state S_{t+1} through the environment. Using the information flow formalism this channel is fully described by the conditional probability distribution $p(s_{t+1}|\widehat{a}_t)$, meaning that we are interested in all the possible outcomes of S_{t+1} for each possible value $\widehat{a}_t$ when this value is injected independently from any other variables (such as previous sensor states).

As mentioned before, the main quantity that we are interested into is the maximum *potential* information flow, i.e. the maximum amount of information that can flow from $\widehat{A}_t$ to S_{t+1}. This quantity is potential, because the agent might not actually inject this amount of information when behaving according to its controller. However, if its actions were decoupled from the past using some kind of 'free will', then it would be able to send a different amount of information. If one now considers the interventional channel $p(s_{t+1}|\widehat{a}_t)$, this maximum information flow can be computed by finding the probability distribution $p(\widehat{a}_t)$ that maximizes this quantity. This is actually equivalent to finding the capacity of the aforementioned channel. Therefore, in the rest of this thesis, unless explicitly stated, the 'hat' notation will be dropped and the channel considered will be simply denoted as $p(s_{t+1}|a_t)$. Nevertheless, one has to understand that information flow principles are always implied behind this notation.

3.3 Channels and Capacities in the Perception-Action Loop

As has been explained in the previous section, one of our key measure is the maximum potential information flow, which can be expressed as the capacity of a channel.

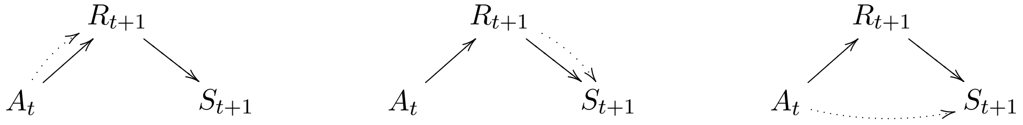

Figure 3.5: Channels of interest in the perception-action loop. Solid arrows: causal links. Dashed arrows: channels. From left to right: motor channel, sensor channel, sensorimotor channel.

If we look at a simple section of the perception-action loop, one can identify various channels of interest (see Fig. 3.5):

- The *motor channel* $A_t \to R_{t+1}$ (or actuation channel): this channel describes the ability of the agent to modify the state of the environment by performing specific actions. In terms of information flow we are interested into how much information can be injected into the environment by the actions. A crucial property of this channel is that its output depends on the current state of the channel, i.e. the knowledge of R_t effects the properties of this channel.

- The *sensor channel* $R_{t+1} \to S_{t+1}$: this channel defines how much information the agent can perceive about the state of the environment. Because it has no side-information, it is less complex that the motor channel, however its properties impose important constraints on what is achievable by the agent.

- The *sensorimotor channel* $A_t \to S_{t+1}$: it is the combination of the two previous channels. By essence it is the only one which is directly 'visible' by the agent. Indeed the agent never has direct access to the state of the environment, the only aspects it has access to are the actions it performs and the sensor readings. This channel is the base of the quantities used in this thesis. Its capacity measures how much information an agent can inject by its actions and later reacquire through its sensors. Because it contains the motor channel, it is also dependent on the state of the environment.

It has to be noted that similar channels, and especially the motor channel, have been studied in a control-theoretic perspective by (Touchette and Lloyd 2004, Touchette and Lloyd 2000). In this thesis, the term motor channel is preferred to the standard control-theoretic term 'actuation channel' because it better fits the general semantic of embodied artificial agents.

The list of channels presented above is not exhaustive, one could for example look at the channel that goes from sensors to actuators (i.e. $S_t \to A_t$). This channel depends on the 'decision mechanism' that the agent employs. However, because this thesis only looks at aspects of the agent that are independent from its decision mechanism, this channel is not treated (one can look at the concept of relevant information (Polani, Nehaniv, Martinetz and Kim 2006) for a treatment related to this channel).

The focus of this work is on the sensorimotor channel and its capacity C_{SM}:

$$C_{SM} = C(A_t \to S_{t+1}) = \max_{p(a_t)} I(A_t; S_{t+1}). \tag{3.3}$$

This channel can be divided in two subchannels: the motor channel with capacity

$$C_M = C(A_t \to R_{t+1}) = \max_{p(a_t)} I(A_t; R_{t+1}), \tag{3.4}$$

and the sensor channel with capacity

$$C_S = C(R_{t+1} \to S_{t+1}) = \max_{p(r_{t+1})} I(R_{t+1}; S_{t+1}). \tag{3.5}$$

We now go on to show that both subchannels are bottlenecks for the sensorimotor channel (this is an addition to the work of (Klyubin 2007)):

Theorem 3.3.1. *The capacity of the sensorimotor channel is bounded from above by the capacity of its subcomponents:*

$$C_{SM} \leq \min(C_S, C_M). \tag{3.6}$$

Proof. Assume a capacity achieving distribution $p^*(a_t)$ for the sensorimotor channel. This channel can be represented by the Markov chain $A_t \to R_{t+1} \to S_{t+1}$. Applying the data processing inequality one has

$$I_{p^*(a_t)}(A_t; S_{t+1}) \leq I_{p^*(a_t)}(A_t; R_{t+1})$$

By definition of the channel capacity we have $C_{SM} = I_{p^*(a_t)}(A_t; S_{t+1})$ and $I_{p^*(a_t)}(A_t; R_{t+1}) \leq C_M$. Therefore we obtain:

$$C_{SM} \leq C_M.$$

We can proceed similarly for the sensor channel. Applying the data processing inequality we get:

$$I_{p^*(a_t)}(A_t; S_{t+1}) \leq I_{p^*(a_t)}(R_{t+1}; S_{t+1}).$$

Again, the left-hand term is equivalent to C_{SM}. The right-hand term has to be understood as measuring the mutual information in the sensor channel for an input distribution $q(r_{t+1}) = \sum_{a_t} p^*(a_t) p(r_{t+1}|a_t)$. Because $q(r_{t+1})$ is not necessarily a capacity-achieving distribution, we have $I_{p^*(a_t)}(R_{t+1}; S_{t+1}) \leq C_S$. Therefore we can write:

$$C_{SM} \leq C_S.$$

□

It has to be noted that the capacity-achieving distributions for the sensorimotor channel may significantly differ from those of the motor or sensor channels. For example one can think of a situation where the agent has a subset of actions that contribute mostly to the capacity of the motor channel but whose effects cannot be observed by the agent. Similarly, a symmetric situation would be that there is a set of states of the environment that contribute mostly to the capacity of the sensor channel, but that these states cannot be reached by the actions available to the agent.

Because the sensorimotor channel encompasses both the sensor and the motor channels, achieving capacity requires to find an overlapping region where actions have an effect on the environment that is visible through the sensors.

3.3.1 Motor Channels and Side-Information

As mentioned in the previous section, the state of the environment may have an effect on the output of the motor and sensorimotor channels. Therefore, using the channel with or without this information, referred to as *side-information*, results in different capacities.[1]

In order to understand the impact of side-information let us focus on a simplified section of the perception-action loop where only the motor channel is considered (see Fig. 3.6).

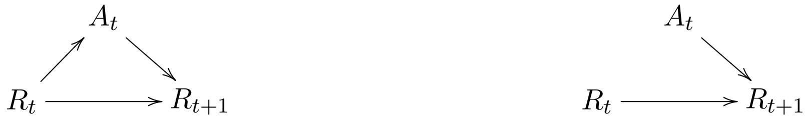

Figure 3.6: Simplified perception-action loop with only the motor channel spanning over one timestep. Left: channel with previous state of the environment accessible as side-information. Right: channel without side-information.

[1] It is important to note that in the perception-action loop framework we do not need to distinguish between having side information at the emitter, at the receiver, or both, as is usually done in the information theory literature. Indeed the agent acts as both a sender and a receiver, so either side information is inaccessible or it is accessible to both.

According to the model of the perception-action loop, the outcome of each action is defined by the action a_t and the original state r_t, i.e. the actuation channel $p(r_{t+1}|a_t, r_t)$. Depending on the distribution over the original states $p(r_t)$, capacity may differ. In fact, for each possible state r_t of the channel, the capacity of the motor channel when the state is fixed is:

$$C_M(r_t) = \max_{p(a_t|r_t)} I(A_t; R_{t+1}|r_t). \tag{3.7}$$

The average capacity is then defined as:

$$C_M(R_t) = \sum_{r_t} p(r_t) C_M(r_t). \tag{3.8}$$

Because the state of the channel is known and used in the decision-making process, this looks similar to performing closed-loop control. However, we do not make any assumption at this stage about whether the next steps will also have access to similar information. These aspects are studied with the concepts of horizons and feedback (see Sec. 3.4.1 and 3.4.2). Here we just assume that the state of the environment is known at time t.

In the case where the agent does not know the state of the channel, it becomes a mixed actuation channel. More precisely the motor channel from the controller perspective is then described by:

$$p(r_{t+1}|a_t) = \sum_{r_t} p(r_{t+1}|a_t, r_t) p(r_t). \tag{3.9}$$

As the channel capacity is a convex function, one can use Jensen's inequality to show that

$$C_M \leq C_M(R_t). \tag{3.10}$$

The relation between the different channels when side-information is accessible is similar to the one shown in theorem 3.3.1 (the following theorem is also an addition to the work of (Klyubin 2007)):

Theorem 3.3.2. *When the previous state of the environment R_t is known to the agent, the capacity of the sensorimotor channel is bounded from above by the capacity of its motor channel with access to R_t and the capacity of its sensor channel:*

$$C_{SM}(R_t) \leq \min\big(C_S, C_M(R_t)\big). \tag{3.11}$$

Proof. First, let us notice that because S_{t+1} depends only on R_{t+1}, the capacity of the sensor channel is not impacted by the knowledge of the previous state, i.e. $C_S(R_t) = C_S$.

Now, for each r_t, the channel can be considered as having no state information. In this case theorem 3.3.1 applies leading to

$$C_{SM}(r_t) \leq \min\big(C_S(r_t), C_M(r_t)\big).$$

Because the conditional capacity is the average over all states of the capacity of the channel in this given state, it follows directly that:

$$\sum_{r_t} p(r_t) C_{SM}(r_t) \leq \sum_{r_t} p(r_t) C_M(r_t)$$

and

$$\sum_{r_t} p(r_t) C_{SM}(r_t) \leq \sum_{r_t} p(r_t) C_S(r_t)$$

Therefore we have:

$$C_{SM}(R_t) \leq C_M(R_t)$$

and

$$C_{SM}(R_t) \leq C_S.$$

□

3.3.2 Contexts

The concept of side-information is presented in (Klyubin 2007) under the name of *context*. In the information theory literature, a channel with side-information is a channel whose outcome depends not only on its input but also on its current state. Some of this side-information may or may not be accessible to the sender or the receiver of the channel. Therefore, side-information refers to existing information that has an impact on the output of the channel. This information is necessarily contained in the state of the channel, but it can also be fully or partially replicated somewhere else.

One place where it can be replicated is the context of the agent. It is a set of variables that is accessible to the agent prior to performing its action and upon which it can make a decision. It can typically be the current sensor state S_t, the previous action A_{t-1}, a memory M_t constructed from past sensorimotor experience, or any combination of those (e.g. see Fig. 3.2). The idea is that this information, however it has been acquired, acts as an informational context for the controller. Ideally, the context is correlated with the current state of the environment and can then be used to increase the capacity of the channel. In this case the context can be said to contain side-information.

As a general rule of thumb, one can think of side-information as information about the state of the channel, whereas a context is a random variable accessible to the agent that contains some of this side-information.

The concepts of context and side-information are a crucial aspect of this thesis. The next chapter will present various ways for an agent to actually extract a useful context in order to increase its control on the environment.

Depending whether a context is available to the agent, (Klyubin 2007) distinguishes between the following capacities:

- *context-free capacity*: no side-information is available for the controller, e.g. $C(A_t \to S_{t+1})$.
- *contextual capacity*: a variable or a set of variables is accessible to the controller and may contain side-information correlated with the state of the channel, e.g. $C(A_t \to S_{t+1}|M_t)$.

Two different context classes are identified in (Klyubin 2007): *maximal* and *ideal* contexts. A maximal context M_t is such that there does not exist any context M'_t that would increase the capacity more that M_t does. An ideal context is a maximal context with the smallest possible state space.[2]

In the case of the perception-action loop presented on Fig. 3.1, the state of the environment R_t is a maximal context. However, depending on the environment, it may or may not be an ideal context.

Of course there is a continuum between context-free and contextual capacity with a maximal context. The amount of information about the state of the channel that the context captures is a crucial parameter. This aspect will be studied in greater detail in section 4.3, but a simple example can convey its importance.

Consider a XOR gate. The agent controls one of the inputs, sending $\{0, 1\}$ uniformly, and observes the output. The second input of the XOR gate is the state of the environment also uniformly distributed over $\{0, 1\}$. When no context is available to the agent (see Fig. 3.7 left), it is impossible for the agent to distinguish its own information from the output because of the noise introduced by the state of the environment. On the other

[2] This is quite similar to the concept of causal states in ϵ-machines (Shalizi and Crutchfield 2002), but in the perception-action loop actions are taken into account, and only the information which is relevant to their outcome is captured in the ideal context.

$$\begin{tikzcd} & A_t \arrow[dr] & \\ R_t \arrow[rr] & & R_{t+1} = A_t \oplus R_t \end{tikzcd} \qquad \begin{tikzcd} & A_t \arrow[dr] & \\ R_t \arrow[ur] \arrow[rr] & & R_{t+1} = A_t \oplus R_t \end{tikzcd}$$

Figure 3.7: Impact of context in the XOR channel. R_t and A_t are uniformly distributed over $\{0,1\}$. Left: context-free case, $I(A_t; R_{t+1}) = 0$. Right: contextual case with R_t as a context, $I(A_t; R_{t+1}|R_t) = 1$.

hand, if the agent uses the state of the environment as a context, it can perfectly distinguish between its own information and the information coming from the environment (see Fig. 3.7 right).

3.3.3 Example: the Line World

In order to illustrate how the previously described quantities behave, we introduce a simple world, i.e. an environment and an embodiment. This world can be seen as a bounded unidimensional discrete space (hence 'line world'). The embodiment is defined as follows:

- Actions are taken in the set {move left,move right, stay}.
- Sensor values are an exact copy of the state of the channel: $S_{t+1} = R_{t+1}$.

The outcome of actions is deterministic and the agent collides with the boundaries of the line world (see Fig. 3.8).

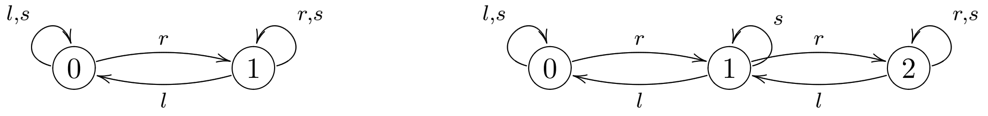

Figure 3.8: Automata of the line world of size 2 and 3. States of the automaton are the states of the environment (absolute position of the agent). Transitions are deterministic and labelled with the three possible actions: left, right, stay denoted as $\{l, r, s\}$. The agent sensor is the state of the channel.

The causal Bayesian graph of the perception-action loop is depicted on Fig. 3.9 along with the channel under consideration $A_t \to S_{t+1}$.

Let us first analyse the line world of size 2 (LW(2)) represented on Fig. 3.8 left. The following table shows the capacity of the sensorimotor channel for each possible state of the environment:

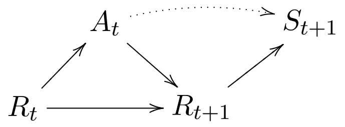

Figure 3.9: Section of the perception-action loop studied in the line world. Solid arrows: causal links. Dashed arrow: channel. The causal link between R_t and A_t represents the potential use of R_t as a context.

r_t	0	1
$C(A_t \to S_{t+1}\|r_t)$	1	1

Regardless of the distribution over the states $p(r_t)$ is, the average capacity of the motor channel is $C(A_t \to S_{t+1}|R_t) = 1$. If we now consider the capacity of the motor channel without knowing the state of the environment we also have $C(A_t \to S_{t+1}) = 1$. This is easy to see, simply because wherever the agent is, if it emits action 'left' it will end up in state 0, and respectively in state 1 if it emits action 'right'. Only the action 'stay' has an outcome which depends on the current state.

If we now look at LW(3), the capacities for each state of the environment are:

r_t	0	1	2
$C(A_t \to S_{t+1}\|r_t)$	1	1.58	1

One can see that the motor channel has a higher capacity when the agent is at the centre of the line. This is due to boundary effects. Indeed, when the agent is at a boundary, only two outcomes are possible: either it stays at the boundary, or it moves one tile away from it. On the other hand, when the agent is at the centre, three outcomes are possible: it can stay there, move toward the left boundary, or toward the right boundary.

Therefore, by behaving in this environment, the agent can change the capacity of its sensorimotor channel. Put another way, the distribution over the states of the environment has an impact on both the average motor capacity with and without knowledge of the state. For the former, the average capacity is simply the weighted average of the sub-channels' capacities, therefore $C(A_t \to S_{t+1}|R_t) \in [1; 1.58]$.

When the state of the environment is unknown, things get more complicated. The simplest cases are those for which the distribution over the states is concentrated on only one of them. In that case the average motor channel has exactly the same distribution as the sub-channel which constitutes the support of $p(r_t)$. Hence in the setup of LW(3), the maximum capacity 1.58 is reached when $p(R_t = 1) = 1$.

We now look at the following distribution: $p(R_t = 0) = \frac{1}{2}$ and $p(R_t = 2) = \frac{1}{2}$. When the state is unknown to the agent, the corresponding channel is an equiprobable mix of sub-channels 0 and 2, leading to the following conditional distribution $p(r_{t+1}|a_t) = \sum_{r_t} p(r_t)p(r_{t+1}|a_t, r_t)$:

r_{t+1} a_t	0	1	2
left	$\frac{1}{2}$	$\frac{1}{2}$	0
stay	$\frac{1}{2}$	0	$\frac{1}{2}$
right	0	$\frac{1}{2}$	$\frac{1}{2}$

The context-free capacity of this channel is then $C(A_t \to S_{t+1}) = 0.58$ bits, which is a significant reduction compared to the contextual capacity.

One has to understand that the distribution over the states of the environment is a direct consequence of the behaviour of the agent. If we consider the distribution of the previous example as a starting distribution $p(r_0)$, it is easy to imagine a policy which after a few time-steps would lead to a distribution $p(r_t)$ that gives a capacity greater than 0.58 bits. An example of such a policy is to move left all the time. In that case, for any $t \geq 2$ we have the distribution $p(R_t = 0) = 1$. The capacity of the motor channel without knowledge is then equal to that of the corresponding sub-channel:

$$C(A_t \to S_{t+1}) = C(A_t \to S_{t+1}|R_t = 0) = 1.$$

3.4 Time-Extended Channels

Dealing with channels that may include more than one action is a delicate topic. Unless specified otherwise, the rest of this thesis does not make use of the content of this section, so it is safe for the reader to skip it. However, one of the contribution of the thesis to be found in this section is the connection between the perception-action loop formalism introduced in (Klyubin 2007) and the concepts of directed information (Massey 1990) and feedback capacity (Tatikonda and Mitter 2009).

3.4.1 Horizons

So far we have only looked at communication channels spanning one time-step. However one can be interested in looking at channels involving several time-steps, i.e. having a longer temporal horizon. There are numerous such channels which we will classify according to the terminology introduced in (Klyubin 2007). A distinction is made between interleaved and non-interleaved channels:

- Non-interleaved channel: the channel goes from a set of consecutive actions to sensors that appear just after the last action of the set, for example the motor channel that goes from actions A_t, A_{t+1}, A_{t+2} to R_{t+3} (see Fig. 3.10).
- Interleaved channel: intermediate sensor variables are taken into account as part of the channel output. For example the motor channel going from actions A_t, A_{t+1}, A_{t+2} to $R_{t+1}, R_{t+2}, R_{t+3}$ (see Fig. 3.11).

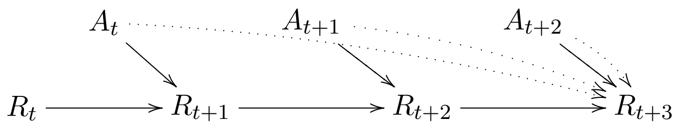

Figure 3.10: Example of the non-interleaved channel $A_t, A_{t+1}, A_{t+2} \rightarrow R_{t+3}$ without context. Solid arrows: causal link. Dashed arrows: channel. Intermediate Rs are left out of the channel output.

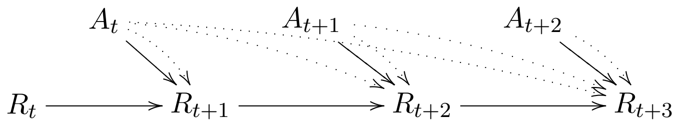

Figure 3.11: Example of the interleaved channel $A_t, A_{t+1}, A_{t+2} \rightarrow R_{t+1}, R_{t+2}, R_{t+3}$ without context. Solid arrows: causal link. Dashed arrows: channel. Intermediate Rs are included in the channel output.

One can also include more destination variables, for example by adding R_{t+4} in both examples above, without changing the class that the channel belongs to, as long as these destination variables are not parent of the source variables on the Bayesian graph.

We define an horizon as the number of timesteps of the perception-action loop that are considered in the channel. It is useful to introduce a distinction between two horizons: (i) the source horizon that specifies how many actions are considered and (ii) the destination horizon which specifies how many destination variables appearing *after* the actions are included.

For instance, the channel in Fig. 3.10 has source horizon 3 and destination horizon 1. With a destination horizon of 2 this channel would be $A_t, A_{t+1}, A_{t+2} \to R_{t+3}, R_{t+4}$, and its interleaved counterpart would be $A_t, A_{t+1}, A_{t+2} \to R_{t+1}, R_{t+2}, R_{t+3}, R_{t+4}$.

In the following chapters of this thesis, only non-interleaved channels will be considered.

3.4.2 Directed Information and Feedback

When the perception-action loop is studied for only one timestep, the quantities are quite straightforward to define. The context-free capacity of the sensorimotor channel is defined as

$$C(A_t \to S_{t+1}) = \max_{p(a_t)} I(A_t; S_{t+1}). \tag{3.12}$$

The contextual capacity of the sensorimotor channel with context M_t (any set of variables accessible to the agent prior to using the channel) is defined as

$$C(A_t \to S_{t+1}|M_t) = \max_{p(a_t|m_t)} I(A_t; S_{t+1}|M_t). \tag{3.13}$$

However, when larger horizons are considered, these capacities are more complicated to define. Different usages of the channel are possible and the expression for the capacity has to take the usage into account. Larger horizons are not used with feedback in the remaining of this thesis, so the reader can safely skip this section. But for the sake of completeness, and because this aspect is an improvement on the framework described in (Klyubin 2007), we now describe how to deal with larger horizons and integrate feedback.

Consider the simplest case of the non-interleaved motor channel with source-horizon 2 and destination-horizon 1 whose causal Bayesian graph is depicted on Fig. 3.12 left. The channel of interest is therefore $A_t, A_{t+1} \to R_{t+2}$. This channel can be used in many different ways, one of them is by using a context variable as described in the previous section.

However, when the source horizon is increased, new usages have to be distinguished. Following (Klyubin 2007), a first distinction between open-loop an closed-loop usage can be made:

- *Open-loop* usage: input variables are not allowed to depend on intermediate output variables, the channel is used without feedback (see Fig. 3.12 top).

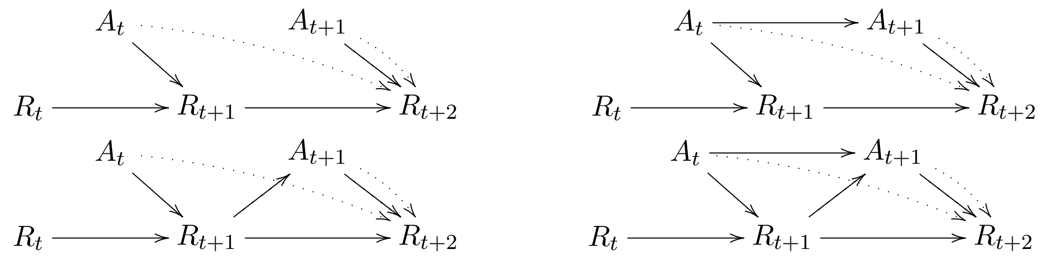

Figure 3.12: Four possible uses of the non-interleaved motor channel with source horizon 2 and destination horizon 1 without context. Solid arrows: causal links. Dashed arrows: channel. Top/bottom: open-loop/closed-loop usage. Left/right: independent actions/action sequences.

- *Closed-loop* usage: input variables may depend on intermediate output variables, the channel is used with feedback (see Fig. 3.12 bottom).

Also we can introduce a further distinction between independent actions and action sequences usage:

- *Independent actions* usage: input variables are not allowed to directly depend on each other (actually previous inputs) (see Fig. 3.12 left).
- *Action sequences* usage: input variables may directly depend on previous input variables (see Fig. 3.12 right).

These distinctions are important in two respects. First they define the set of input distributions upon which the maximization is performed to get the capacity. Secondly they also specify how the mutual information has to be computed. For the different cases presented above we have:

- Open-loop independent actions (Fig. 3.12 top left): input distributions of the form $p(a_t)p(a_{t+1})$, mutual information $I = I(A_t; R_{t+2}) + I(A_{t+1}; R_{t+2})$.
- Open-loop action sequences (Fig. 3.12 top right): input distributions of the form $p(a_t)p(a_{t+1}|a_t)$, mutual information $I = I(A_t; R_{t+2}) + I(A_{t+1}; R_{t+2}|A_t)$ (this can also be expressed as $I(A_t, A_{t+1}; R_{t+2})$).
- Closed-loop independent actions (Fig. 3.12 bottom left): input distributions of the form $p(a_t)p(a_{t+1}|r_{t+1})$, mutual information $I = I(A_t; R_{t+2}) + I(A_{t+1}; R_{t+2}|R_{t+1})$.
- Closed-loop action sequences (Fig. 3.12 bottom right): input distributions of the form $p(a_t)p(a_{t+1}|r_{t+1}, a_t)$, mutual information $I = I(A_t; R_{t+2}) + I(A_{t+1}; R_{t+2}|R_{t+1}, A_t)$.

To convince oneself that the last expression for the mutual information encompasses the previous ones, it suffices to see that the second term $I(A_{t+1}; R_{t+2}|R_{t+1}, A_t)$ can be replaced by $I(A_{t+1}; R_{t+2}|A_t)$ when no feedback is used, by $I(A_{t+1}; R_{t+2}|R_{t+1})$ when actions do not directly depend on previous ones, and by $I(A_{t+1}; R_{t+2})$ when the second action is independent from the first action and the intermediate input.

As one can see from the previous list, the 'naive' expression for the mutual information $I(A_t, A_{t+1}; R_{t+2})$ is only valid for the open-loop case. Let us introduce the following notation which stands for the expression that properly integrates feedback:

$$I(A_t, A_{t+1} \to R_{t+2}) = I(A_t; R_{t+2}) + I(A_{t+1}; R_{t+2}|R_{t+1}, A_t). \tag{3.14}$$

Generalizing this expression for source horizon N, one obtains:

$$I(A_t^N \to R_{t+N}) = \sum_{n=0}^{N-1} I(A_{t+n}; R_{t+N}|R_{t+1}^{n-1}, A_t^{n-1}). \tag{3.15}$$

In the case of the interleaved channel, the intermediate output variables have to be accounted for in the mutual information, leading to the expression:

$$I(A_t^N \to R_{t+1}^N) = \sum_{n=0}^{N-1} I(A_{t+n}; R_{t+1+n}^{N-n}|R_{t+1}^{n-1}, A_t^{n-1}). \tag{3.16}$$

This quantity has been identified in (Massey 1990) and is referred to as *directed information*. Although Klyubin was unaware of this work, he managed to formulate an equivalent quantity by introducing an external variable Z as the source of injected information. Its formulation and the one presented above are actually equivalent to Massey's formulation:

$$I(A_t^N \to R_{t+1}^N) = \sum_{n=0}^{N-1} I(A_t^{n+1}; R_{t+1+n}|R_{t+1}^n). \tag{3.17}$$

Directed information has been introduced in order to account for communication channels used with feedback. Standard treatment of feedback, as can be found in (Cover and Thomas 2006), is usually only presented for memoryless channel. It is shown that the feed-forward capacity of such a channel is the same as the feedback capacity. Put another way, feedback does not increase capacity of a discrete memoryless channel. However, it can help simplify the encoding.

Even though it was generally admitted that feedback could improve capacity for a channel with memory, it took almost ten years (Tatikonda 2000) to use Massey's result and provide a formulation of the feedback capacity along with algorithms to compute its value. The feedback capacity is then expressed as a maximization of directed information. We refer the reader to (Tatikonda 2000, Tatikonda and Mitter 2009) for a detailed treatment of feedback capacity.

An important relation has been identified by (Massey 1990):

$$I(A_t^N \rightarrow R_{t+1}^N) \leq I(A_t^N ; R_{t+1}^N) \tag{3.18}$$

with equality if the channel is used without feedback (which is the case if only one step is considered).

Coming back to our quantities of interest, it is now clear that the capacity has to be expressed in the general case as a maximization of directed information. Standard mutual information being only valid in the open-loop case. Therefore, the general formulation of the capacity for an non-interleaved channel with source horizon of length N and destination horizon of length M is:

$$C\big[A_t^N \rightarrow R_{t+N}^M\big] = \max_{\mathcal{P}} I(A_t^N \rightarrow R_{t+N}^M) \tag{3.19}$$

and for the interleaved channel:

$$C\big[A_t^N \rightarrow R_{t+1}^{N+M-1}\big] = \max_{\mathcal{P}} I(A_t^N \rightarrow R_{t+1}^{N+M-1}) \tag{3.20}$$

where $\mathcal{P}$ is the set of input distributions over which the maximization is performed (which depends on the channel usage).

In order to take a context into account, for example M_t, one has simply to condition the directed information with the context variable. For the interleaved case this would lead to:

$$C\big[A_t^N \rightarrow R_{t+1}^{N+M-1} | M_t\big] = \max_{\mathcal{P}} I(A_t^N \rightarrow R_{t+1}^{N+M-1} | M_t) \tag{3.21}$$

where the conditional directed information is defined as:

$$I(A_t^N \to R_{t+1}^N | M_t) = \sum_{n=0}^{N-1} I(A_t^{n+1}; R_{t+1+n} | R_{t+1}^n, M_t). \tag{3.22}$$

An example of such a situation is depicted on Fig. 3.13.

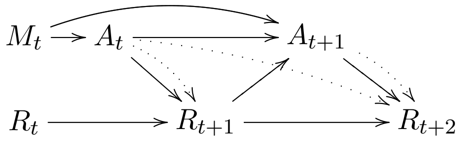

Figure 3.13: Closed loop and action sequences usage of an interleaved motor channel with source horizon 2 and destination horizon 1 using context M_t. Solid arrows: causal links. Dashed arrows: channel. The corresponding capacity is defined as $C(A_t, A_{t+1} \to R_{t+1}, R_{t+2} | M_t)$. The set of input distributions is of the form $p(a_t|m_t)p(a_{t+1}|r_{t+1}, a_t, m_t)$.

Event though context and feedback may appear similar, there is a crucial difference between them. A context refers to any information that is accessible to the agent *prior* to using the channel. On the other hand, feedback refers to information that becomes available *while* the channel is being used.

In the remaining part of this thesis we mainly consider channels with source and destination horizons of 1 or that do not make use of feedback. Therefore, standard mutual information can be used instead of directed information.

3.5 Summary

This chapter introduced the basic framework behind the information-theoretic study of the perception-action loop. The agent and the environment are described as probabilistic relationships between sensors, actuators, memory, and state of the environment at different time-steps. In the resulting causal Bayesian graph, one can identify channels with different properties: a sensor channel, a motor channel and a sensorimotor channel that encompasses both. On the first hand, the capacity of the sensor channel gives an absolute limit on the amount of information that can be acquired about the environment. On the other hand, the motor channel puts a limit on the amount of information that can be injected by the agent into the environment. Put together, a third channel is obtained, the sensorimotor channel, which has its own limitations on the amount of information that can by imprinted by the agent onto the environment and later reacquired through the sensors. This channel is at the core of the concept of empowerment described in the next chapter.

There are many possible sensorimotor channels that can be considered. One of the parameters is the size of the temporal horizon, i.e. how many actions can the agent perform and how many sensor steps ahead is it looking at. Another important factor is the information which is available to the agent for its decision making. The associated quantities of interest, i.e. capacities, have been properly defined using the concept of directed information (Massey 1990) and its use in feedback channels (Tatikonda 2000, Tatikonda and Mitter 2009). These advances allowed to uniformize the information-theoretic framework presented in (Klyubin 2007) and avoid extra variables that were needed to formulate the quantities.

Chapter 4

Empowerment and its Maximization

4.1 Empowerment

The concept of information-theoretic empowerment has been introduced in (Klyubin 2007) and is defined as

'an agent-centric quantification of the amount of control or influence the agent has and perceives.'

This idea is motivated by two main considerations:

- Evolution only gives very sparse feedback to guide the adaptation of behaviours. Especially if one considers an individual's lifetime, other mechanisms have to be sought, such as reinforcement learning. When used in an engineering perspective, such mechanisms have the problem of introducing a semantic bias from the designer. This is because he has to specify explicitly what is 'good' or 'bad' for the agent. In a lot of situations such information is not available, so there must be other, agent-centric measures that 'soften' the landscape and allow the agent find and strive for relatively good situations.
- Whatever an agent has to do in order to survive, and therefore to be considered adapted, it has to be *able* to do it. If the agent does not physically have the ability to do it, then it will not survive. For example a bacteria that feeds on sugar but is unable to perceive its presence (directly or indirectly) will not be very competitive

against one that can follow gradients of sugar. The same applies to a bacteria that has no proper mobility and can only move in a completely random way. Being able to perceive the location of sugar and to move towards it brings a better control on which situations the bacteria will end up into. Therefore, this ability to control the environment is a prerequisite to performing any adapted behaviour [1].

Empowerment, i.e. the ability to control the environment and perceive this control, fits perfectly in the kind of quantity that we are looking for. It has the following properties:

- it is an agent-centric quantity: only information accessible by the agent is necessary (i.e. sensorimotor data),
- it is local: no global knowledge about the world is necessary, and a temporally limited amount of sensorimotor data is sufficient to obtain estimates,
- it is well-defined and computable: because of its channel formulation, standard information-theoretic quantities can be used,
- it is semantically unbiased: the designer does not introduce any external value system, this is entirely resulting from the agent/environment coupling.

4.1.1 Definition

Empowerment, denoted by $\mathfrak{E}$, is defined as the capacity of the sensorimotor channel (see Fig. 4.1). According to (Klyubin 2007), context-free empowerment with source and destination horizons of 1 is defined as (Klyubin 2007, Klyubin et al. 2008):

$$\mathfrak{E}(A_t \to S_{t+1}) = C(A_t \to S_{t+1}). \tag{4.1}$$

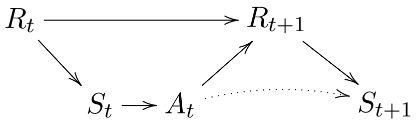

Figure 4.1: Section of a typical perception-action loop with source and destination horizons 1. Solid arrows: causal links. Dashed arrow: channel.

When the agent is allowed to use a context (i.e. a set of variables that the agent has access to before using the channel), for example in the basic perception-action loop of

[1] Of course, sometimes no control is needed in order to achieve successful behaviour. For example if the bacteria is living in an environment uniformly and constantly filled with sugar, sensing and moving add nothing to the fitness. Any 'blind' bacteria will be as good as any other. This can be quantitatively measured using the information-theoretic formulation of *relevant information* presented in (Polani, Martinetz and Kim 2001, Polani et al. 2006)

Fig. 4.1 where the context is S_t, contextual empowerment is defined as the capacity of the sensorimotor channel conditioned on the context variables. In this case it is defined as

$$\mathfrak{E}(A_t \to S_{t+1}|S_t) = \sum_{s_t} p(s_t) C(A_t \to S_{t+1}|s_t). \tag{4.2}$$

The set of input distributions over which the maximization is performed for finding the channel capacity depends on the actual usage of the channel. It was shown in the previous chapter (i.e. Section 3.4) that the most general formulation of the capacity has to use directed information in order to properly account for feedback process. Using this formulation, the quantities described above can be directly generalised to longer horizons, interleaved/non-interleaved channels, and open/closed loop control.

However, the rest of this thesis only investigates situations in which temporal horizons are of size 1 or in which no feedback is used. Therefore, because of the equivalence between directed information and mutual information when the channel is used without feedback, standard mutual information will be used to express and compute the quantities of interest.

4.1.2 Interpretation

Because empowerment is the basis of this work, how it can be interpreted has to be very clear to the reader. To understand why empowerment is a crucial ‘commodity’ for cognitive agents it is useful to first consider cases where this commodity is lacking.

Situations with no empowerment

The first situation can be referred to as the *static sensorimotor channel* (see Fig. 4.2 left). In such a channel, the sensory outcome of any actions is always the same symbol s_t. Therefore whichever action a_t is picked by the agent, the resulting sensor reading will be the same. This can occur in two situations. Either the motor channel is not able to inject any information in the environment, or the sensor channel is not able to pick up any of this information.

A second situation for zero empowerment is the *random sensorimotor channel* (see Fig. 4.2 right). In this channel the outcome of any action is a randomly chosen sensor value with no dependence on the previous action. Similarly to the static sensorimotor channel, there is no way for the agent to behave so as to alter its performance at the task. Indeed the completion of the task, such as reaching a particular sensory state, has no link whatsoever with the policy that the agent implements.

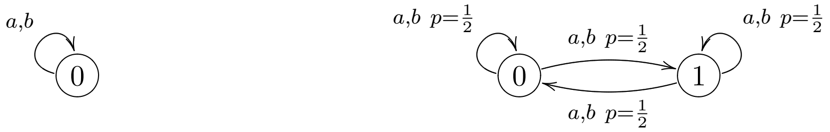

Figure 4.2: Examples of environments with no empowerment. States of the automaton are equivalent to states of the environment and sensor readings. Actions are taken in the set $\{a, b\}$ and are specified on the transitions of the automaton (along with the probability if not deterministic). **Left:** static sensorimotor channel. **Right:** random sensorimotor channel.

It is easy to see that, independently of the task that we expect the an agent to perform in such a channel, there is no strategy that accomplishes this task better than any other. Indeed, whatever the policy of the agent is, the performance at completing the task will be the same. So, if one assumes that information processing has a cost (as advocated in (Laughlin, de Ruyter van Steveninck and Anderson 1998)), then agents behaving in such channels should exhibit no cognitive abilities, for the simple reason that, however complex the cognitive processing is, it brings absolutely no benefit to the agent. Therefore the parsimony principle effective at the evolutionary scale should get rid of the cognitive apparatus of such an agent.

Situations with high empowerment

Imagine the following setup: the agent lives on a discrete line; it perceives its absolute position; it can move instantaneously to any position on the line, making it some kind of perfect entity in this world (see Fig. 4.3). If the environment has N positions, its empowerment will be exactly $\log N$ bits (the sets of sensor states and actions need at least the same cardinality). However, if one adds constraints on the allowed moves, for example being restricted to the neighbouring positions, then empowerment will be reduced accordingly. In some way, empowerment measures the number of perceivable trajectories available to the agent.

If the perfect agent described above had to achieve a control task, for example reaching a target, it would have absolutely no difficulty into achieving it in a single step without much processing. Using a context is not even required to achieve maximum empowerment. On the other hand, if the agent's mobility is constrained, then achieving the task will require more steps and in some cases it will also require to use a context (such as the

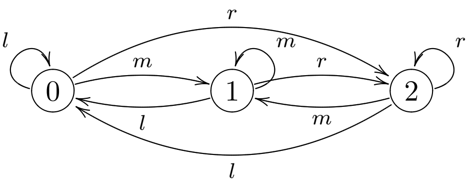

Figure 4.3: Automaton representing a variation of the line world of size 3. States of the automaton are equivalent to states of the environment and sensor readings. Transitions are deterministic and labelled with the three possible actions: left, middle, right $\{l, m, r\}$. Each action allows the agent to reach a unique position on the line, regardless of the starting position. Empowerment in this environment is maximum (perfect controllability and observability).

current absolute position).

4.1.3 Potential and Actual Information Flow

It is important to understand that empowerment, because it is defined as a capacity, is a potential quantity. The actual controller of the agent is not considered. Indeed, the empowerment is the maximum amount that the agent *could* send using *some* controller. It has to be distinguished from the actual information flow, i.e. the amount of information that the agent actually sends by using its controller. Because of this property, empowerment is a policy-independent quantity [2]. But what is kept is the operational constraints of this controller, i.e. the embodiment and the information accessible to the controller.

4.1.4 An Objective Measure which Incorporates Subjectivity

Empowerment has the interesting property of being an objective measure. More precisely, if the environment and the constraints on the controller are given, empowerment is an exactly defined quantity. It can also be computed using algorithms for finding channel capacities.

Also, by considering the information that is accessible to the controller as a context, it accounts for the subjectivity of the agent. Empowerment quantities crucially depend on this information. This is fundamental in the sense that this is all that the agent can ever ”know”. The knowledge that the agent has about the state of the environment is what constitutes its subjective vision of the world.

[2] However, empowerment can be modulated by the policy, this is the topic of the next sections.

However, this general objectivity is available only to an observer external to the system, that knows exactly the real distributions that describe the causal Bayesian network. If empowerment is to be calculated inside such an agent, then only estimates of empowerment are available. These estimates can be constructed by collecting statistics about previous sensorimotor experience and using capacity computation algorithms. Working with estimates adds a second kind of subjectivity that the external observer does not suffer.

4.1.5 Relation to Control Theory

As mentioned in the definition, empowerment is a measure of the *potential perceivable control* an agent has on its future. But here control does not mean reducing the entropy of a process (as in Touchette2000,Touchette2004), it is about being able to inject information, therefore to create future entropy that is correlated with the agent's actions.

However, there are some close relationships between these perspectives. Capacity of the sensorimotor channel and entropy reduction are two aspects of the same process. Having 1 bit of capacity means that the entropy of the outcome can be reduced by 1 bit from the maximum possible.

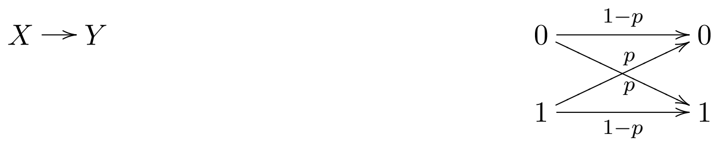

Figure 4.4: Binary symmetric channel. The binary input X is copied into the binary output Y with probability $1 - p$ and inverted with probability p. **Left:** causal Bayesian graph. **Right:** input/output mapping.

Consider the noiseless binary symmetric channel ($p = 0$) (see Fig. 4.4). Because the binary input is transmitted without noise through the channel, the capacity is 1 bit (achieved for uniform distribution). The maximum entropy of the output is also 1 bit (again with uniform distribution at the input). This means that an entropy reduction of 1 bit can be achieved, basically by using a deterministic distribution for the input, leading to 0 bits of entropy at the output.

If noise is added ($p = \frac{1}{4}$), the capacity becomes approximately 0.19 bits. The maximum entropy of the output is still 1 bit, therefore the entropy reduction is at most 0.19 bits. It follows that the entropy of the channel output can be reduced to approximately 0.81 bits.

4.2 Empowerment Maximization

This section describes three mechanisms that allow an agent to modify its empowerment. Each mechanism implies to alter different parameters of the model:

- Evolving sensors and actuators: this requires the agent to change its embodiment or the potential interaction it can have with its environment.
- Navigating through the environment towards empowered states: this requires to adapt the controller of the agent
- Acquiring a context: this requires to adapt a memory mechanism that transforms past sensorimotor experiences into contextual information about the actual state of the sensorimotor channel.

4.2.1 Evolution of Sensors and Actuators

Having proper sensors and actuators is a requisite for having empowerment. Not only these have to be separately adequate, i.e. sensors have to be able to extract information from the environment state, and actuators have to allow the injection of information in the environment, but they also have to match each other. For example, an agent which is able to see but whose actuators allow only to emit sound has no empowerment, even though both sensors and actuators meet the required conditions. From the perspective of empowerment, the sensorimotor system is an integrated system. It is the potential interplay between sensors and actuators through the environment that defines empowerment.

It has been suggested that information processing in living beings, and mainly acquisition of sensory information, is an expensive phenomenon (see (Laughlin 2001, Laughlin et al. 1998)). The authors give the example of the fly's eye whose operation consumes around 10% of the metabolic energy of the organism. From an evolutionary perspective, such a cost has to bring some benefits, otherwise mutant which have no eyes should be selected. It is suggested in (Jeffery 2009) that this phenomenon has led to the evolution [3] of the blind cave fish (*Astyanax mexicanus*). Indeed because it lives in a dark environment, developing an eye and processing the associated sensory information has a huge metabolic cost which can be avoided. This is a typical case where the sensor channel has evolved (or rather de-evolved) because the visual aspects of perception are of no use in

[3] Is not known yet how much of this phenomenon is due to actual evolution or to the developmental process of the fish. Other fishes from the same specie that live in lighted areas keep their eyes. Actually it seems that the development of the eye starts in both cases, but it is stopped early in its development when there is no visual stimulation.

this environment.

By considering empowerment as a utility for evaluating a sensorimotor apparatus, this allows to take into account both the sensor part and the motor part as an integrated system. It is then possible to quantify the contribution of various sensor of motor systems to empowerment. Following the parsimony principle, parts which contribute less to empowerment can be discarded so that information processing focuses on the highly contributing parts. The advantage of empowerment in this case is that it provides immediate and local feedback to the adaptation. Instead of waiting for the fitness feedback from evolution, empowerment can be estimated and used in order to select sensorimotor apparatus.

Such a sensorimotor evolution is experimented in (Klyubin 2007). An agent is allowed to move on a 2D grid with a chemical gradient at its centre. The sensor of the agent is made of a set of points which are spatially distributed in the vicinity of the agent. The sensor reading gives the index of the point which reads the highest gradient. Using a genetic algorithm, spatial positions of the sensor points are evolved for different starting positions of the agent. The fitness function measures the empowerment of the agent (4-steps open-loop non-interleaved empowerment) and includes a penalty term for the number of sampling points. The idea is to find sensors that maximize empowerment while limiting the amount of information acquired. Results of the evolution show that specific sensors are generated that depend on the initial location of the agent, typically these change from an agent-centred cluster of sampling points for initial position starting near the centre of the gradient, to an arched layout for starting positions away from the centre. The same methodology is also used to evolve actuators whose layout depend on the initial position of the agent. Of course both aspects can also be evolved together.

4.2.2 Modification of the Environment

Consider the following situation. There is an agent existing in an abstract state-space of set R. If the agent has nothing specific to perform, no task given, is there a natural preference for some states? Empowerment brings a positive answer to this question.
By being able to compute and compare the empowerment of an agent in different states of the environment, one can construct a value system which gives a preference to states that have high empowerment. In a general sense, these states are preferable for the agent because it can 'move' to other states (and perceive this move) more easily.

One very illustrative example is presented in (Klyubin 2007) (see Fig. 4.5). In this example, the agent is moving in a 2D-maze, at each move it can only reach its four immediate neighbours. Its sensor reads the absolute location of the agent in the maze. By looking at the empowerment of the agent at each position on the map and plotting it on the same map one can see that empowered places are those with high centrality properties (an aspect which is studied in (Anthony, Polani and Nehaniv 2008)). Basically places from which many other places are reachable in a short amount of moves are preferred. Consequently, dead-ends and places surrounded by many walls are negatively valued.

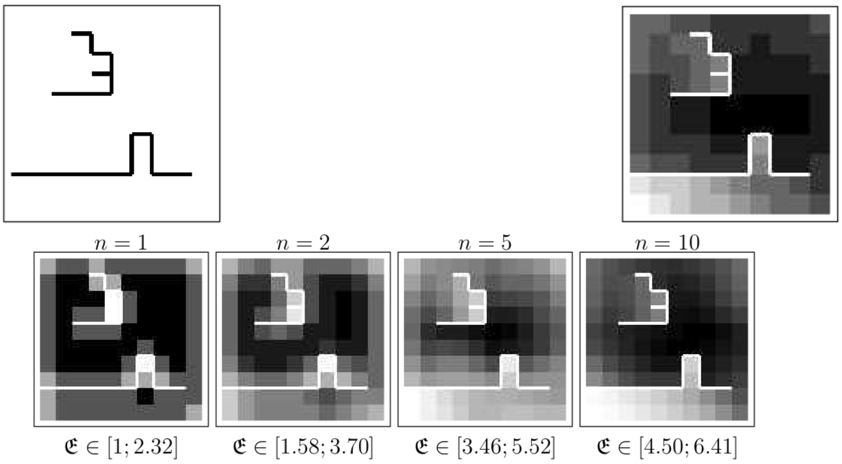

Figure 4.5: Klyubin's maze experiment. **Top left:** 10×10 maze with walls in black. **Top right:** map of average distance to other places. Shortest distances in dark. **Bottom:** empowerment maps for various source horizons n (dark for high values). In this environment, empowerment is directly related to the average distance to other places. Reproduced from (Klyubin 2007)

It is easy to understand that, in such an environment, empowerment can be maximized by navigating between the states towards those at the centre of the maze. Therefore an agent starting in any position would be expected to follow the empowerment increasing gradient and then stop in a highly empowered state.

The maze example can be used as an analogy to other scenarios. Any environment can be described as a state-space (even though it does not necessarily imply an euclidean space as in the maze). If the agent possesses a model of the environment, it is able to compute the empowerment and 'navigation' through the states of the environment in order to reach highly empowered states. This 'navigation' is meant in a very general way, it is not necessarily spatial. It is basically about controlling the environment in order to reach these specific states.

Empowerment-Maximizing Policy

Let us assume that the agent is using the last sensor state as a context. After some time experimenting with its perception-action loop and collecting statistics, it is able to construct a subjective model of the environment: $p(s_{t+1}|a_t, s_t)$ and $p(s_t)$. Using these statistics, the agent can estimate the empowerment $\mathfrak{E}(A_t \to S_{t+1}|s_t)$ for each context state. Using this estimate and the statistical model, one can compute a behaviour policy that maximizes the expected empowerment of the agent. This can be done using standard dynamic programming techniques such as the Bellman equation. These principles can also be applied to longer horizons, and to situations where the agent is equipped with a different more complex context.

What should be expected from such a policy? In the case where the agent is the only source of noise, it should move towards the empowerment gradient and stop on a highly empowered state. If noise is introduced then the agent will try to counteract this noise by going back towards the highly empowered state thereby correcting the perturbation. More generally this can be considered as some kind of homeostatic behaviour where the 'viability' region is a direct consequence of the structure of the state-space and the agent's abilities to navigate through it.
This kind of behaviour can be put in contrast with other self-motivated techniques such as homeokinesis (Der et al. 1999) or maximization of predictive information (Ay et al. 2008) where static situations are avoided and dynamic ones are sought.

It is important to note that it does not matter whether the last sensory state accurately represents the state of the environment. As long as it captures some information about it, it can be used to navigate. Of course a more accurate variable would improve on empowerment and the navigation.

4.2.3 Maximization Using Contexts

The previous chapter described the effect of a context, a variable accessible to the agent that captures side-information about the state of the sensorimotor channel. Possessing a context can strongly increase the empowerment of the agent. This was illustrated by the XOR channel example of section 3.3.1 in which context-free empowerment is zero whereas it becomes maximum when the context is known.

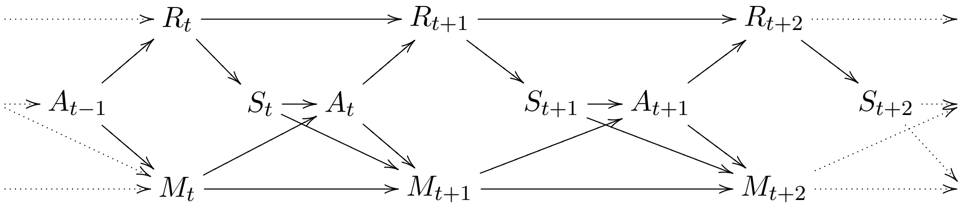

Figure 4.6: Perception-action loop of an agent that has a memory accounted for by M_t. It is used as a context together with the current sensor reading S_t. The memory is a function of its previous state M_{t-1}, the last action A_{t-1} and the previous sensor state S_{t-1}.

A variable that could be used as a context by the agent is the last sensor reading S_t. However one can consider cases where this information is not accessible to the agent. Many other situations can be envisaged. One is to model a memory variable M_t such as the one presented on Fig. 4.6 and use this variable as a context, or a combination of the last sensor reading and the memory. For clarity we will refer the variable C_t as being a general context. It can be a compound of any random variables appearing before the agent's action on the causality path.

Ideally, the context captures some of the side-information of the sensorimotor channel. Because this information is contained in R_t, it has to be correlated with it. However, from an agent-centric point of view, the only accessible variables that may contain such information is the set of all previous sensor and motor states. Therefore, in order for the agent to construct such a context variable it has to store, process and potentially compress past sensorimotor data through some mechanism.

Using empowerment as a utility for selecting this mechanism is again a very useful technique. This provides the agent with local and immediate feedback about the performance of a mechanism in terms of how much empowerment is increased compared to other mechanisms. Therefore, empowerment maximization can be done by searching the space of context extraction mechanisms. Such an approach was also presented in (Klyubin 2007) where a context automaton is evolved for an AIBO robot that allows it to identify whether there is a book upfront or not using sensorimotor data.

In the next section, I will present an in-depth theoretical treatment of the concept of context and derive new bounds on context-based empowerment maximization. Using these results, a new empowerment estimation technique is presented that avoids the calculation of actual empowerment and allows faster searches in the space of context-extraction mechanisms. Improving the methodology used in (Klyubin 2007) by using this new estimate is one of the contributions of this thesis.

A more significant contribution is the derivation of analytical solutions to empowerment maximization. These analytical solution are obtained by using an information-bottleneck like approach (Tishby et al. 1999) that compresses the sensorimotor history into a context variable. Two solutions are presented: one for the maximization of empowerment, and the other for the maximization of empowerment estimates introduced in the next section.

Based on these solutions, two iterative algorithms are presented that allow to efficiently compute a context variable given some statistics about the perception-action loop and the sensorimotor experience.

It is important to note that context extraction and environment navigation can be used together. Starting with a completely ignorant agent, one can collect sensorimotor data (through random motor 'babbling' or another exploration policy) and use this data to create a context variable. Using the collected statistical model of the environment and the context, the agent can 'navigate' between states in order to reach those that have high empowerment.

4.3 Context Extraction

4.3.1 Theoretical Treatment

Let us consider the causal Bayesian graph presented on Fig. 4.7. For simplicity reasons time indices have been dropped. The variables are:

- the original full state of the environment R,
- the context variable C that may capture information from R,
- the action variable A that may depend on the context, and
- the sensor state S resulting from the action and the original state of the channel[4].

The sensorimotor channel that goes from A to S has different empowerment values depending how it is used:

[4] In the full perception-action loop model, the causal link from R to S has actually to go through the environment, i.e. $R_t \rightarrow R_{t+1} \rightarrow S_{t+1}$. However, because this intermediate variable is inaccessible in our model, it can be made implicit for this theoretical treatment.

$$\begin{tikzcd}
 & C \arrow[r] & A \arrow[dr] & \\
R \arrow[ur] \arrow[rrr] & & & S
\end{tikzcd}$$

Figure 4.7: Simple stateful sensorimotor channel with a context.

- the context-free empowerment $\mathfrak{E}(\cdot)$,
- empowerment with context C: $\mathfrak{E}(\cdot|C)$, and
- empowerment with maximal context: $\mathfrak{E}(\cdot|R)$ (this requires changing the Bayesian graph).

Because having a context can only increase capacity, and because R is a maximal context, these quantities can be ordered as

$$\mathfrak{E}(\cdot) \leq \mathfrak{E}(\cdot|C) \leq \mathfrak{E}(\cdot|R). \tag{4.3}$$

Capacity Gain from Context

In the work of (Klyubin 2007), it was made clear that having a context increases empowerment. However, the amount of empowerment that can be gained from the context was only investigated in one experiment. This is the question I address in this section from a theoretical perspective.

In order to understand how the context can increase empowerment it is easier to take the channel perspective. By doing so, one can distinguish different information rates and capacities:

- the context-free capacity $\max_{p(a)} I(A;S)$ which is achieved for $p^*(a)$,
- the rate of information sent if the context is added a posteriori: $I_{p^*(a)}(A;S|C)$,
- the context-capacity when the agent is not allowed to pick actions according to the context: $\max_{p(a)} I(A;S|C)$ (in the information-theory literature this is referred to as the capacity when side-information is known at the receiver),
- the context-capacity when the agent can pick actions according to the context: $\max_{p(a|c)} I(A;S|C)$ (or capacity when side-information is available at both the emitter and the receiver).

Again these quantities can be ordered as follows:

$$\max_{p(a)} I(A;S) \leq I_{p^*(a)}(A;S|C) \leq \max_{p(a)} I(A;S|C) \leq \max_{p(a|c)} I(A;S|C). \tag{4.4}$$

We are interested in how much capacity can be gained by using the context C. This gain can be expressed as

$$\Delta C = \max_{p(a|c)} I(A; S|C) - \max_{p(a)} I(A; S). \tag{4.5}$$

It is possible to identify a lower bound on the capacity gain ΔC.

Theorem 4.3.1. *The capacity gain ΔC is bounded from below by the synergy when the channel is used at context-free capacity:*

$$\Delta C \geq I_{p^*(a)}(A; C|S) \tag{4.6}$$

where $p^*(a)$ is the capacity achieving distribution for channel $p(s|a)$ and the mutual information is computed according to the joint distribution $p(s, a, c) = p(s|a, c)p^*(a)p(c)$. Equality is achieved if the context-free capacity-achieving distribution matches an achieving distribution of capacity with receiver side-information.

Proof. Let $p^*(a)$ be the capacity achieving distribution for channel $p(s|a)$. Using inequation 4.4, one has

$$\Delta C \geq I_{p^*(a)}(A; S|C) - I_{p^*(a)}(A; S). \tag{4.7}$$

This quantity can be transformed through the multi-information:

$$\begin{aligned} I_{p^*(a)}(A; S|C) - I_{p^*(a)}(A; S) &= I_{p^*(a)}(A; S; C) \\ &= I_{p^*(a)}(A; C|S) - I_{p^*(a)}(A; C) \\ &= I_{p^*(a)}(A; C|S) \end{aligned}$$

because A and C are independent in the joint distribution. □

This lower bound, which will be referred to as the *context-synergy at context-free capacity* , allows us to obtain estimates of empowerment for a given context. This result is used in the next section in order to devise new algorithms for generating context-extraction mechanisms.

Side-Information Capture Bound

For the context to increase empowerment, it has to capture some of the side-information of the sensorimotor channel. We now go on to show how the amount of information captured is related to the capacity gain.

Theorem 4.3.2. *The context-synergy at context-free capacity* $I_{p^*(a)}(A;C|S)$ *is bounded from above by the amount of information that the context* C *captures about* R*:*

$$I_{p^*(a)}(A;C|S) \leq I(R;C). \tag{4.8}$$

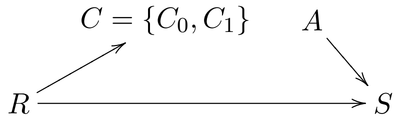

Figure 4.8: Causal Bayesian network used for the proof of theorem 4.3.2. The agent's policy $p^*(a)$ is independent from the context. The context is split into a purified part C_1 containing all and only the correlations with R and the noise part C_0.

Proof. Let us split the context C in two variables C_0 and C_1. This *purification* is done by putting all the information that C captures about R into C_1, and only this information (no extra noise), while all the noise is put in C_0 (and there is no information about R). The corresponding causal Bayesian is presented on Fig. 4.8. Variables C, C_0 and C_1 are related by

$$H(C) = H(C_0) + H(C_1), and \tag{4.9}$$

$$I(C;R) = I(C_1;R) = H(C_1). \tag{4.10}$$

Decomposing the context-synergy at context-free capacity we obtain (the $p^*(a)$ notation is dropped for clarity):

$$\begin{aligned}
I(A;C|S) &= I(A;C_0,C_1|S) && (4.11)\\
&= H(C_0,C_1|S) - H(C_0,C_1|S,A) && (4.12)\\
&= H(C_0|S) + H(C_1|S) - I(C_0;C_1|S) && (4.13)\\
&- H(C_0|S,A) - H(C_1|S,A) + I(C_0;C_1|S,A). && (4.14)
\end{aligned}$$

As C_0 is an independent node, conditioning does not change its entropy. Moreover C_0 and C_1 are totally uncorrelated from each other, whatever the conditioning. Therefore both mutual informations vanish. We can therefore rewrite

$$\begin{aligned}
I(A;C|S) &= H(C_0) + H(C_1|S) - H(C_0) - H(C_1|S,A) && (4.15)\\
&= H(C_1|S) - H(C_1|S,A) && (4.16)\\
&= I(A;C_1|S). && (4.17)
\end{aligned}$$

This mutual information, whether conditional or not, is bounded from above by the entropy of each variable. Therefore we can write that $g(C) \leq H(C_1)$ which by definition is equal to the amount of information captured by C_1 (and C) about R, proving the inequality. □

Interpretation

These two theorems can be interpreted in the following way. The first one shows that one of the components of empowerment increase due to the context is at least equal to the context-synergy of the context-free capacity. Therefore if we have a context whose synergy is equal to one bit then the empowerment when using this context can be increased by at least one bit. It is important to understand that this empowerment increase is obtained in a completely passive way, i.e. without the agent actively using the context. The empowerment may be further increased by allowing the agent to act according to the context. However it was not possible in this thesis to identify an upper bound on the total empowerment gain.

The second theorem shows that this passive increase of empowerment through context extraction is exactly limited by the amount of side-information that is captured by the context. Put another way, passively increasing empowerment by one bit requires to capture at least one bit of side-information. However, the total empowerment increase (including the active part) may be more than the amount of captured information.

4.3.2 Evolving a Context-Automaton

As mentioned before, the side-information is accessible to the agent only through its past sensorimotor data. Therefore at time t the best context available is the set of variables $\{S_t, A_{t-1}, S_{t-1}, A_{t-2}, ...\}$. From a practical perspective, there are memory constraints that prevent the agent from using this set of variables as a context. Instead the agents has to process this information on the fly in order to extract a context variable. This can be represented by the causal Bayesian graph depicted on Fig.4.9.

Evolving such an automaton has been performed in (Klyubin 2007). The author uses a genetic algorithm to search the space of automaton which are described by an arbitrary number of states $|\mathcal{M}|$ and a probabilistic transition matrix $p(m_t|m_{t-1}, a_{t-1}, s_t)$. In order to evaluate a set of candidates, one has first to collect random sensorimotor history for an arbitrary number of time-steps. The general evaluation procedure is the following:

- Step 1: process the sensorimotor history with the candidate automaton.

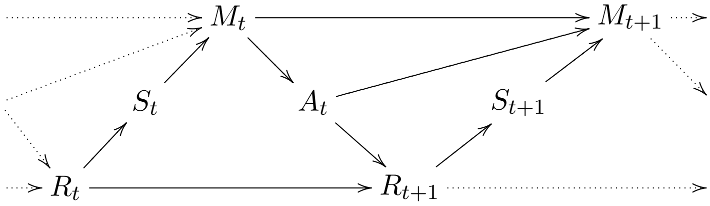

Figure 4.9: Perception-action loop of an agent with a context automaton capturing information in M_t.

- Step 2: collect the resulting distributions $p(s_{t+1}|a_t, m_t)$ and $p(m_t)$.
- Step 3: compute the empowerment with context $\mathfrak{E}(A_t \to S_{t+1}|M_t)$.

Using this methodology, Klyubin was able to evolve a context-automaton for an AIBO robot. This automaton is able to identify whether there is a book in front of the robot or not. The author also noticed that the empowerment gain and the amount of captured information $I(R_t; M_t)$ were correlated. This correlation has been further explained in the previous theoretical treatment.

It is to be noted that the last step of the evaluation of a candidate automaton acts as a bottleneck in the search process. This is because computing empowerment with context requires the use of an optimization algorithm such as Blahut-Arimoto for each possible context state. Such algorithms are typically computationally intensive and as they have to be used for each possible candidate, they may slow down the search process.

Thanks to the lower bound on empowerment gain defined in theorem 4.3.1, it is possible to reduce this bottleneck. The idea is that instead of computing context empowerment, one can estimate it by computing the context-synergy at context-free capacity. Before evaluating candidates one has first to compute the context-free capacity achieving distribution $p^*(a_t)$. Then the evaluation procedure becomes:

- Step 1: process the sensorimotor history with the candidate automaton.
- Step 2: collect the resulting distributions $p(s_{t+1}|a_t, m_t)$ and $p(m_t)$.
- Step 3: compute the context-synergy at context-free capacity $I_{p^*(a)}(A_t; M_t|S_{t+1})$.

Unlike the original one, this algorithm does not provide the exact capacity for a given context automaton but a lower bound. Therefore it might not converge to the best solution, however it will converge to good solutions for a highly reduced computational cost, especially when looking for complex automatons.

Experiment 4.3.1. ***Performance comparison of context-automaton evolution using full empowerment or empowerment estimates***
Objective: *Compare the performance of the original algorithm and the empowerment estimation one.*
Main results: *The algorithm based on empowerment estimates find solutions as good as the original algorithm while dividing the computation time by almost ten.*

The environment in this experiment is a 8 × 8 grid on which the agent can move (see Fig. 4.10). The four possible actions are moving north, east, south and west. If the agent collides with the border of the grid then it stays at its position. The sensor of the agent reads the immediate presence of walls as a set of four binary values, one for each direction.

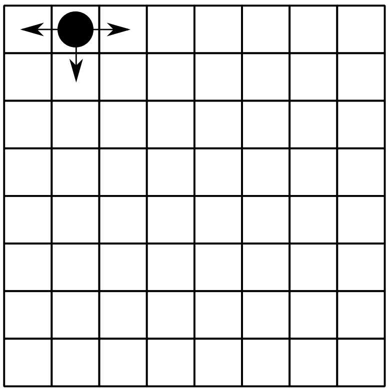

Figure 4.10: 8 × 8 grid world. The agent is represented as a disc and is allowed to move in the 4 directions. The agents collides with the boundaries of the grid.

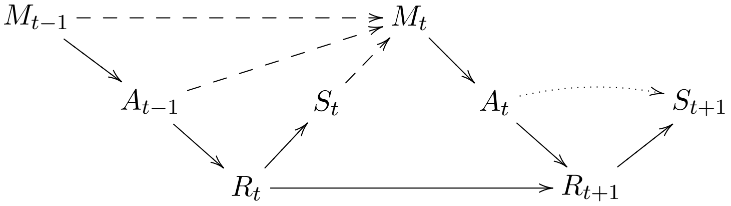

Figure 4.11: Section of the perception-action loop in the context automaton search experiment. Solid arrows: causal links. Dashed arrows: the distribution that is being searched $p(m_t|m_{t-1}, a_{t-1}, s_t)$. Curved dashed arrow: sensorimotor channel under consideration.

The search methodology is the following. Sensorimotor data are collected using a uniform action policy during 100000 time-steps. The search space is restricted to deterministic automata with 6 states. It is searched using a random search with temperature $T = 0.0001$. For the original algorithm, the maximized function is the contextual empowerment $\mathfrak{E}(A_t \to S_{t+1}|M_t)$ (see Fig. 4.11)which is computed using the Blahut-Arimoto algorithm with stopping criterion $\epsilon = 0.0001$ (see (Blahut 1972)). The new algorithm uses the empowerment estimate $I_{p^*(a)}(A_t; M_t|S_{t+1})$.

The following quantities are measured and averaged over 100 trials in order to obtain statistically significant results:

- the computation time per evaluation (only the differing step of the algorithm)
- the increase in empowerment after 1000 evaluation steps

Results of this experiment are shown on the following table[5]:

Quantity	Mean	Standard Deviation
Computation Time - Original (ms)	13.01	1.58
Computation Time - Estimates (ms)	1.72	0.08
Empowerment gain - Original (bits)	0.0925	0.0147
Empowerment gain - Estimates (bits)	0.0927	0.0149

One can see that both methods lead to the same performance in terms of empowerment gain, however the estimates-based algorithm is 7.55 times faster. Of course such values may differ when considering other scenarios or states, but the same tendency was found in other experimental setups (not described here).

Experiment 4.3.2. ***Analysis of the best evolved automata.***
Objective: *Identify what kind of context is provided by the best automata that could be evolved using empowerment or empowerment estimates.*
Main results: *The evolved automaton captures information about the approximate location of the agent in space. When using empowerment estimates, a preference has been given to capturing the horizontal position.*

[5] Computational times are measured using Java on an Intel Core 2 Duo processor running at 1.83 GHz. Single-threaded user time is measured, quantifying only the amount of time spent executing the actual implementation of the algorithm.

The methodology is exactly the same as the previous experiment, apart from the fact that the number of evaluation steps is not arbitrarily bounded. Instead, several searches have been performed and the best automaton ever found has been extracted. A graphical representation of the resulting context automaton when using empowerment maximization is presented on Fig. 4.12. A similar result is given for empowerment estimates maximization on Fig. 4.13.

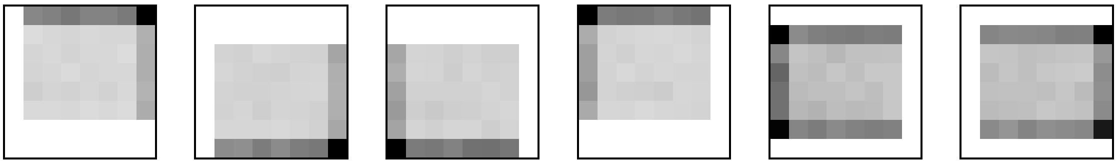

Figure 4.12: Mapping obtained from the best context-automaton evolved using the original algorithm. Each of the six pictures represent the probability of being at a specific location on the grid when the automaton is in a given state, i.e. $p(r_t|m_t)$. Dark means high probability. Context-free empowerment: 0.25 bits, with context: 0.42 bits

As one can see on the figures, the context-extraction algorithm using actual empowerment has created a 6-states automaton that captures information about the spatial position of the agent. In fact a state is dedicated to almost each corner of the grid. For example, if the context automaton is in state 0 (the leftmost one) then there is a high likelihood that the agent is in the upper-right part of the box. Each of the areas distinguished by the context automaton has distinct properties in terms of the perception-action loop.

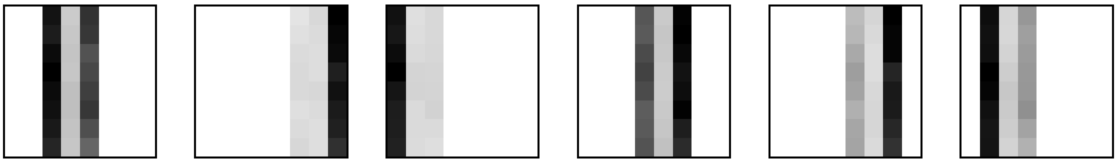

Figure 4.13: Mapping obtained from the best context-automaton evolved using the empowerment estimates algorithm. Each of the six pictures represent the probability of being at a specific location on the grid when the automaton is in a given state, i.e. $p(r_t|m_t)$. Dark means high probability. Context-free empowerment: 0.25 bits, with context: 0.53 bits.

The best context-automaton obtained when maximizing empowerment estimates is quite atypical. Generally the results are qualitatively similar to the maximization with actual empowerment. However, in that case, a high empowerment could be reached by capturing only the horizontal dimension of the grid. One can see however that the areas distinguished by this context automaton are much sharper than those of Fig.4.12.

Interestingly this atypical automaton is the result of the empowerment estimation. What happened is that the context-free capacity achieving distribution $p^*(a_t)$ was biased towards mostly using left and right motor commands (this is due to the noise of the sampling process). Because of this bias, contexts capturing the horizontal position are preferred during the search process.

4.3.3 Bottleneck Approach to Context Extraction

Assuming that the amount of information the context can capture is limited, one can ask how much empowerment is maximally achievable under such a constraint. The best context that is accessible to the agent is the past sensorimotor history. The random variable H will denote the set of variables that constitute the recent sensorimotor experience. This set can be of any horizon and it may or may not include sensors or actuators. Now we are interested in finding a context variable C that will compress as much as possible the sensorimotor history while keeping maximum empowerment (see Fig. 4.14). More precisely we are looking for the probabilistic mapping $p(c|h)$ that maximizes

$$\mathfrak{E}(A_t \to S_{t+1}|C) - \lambda I(H;C) \tag{4.18}$$

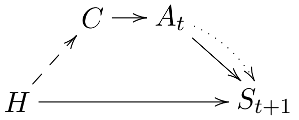

Figure 4.14: Simplified causal Bayesian graph of the bottleneck approach to context extraction. Solid arrows: causal links. Dashed arrow: conditional distribution sought for, i.e. $p(c|h)$. Curved dashed arrow: sensorimotor channel of interest. H is the sensorimotor history (any kind of combination of past sensors and actuator variables). C is the context variable. Intermediate variables inaccessible to the agent (e.g. the state of the environment) are hidden.

This approach is inspired from the *information bottleneck* method presented in (Slonim 2002, Tishby et al. 1999) (see Fig. 4.15). The main idea behind this method is the following. One has a random variable X which is correlated with a label variable Y. The goal is to define a mapping for a bottleneck variable $\tilde{X}$ that compresses the information while preserving as much correlation with the label Y as possible. In order to maximize the function

$$I(\tilde{X};Y) - \lambda I(X;\tilde{X}) \tag{4.19}$$

where λ is a trade-off parameter, the authors derive an iterative algorithm with good convergence properties.

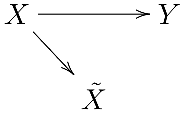

Figure 4.15: Causal Bayesian graph of the information bottleneck. X is the input data that we want to compress. Y is the 'labels' attached to this data. $\tilde{X}$ is the compressed representation of X that correlates as much as possible with the labels Y while minimizing that amount of information captured from the input X.

In this work the methodology is similar, but instead of preserving correlation with a label, the goal is to preserve the empowerment of the agent.

The context variable extracted by using such a technique is very similar to the concept of *causal states* of ϵ-machines (see (Shalizi and Crutchfield 2002, Shalizi 2001)). The context states have the same property as causal states in the sense that they make the future more independent from the past (given the state). However, a crucial difference is that causal states apply to a standalone process, whereas context states capture information that is causally related to the perception-action loop. This means that some causal states of the environment are not taken into account because they have no impact on the perception-action loop. A different approach that takes actuation into account is presented in (Still 2009).

4.3.4 Iterative Algorithms for Bottleneck Context Extraction

This section introduces and important contribution of this thesis. Two algorithms are presented that use a bottleneck approach to extract empowerment-maximizing contexts. The first algorithm is a simple version that uses the lower bound identified in theorem 4.3.1. The advantage is that it is simple to implement and it requires less computational power. The second one performs full empowerment maximization and for this requires two maximization procedures.

Context-Extraction From Empowerment Estimates

The principle of the first algorithm is to maximize the lower bound on empowerment gain while minimizing the amount of information that the context captures. The quantity maximized is then

$$I_{p^*(a_t)}(A_t; C|S_{t+1}) - \lambda I(H; C) \tag{4.20}$$

where λ is a trade-off parameter.

Solutions to this maximization problem can be derived using Lagrange multipliers, leading to (time indices are dropped for the sake of clarity):

$$p(c|h) = \frac{p(c)}{\mathcal{Z}(h)} \exp\left(\frac{1}{\lambda}\sum_{s,a} p(s,a|h)\log p(a|c,s)\right) \tag{4.21}$$

with $\mathcal{Z}(h) = \sum_c p(c) \exp\left(\frac{1}{\lambda}\sum_{s,a} p(s,a|h)\log p(a|c,s)\right)$.

The following iterative algorithm can then be proposed:

Input:

Perception-action loop conditional distribution $p(s|a,h)$ and histories distribution $p(h)$.

Trade-off parameter λ, convergence parameter ϵ, and output set $\mathcal{C}$.

Output:

Mapping of histories to context $p(c|h)$.

Initialisation:

$\forall h : p^{(0)}(c|h) \leftarrow$ random distribution.

$p^*(a) \leftarrow$ capacity-achieving distribution for channel $p(s|a) = \sum_h p(s|a,h)p(h)$.

$i \leftarrow 1$

Algorithm:

```
While TRUE
    // Update conditional distribution $p(a|c,s)$ with current context-mapping $p(c|h)$.
    $\forall a,c,s : p^{(i)}(a|c,s) \leftarrow \frac{\sum_h p^*(a)p(h)p^{(i-1)}(c|h)p(s|a,h)}{\sum_{h,a} p^*(a)p(h)p^{(i-1)}(c|z)p(s|a,h)}$.
    // Compute new context-mapping $p(c|h)$ according to equation 4.21.
    $\forall h,c : E_h^{(i)}(c) \leftarrow p^{(i-1)}(c)\exp(\frac{1}{\lambda}\sum_{s,a} p(s|a,h)p^*(a)\log p^{(i)}(a|c,s))$.
    // Normalization step.
    $\forall h,c : p^{(i)}(c|h) \leftarrow \frac{E_h^{(i)}(c)}{\sum_c E_h^{(i)}(c)}$.
    // Terminates the algorithm if the mapping has not changed significantly.
    If $\max_h D_{JS}[p^{(i)}(C|h)||p^{(i-1)}(C|h)] \leq \epsilon$ break.
```

$i \leftarrow i + 1$.

It is important to note that, similarly to the information bottleneck technique, there is no guarantee that the algorithm will converge towards the global maximum. This is why the mapping is randomly initialised, and several searches should be made with different initialisations to avoid being stuck in local maxima.

Context-Extraction From Full Empowerment

The second algorithm directly maximizes the context empowerment. However, this implies that maximizing over a quantity which is itself a maximization. The quantity maximized is then

$$\mathfrak{E}(A_t \to S_{t+1}|C) - \lambda I(H;C) \tag{4.22}$$

where λ is a trade-off parameter.

Solutions to this maximization problem can be derived using Lagrange multipliers, leading to (time indices are dropped for the sake of clarity):

$$p(c|h) = \frac{p(c)}{\mathcal{Z}(h)} \exp\left(\frac{1}{\lambda}\sum_{a,s} p(a|c)p(s|a,h)\log\frac{p(a|s,c)}{p(a|c)})\right) \tag{4.23}$$

with $\mathcal{Z}(h) = \sum_c p(c)\exp\left(\frac{1}{\lambda}\sum_{a,s} p(a|c)p(s|a,h)\log\frac{p(a|s,c)}{p(a|c)})\right)$.

The resulting algorithm is (time indices are dropped):

Input:

Perception-action loop conditional distribution $p(s|a,h)$ and histories distribution $p(h)$.

Trade-off parameter λ, convergence parameter ϵ, and output set $\mathcal{C}$.

Output:

Mapping of histories to context $p(c|h)$.

Initialisation:

$\forall h : p^{(0)}(c|h) \leftarrow$ random distribution.

$\forall c : p^{(0)}(a|c) \leftarrow$ random distribution.

$i \leftarrow 1$

Algorithm:

```
While TRUE
    // Blahut-Arimoto iteration (improves policy $p(a|c)$).
        // Compute joint distribution $p(s,a,c)$ with current context-mapping $p(c|h)$ and policy $p(a|c)$.
      $\forall s,a,c : p^{(i)}(s,a,c) \leftarrow \sum_h p^{(i-1)}(a|c)p^{(i-1)}(c|h)p(h)p(s|a,h)$.
        // Update policy $p(a|c)$ using standard Blahut-Arimoto.
      $\forall a,c : E_c^{(i)}(a) \leftarrow \exp(\sum_s p^{(i)}(s|a,c)\log p^{(i)}(a|c,s))$.
        // Normalization step.
      $\forall c : Z^{(i)}(c) \leftarrow \sum_a E_c^{(i)}(a)$.
      $\forall a,c : p^{(i)}(a|c) \leftarrow \frac{E_c^{(i)}(a)}{Z^{(i)}(c)}$.
    Bottleneck iteration (improves context-mapping $p(c|h)$).
        // Compute joint distribution $p(s,a,c)$ with current context-mapping $p(c|h)$ and policy $p(a|c)$.
      $\forall s,a,c : p^{(i)}(s,a,c) \leftarrow \sum_h p^{(i)}(a|c)p^{(i-1)}(c|h)p(h)p(s|a,h)$.
        // Update context-mapping $p(c|h)$ using equation 4.23.
      $\forall c,h : E_h^{(i)}(c) \leftarrow \exp(\frac{1}{\lambda}\sum_{a,s} p^{(i)}(a|c)p(s|a,h)\log\frac{p^{(i)}(a|sc)}{p^{(i)}(a|c)})$.
        // Normalization step.
      $\forall c : Z^{(i)}(h) \leftarrow \sum_a E_h^{(i)}(c)$.
      $\forall c,h : p^{(i)}(c|h) \leftarrow \frac{p^{(i)}(c)E_h^{(i)}(c)}{Z^{(i)}(h)}$.
    // Terminate algorithm if the context-mapping has not changed significantly.
    If $\max_h D_{JS}[p^{(i)}(C|h)||p^{(i-1)}(C|h)] \leq \epsilon$ break.
    $i \leftarrow i+1$.
```

As one can see from the algorithm, two maximization steps are needed. First a Blahut-Arimoto step is performed to update the action policy $p(a|c)$. Then, using this new policy, the mapping $p(c|h)$ is updated in a bottleneck-like step. This new mapping is then used to update the action policy at the next iteration, and so on.

The same caveats as for the previous algorithm have to mentioned. In order to get out of local maxima, one should randomly initialise the mapping and the action policy, and perform multiple searches with different initialisations. No proof of convergence is provided, but empirical results show that both algorithms have good convergence properties.

4.3.5 Bottleneck Approach in the Grid Scenario

The two algorithms presented in the previous subsection are now applied to the grid world scenario used in the two previous experiments 4.3.1 and 4.3.2.

Experiment 4.3.3. ***Analysis of the best contexts obtained from iterative algorithms.***
Objective: *Identify the context mappings that are obtained from both iterative algorithms in the context of the grid scenario.*
Main results: *The obtained mappings are able to achieve maximum empowerment while compressing the historical information by half. The mappings are very sharp and indicate areas for which the sensorimotor channel differs.*

Similarly to the previous experiments, sensorimotor data are collected during a first phase of random behaviour. The history horizon is just one step, meaning that the history variable is defined as $H = \{A_{t-1}, S_t\}$. The number of states for the context variable is arbitrarily set to $|\mathcal{C}| = 6$.

From the sensorimotor data, the following statistics are collected: $p(s_{t+1}|a_t, h)$ and $p(h)$. The algorithms calculate a context-mapping $p(c|h)$. Parameters for the two algorithms the following: trade-off parameter $\lambda = 0.01$, stopping criterion $\epsilon = 0.0001$.

Important quantities that are common to this scenario are:

- context-free empowerment: $\mathfrak{E}(A_t \to S_{t+1}) \approx 0.25$ bits,
- the empowerment with history context : $\mathfrak{E}(A_t \to S_{t+1}|H) \approx 0.69$ bits, and
- the entropy of the sensorimotor history : $H(H) \approx 3.86$ bits.

The parameters measured during both experiments are:

- empowerment with bottlenecked context: $\mathfrak{E}(A_t \to S_{t+1}|C)$, and
- the amount of historical information captured by the context: $I(H; C)$.

The resulting contexts and the quantities associated can be found on figure 4.16.

One can see from the results that the two iterative algorithms perform equally well. In both cases a context-mapping has been found the leads to maximum empowerment (0.69

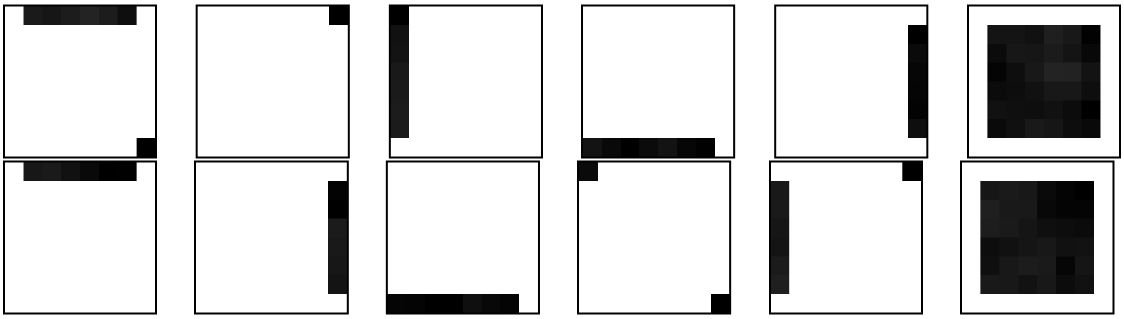

Figure 4.16: Mapping obtained from the best context-mappings computed with the iterative algorithms. Top: algorithm with full empowerment. Bottom: algorithm with empowerment estimates. Each of the six pictures represent the probability of being at a specific location on the grid when the context is in a given state, i.e. $p(r_t|m_t)$. Dark means high probability. Top: empowerment with context: 0.69 bits, information capture: 1.95 bits. Bottom: empowerment with context: 0.69 bits, information capture: 1.99 bits.

bits) while compressing at the same time the amount of historical information (3.86 bits) to almost half of it, namely 1.95 and 1.99 bits.

The resulting contexts are much sharper that the ones that could be obtained by evolving a context-automaton. Indeed what is obtained is an almost hard-mapping from history to context states, where different situations (which are actually distinguished by the last sensor reading) are mapped to distinct context states. Because the 6 states available to the context do not allow to distinguish between all possible situations, some of them, the most uncommon ones, are grouped together (for example some corners are grouped with larger areas).

4.3.6 Limits on Empowerment Gain

Using the iterative algorithms, it is also possible to study the limit of context-empowerment under various constraints on the amount of captured information by using different values for the trade-off parameter λ.

Experiment 4.3.4. ***Empowerment gain limits in the grid scenario.***
Objective: *Identify the relationship between the amount of information captured by the context and the maximum empowerment gain in the grid scenario of experiment 4.3.1*
Main results: *Maximum empowerment gain is almost linearly related to the amount of information captured.*

This experiment uses the same grid scenario as experiment 4.3.1. Sensorimotor data are collected in the same way, but instead of searching for an automaton, the two iterative algorithms presented in the previous section are used to find an optimal mapping for a given λ. By spanning different values for λ, and collecting the corresponding points, the optimal trade-off between information capture and empowerment gain can be plotted. Results can be found on Fig. 4.17.

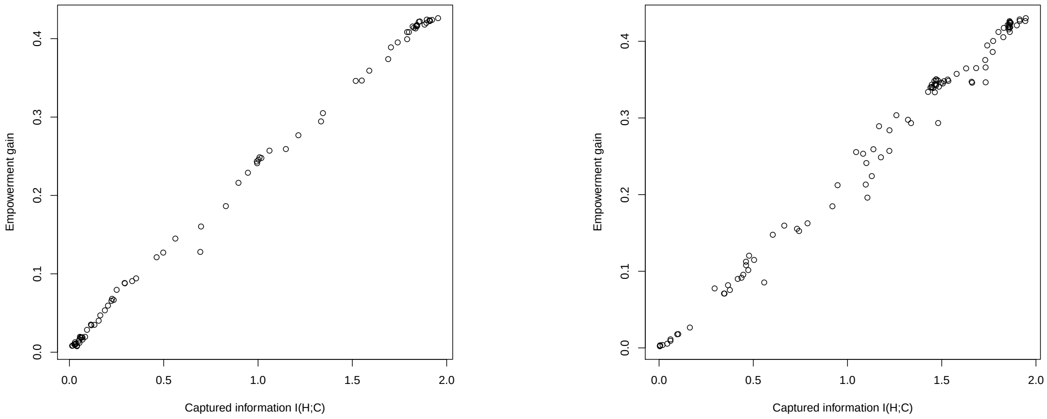

Figure 4.17: Empowerment gain $\mathfrak{E}(A_t \rightarrow S_{t+1}|C) - \mathfrak{E}(A_t \rightarrow S_{t+1})$ versus information captured by the context $I(H;C)$. Left: values obtained using the full empowerment iterative algorithm. λ goes from 0.001 to 0.4. Right: values obtained using the empowerment estimates iterative algorithm. λ goes from 0.001 to 0.1.

The following quantities have also been measured:

- context-free empowerment: $\mathfrak{E}(A_t \rightarrow S_{t+1}) \approx 0.25$ bits,
- the empowerment with history context : $\mathfrak{E}(A_t \rightarrow S_{t+1}|H) \approx 0.69$ bits, and
- the amount of information in the history : $H(H) \approx 3.86$ bits.

In light of the above quantities, one can draw the following conclusions. Firstly, it can be remarked that maximum empowerment (0.69 bits) can almost be achieved by capturing only 2 bits of information. Secondly, roughly half of the information contained in the history is redundant and can be discarded through the bottleneck while still achieving maximum empowerment.

It is interesting to note that there seems to be an almost linear relationship between empowerment gain and the captured information. Similarly, the empowerment gain seems to be upper bounded by the amount of captured information. Such a relationship was already identified in theorem 4.3.2, but it was only applied to the lower bound on empowerment gain. Performing the same kind of experiment in different scenarios may help validate this more general relationship.

4.4 Heuristics for Agent-Centric Estimates of Empowerment

One of the most interesting properties of empowerment is that it can be computed and exploited directly inside the agent, without introducing any external information. However for this to happen, one has to efficiently collect statistics of the perception-action loop in order to estimate various empowerment-related quantities. This section describes two contributions of the thesis that aim at this purpose.

The first one is an exploration strategy based on the adaptive sampling of the perception-action loop. It allows the agent to focus on exploring parts of the environment for which its model is not accurate.

The second contribution is a tool that allows for efficient extraction of causal relationships in time-extended perception-action loops (i.e., causal relationships that may span over long time delays). Its main advantage is that it avoids the combinatorial explosion by 'compressing' the time dimension.

Both contributions are detailed in separate papers: (Capdepuy, Polani and Nehaniv 2008) for the adaptive sampling strategy (see Appendix D), and (Capdepuy, Polani and Nehaniv 2007a) for the second contribution (see Appendix E).

4.4.1 Adaptive Sampling Strategy

A conceptually simple case, the memoryless channel, is used to define the basic principles of our exploration strategy. The perspective taken is to consider an agent that constructs a statistical model of its perception-action loop by collecting samples. This model is represented by a probability distribution $p(s_{t+1}|a_t)$. To construct this model, the agent has to explore the environment by acting on it. At each time-step it picks an action and sends it into the channel, through the environment, and then perceives back a particular sensor value.

By collecting such data it is possible to approximate the real probability distribution of the channel (if it is stationary). However, if one supposes that the channel can sometimes be changed (e.g. external damage, change in the environment) then the agent has to re-evaluate its statistical model to reflect the changes and match the new real model. In the case of the memoryless channel, at each turn, the agent has to pick one of the existing actions to obtain a new sample of the corresponding conditional distribution.

The general idea of the adaptive sampling strategy is to pick actions that are more likely to make the corresponding distribution converge. The convergence is quantified by measuring the delta of entropy δ_H of this distribution for the previous samples. Basically, if the distribution has already converged then $\delta_H = 0$. Therefore this distribution is sampled with a low probability, because it is unlikely that new samples would change it.

However, if recent samples have changed the entropy, this means that further sampling may make it converge, therefore this distribution has a high probability of being sampled.

The δ_H strategy is compared to a random exploration and to an 'Oracle' strategy. The Oracle knows the real distribution and therefore it always picks the best sample. Results in various scenarios show that the δ_H strategy clearly outperforms the random one, and that its performance comes quite close to the optimal Oracle strategy.

It is possible to extend this strategy to channels with memory and agents using a context. In this case each context-state/action pair is associated with a distribution $p(s_{t+1}|a_t, m_t)$ and the sampling strategy has to pick the state/action pair with fastest convergence. Using standard value-iteration algorithms, the agent can 'navigate' between states in order to reach those for which useful sampling can be performed.

This strategy is compared to the random exploration in a grid-world scenario, and results in the construction of a significantly more accurate probabilistic model for the same amount of time exploring the environment.

4.4.2 Time-Extended Perception Action Loops

In the previous chapter, it was shown how sensorimotor channels with extended time horizons can be handled. However, increasing the temporal horizon can rapidly produce combinatorial explosions because of the number of possible sensor states and motor states at each timestep. This section introduces a set of tools that allow to obtain a compressed representation of the time-extended perception-action loop.

To simplify the problem , it is assumed that not all sensor or motor states are taken into account. Instead, we will consider that these are somehow pre-filtered and that only interesting ones are kept (this could be a saliency filter or any other mechanism). One can see this as having a 'do-nothing' action and a 'nothing to see' sensor state which are basically filling up a significant portion of time. With this perspective, the volume of sensorimotor data can be highly reduced, simply because most of the time nothing happens (filtered out), and the resulting information is sparsely distributed over time. We will refer to this information as events (in which actions are included).

Events can be seen as stimuli that the agent encounters in its environment. The agent is then observing a stream of events such as those represented in Fig. 4.18.

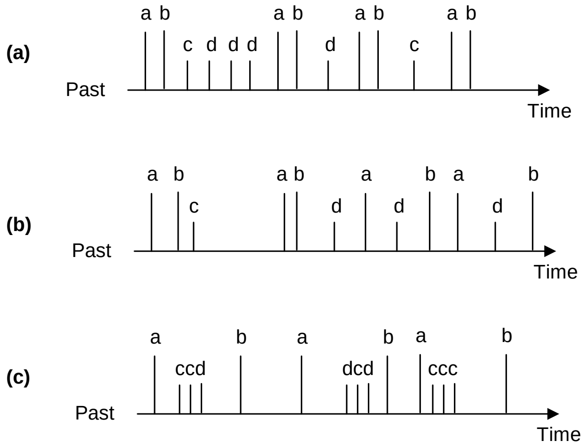

Figure 4.18: Three different streams of events with $\mathcal{E} = \{a, b, c, d\}$. For each of them our aim is to identify the predictive relationship from a to b. Example **(a)** is quite obvious, events a and b follow each other very closely in time, and with a constant delay. In example **(b)** the delay between a and b varies, although it seems that b always follows a. In example **(c)** the delay between a and b is very large, providing room for many events to occur in between, nevertheless as this delay is constant we would like to identify such a relationship.

The goal of this technique is to extract a predictive model of the environment from a stream of symbolic events that can potentially have a predictive relationship with each other. The basis of the model is to collect statistics about pairs of events and their relation in time. The first step is a low-level extraction of information that allows us to generate a compact probabilistic model of the time-delay between pairs of events.

Based on this compressed representation, two measures of anticipation are defined. The first one is based on the regularity of the time-delay between two events measured using the concept of *causal entropy* (Waddell, Dzakpasu, Booth, Riley, Reasor, Poe and

Zochowski 2007), the second one extracts information from the likelihood of consecutive events.

Using various scenarios, it is shown that these two measures have complementary properties. The first one is robust to long delays, but not to the variability of the delay. The second one is robust to variability in the delay, but it requires that the events are relatively close in time. By combining the two approaches, one can extract most of the existing causal or predictive relationships from the stream of events.

The model constructed can be used to perform reward-based behaviour (Capdepuy, Polani and Nehaniv 2007b), but further research is needed in order to transform it into a probabilistic model suitable to expressing empowerment-like quantities.

4.5 Summary

This chapter introduced the concept of *empowerment*, a measure of the amount of perceivable control on agent has onto the environment. This quantity has several interesting properties: it is agent-centric, it can be computed from local information, and it does not rely on a semantic provided by an external designer.

The definitions of empowerment introduced in (Klyubin 2007) have been changed in order to remove reference to information flows, using instead a channel capacity formulation suitable to all possible usages of the sensorimotor channel. This formulation is made possible thanks to the concepts of directed information and feedback capacity presented in the previous chapter.

Empowerment maximization has been motivated and described through three different mechanisms:

- the evolution of sensors and actuators,
- the control of state of the environment, and
- the extraction of a context from past sensorimotor experience.

The last mechanism, context-extraction, has been studied from both a theoretical and an experimental perspective. In the theoretical analysis, I showed that the empowerment gain from using a specific context is bounded from below by the synergy when the channel is used at full capacity. Moreover, it was shown that this bound is itself upper-bounded by the amount of side-information (information about the state of the environment) that the context captures. Together, these two theoretical results allow us to easily estimate the empowerment gain for a given context, and to deduce the amount of information that the context captures about the state of the environment (even though this state is not directly accessible).

The context-extraction methodology presented in (Klyubin 2007) (based on a stochastic search in the space of context-automata) has been applied to a grid-world scenario resulting in structured internal representations of space by the agent. Moreover, I improved this methodology using empowerment estimates based on the lower bound identified in the theoretical analysis. This new approach leads to similar results, but with a significant reduction on the computation time.

I also introduced two iterative algorithms inspired from the information-bottleneck method (Tishby et al. 1999). The idea is to find a mapping of sensorimotor histories into a context variable that maximizes empowerment while minimizing the amount of information captured. The first algorithm maximizes the empowerment estimates derived from the lower bound identified in the theoretical analysis. The second algorithm maximizes the actual empowerment. Experiments in the grid-world scenario leads to very efficient internal representations with an empowerment increase significantly better than the contexts obtained from the context-automaton methodology.

Also, because the trade-off between information capture and empowerment gain can be directly controlled in the iterative algorithms, I could study in more detail the relationship between these two quantities. Results from this analysis in the grid-world scenario indicate a linear relationship between information capture and empowerment gain. However it is not clear at this stage whether this is a general relationship or if it is specific to the scenario studied.

An adaptive sampling technique was proposed in order to improve the collection of statistics from the perception-action loop in changing environments. The main idea is to use the variation of entropy of collected statistics to identify which parts of the model need extra sampling to converge. In a range of scenarios, this technique has been shown to be more efficient than random sampling and almost as efficient as optimal sampling performed by an Oracle (which knowns the actual model of the perception-action loop and performs its sampling accordingly). This technique is described in more details in (Capdepuy et al. 2008) (see Appendix D).

In order to deal with environments in which the causal relationships span over long time-delays, which become problematic in the standard approach because of the combinatorial explosion, two complementary techniques have been proposed. The first one relies on the regularity of the time-delay between specific events. It is robust to long time-delays but not to the variation of the delay. On the other hand, the second technique relies on the likelihood of consecutive events. It is robust to variations in the delay between events, but not to long delays. Used in conjunction, these two techniques allow an agent to identify most causal relationships. However, further research is needed in order to integrate them into the empowerment framework. These techniques are described in more details in (Capdepuy et al. 2007a) (see Appendix E).

Chapter 5

Agents Interactions and Collective Systems

The first part of this thesis studied the informational principles of perception-action loops of single agents. This chapter builds upon results presented in the previous ones to investigate cases where more than one agent are involved.

In the next section, the information-theoretic perspective of multiple agents sharing an environment is depicted. Two aspects of the resulting causal Bayesian network are identified as having an impact on the quantities of interest:

- the ability of the agents to collectively bring the environment into structured states, and
- the simultaneous 'execution' of actions from multiple agents, which we refer to as *interferences*.

Only the first mechanism is studied in this chapter. In order to exclude the second mechanism from the models, agents are forced to behave in an asynchronous way. Because of this, the effect of one agent's actions is necessarily independent from those of the other agents. As interferences are the result of simultaneous actions, asynchronism makes sure none can occur. Interferences will be studied in detail in the next chapter.

Sections 5.2 and 5.3 present two simple scenarios where empowerment leads to either a competitive or a collaborative situation.

Section 5.4 introduces a simple unidimensional space with two agents very similar to the line world example studied in chapter 3. This example is thoroughly studied, using different sensors, embodiments, and empowerment horizons.

In section 5.5, a two dimensional grid world is used with several agents in order to study the impact of different embodiments, and more precisely of different density sensors, on the empowerment of individual agents in a swarm of randomly behaving agents. It is shown that each sensor has specific optimal densities of agents and therefore, that agents striving at maximizing empowerment should aim for such densities.

Section 5.6 takes a similar setup and studies the impact of the spatial distribution of the agents on the empowerment. Results show that, depending on the embodiment of the agents, specific global structures bring maximum empowerment.

The space of possible behaviours is searched in Sec. 5.7 for those that bring maximum empowerment to the agents. It is shown that the obtained behaviours induces global organizations that increase empowerment by creating regularities in the sensorimotor channel.

In section 5.8, the behaviour of agents is constructed so that they locally and selfishly maximize empowerment. By doing so, complex emergent organizations are generated. In return, these organizations induce new empowerment-maximizing behaviours which, when performed by the agents, create new patterns of organization. Computer simulations show that this ‘coevolution’ between organization and behaviour leads to the emergence of a wide range of global structures, but generally fails to bring high empowerment to the agents.

5.1 Information-Theoretic Picture of Interactions

When two or more agents share a common environment, their perception-action loops become intertwined (see Fig. 5.1). As one can see, the information that flows from each agent’s action A and B to their subsequent sensor variables S and T ‘collides’ in the state of the environment R. It is this collision which is at the core of the interaction between agents and that makes their information-theoretic properties interesting.

Because the collision occurs before the next sensor readings are performed, looking at the motor channel is sufficient to investigate the main properties of the interaction. Therefore the remaining of this section will focus on the motor channels of both agents. Put another way, we assume that both agents are able to directly sense R (sensor variables S and T are removed and replaced by R).

Two levels of interaction have to be distinguished. The first one appears at each time-step, and even in channels which are memoryless. Consider for example the very first step

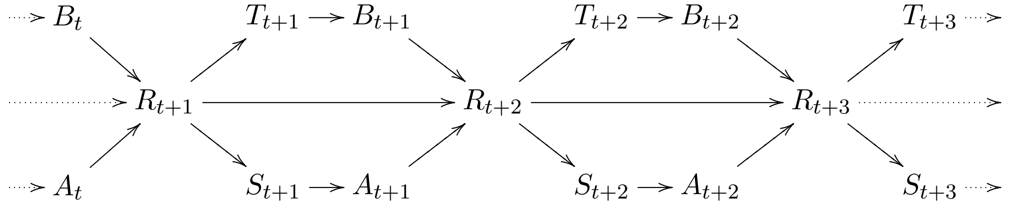

Figure 5.1: Typical perception-action loop of two agents sharing the same environment unrolled over three time-steps. The agents have no memory beyond their immediate sensor readings. R stands for the state of the environment. The first agent's sensor and action variables are respectively S and A, those of the second agent are denoted by T and B.

appearing in Fig. 5.1. One can see that the output of the motor channel R_{t+1} depends on both actions A_t and B_t. From the point of view of agent A this means that the outcome of the selected action depends on the policy performed by B. Therefore, the policy executed by each agent may have an impact on the information-theoretic quantities of interest for the other agent. Moreover, it is obvious that knowledge about the action of the other agent has an impact on these quantities too. This level of interaction will be referred to as *interference*.

The second level of interaction appears when the environment has memory and the agents can change the its state by acting onto it. Indeed, the distribution over the states of the environment is an important parameter of the information-theoretic quantities, whether or not the agents have any knowledge about this state. Therefore, depending on the structure of the environment, agents striving at maximizing their empowerment will have preferred distributions. We refer to this interaction as *shared control*.

Although shared control appears only when controllable memory is introduced in the environment, interferences may also play a role. Not only can the actions of the agents interfere in a given state of the environment, but they may also interfere for controlling the state. The two levels of interaction therefore necessarily arise when the environment has memory.

Also, it is important to understand that, even if the environment itself is memoryless, if one agent has memory (and chooses its actions according to it) then the environment, as perceived by the other agents, also has memory. Therefore the two levels of interaction may arise in memoryless environments with memoryful agents.

In the following sections, only the shared control mechanism will be studied. For this purpose, interferences have to be removed. In order to do this, it will be assumed that the agents are behaving in an asynchronous way. By doing so, no two agents can perform actions during the same time-step. This allows us to decouple the two mechanisms. Interferences are the topic of the next chapter.

5.2 The Bathroom Problem

Consider the following situation. Two agents A and B are in front of a bathroom. Only one agent at a time can be in the bathroom. The environment has three possible states (which both agents perceive): the bathroom is either empty, occupied by A, or occupied by B. The agents have only two possible actions, either stay where they are, or change their location. If the bathroom is already occupied, then the agent outside cannot change its location.

What is the empowerment in this environment? Let us look at the different situations:

- The bathroom is empty. Agent A can either stay outside or move inside, therefore it has 1 bit of empowerment. The same applies to agent B, so it also has 1 bit of empowerment.
- The bathroom is occupied by A. In this case agent A can either stay inside or move outside, hence it has 1 bit of empowerment. However, agent B is not as fortunate. It has no other option than staying outside, and wait for agent A to come out. Put another way, the outcome of both its actions is the same, therefore it has no empowerment.
- The bathroom is occupied by B. This situation is exactly the opposite of the previous one, therefore agent B has 1 bit of empowerment, and agent A has none.

Now, if we take the empowerment maximization perspective, what should the agents do? It is quite obvious that if one wants to maximize the empowerment of both agents, the best strategy is to have them both outside the bathroom. Indeed, in this situation, both agents have the ability to get into the bathroom if they want to.

Things are different if we take the perspective of one agent, say agent A. In this case, the only bad situation is when agent B occupies the bathroom. In the two other situations, it has 1 bit of empowerment. But, if one takes a closer look at this example one of them is preferable. In the case where the bathroom is empty, the empowerment of agent A is guaranteed as long as B also stays outside. If B moves inside the bathroom, A's empowerment drops to zero. This situation is not really optimal. However, if agent A is inside the bathroom, then it keeps its 1 bit of empowerment no matter what agent B decides to do. Indeed, in this case, only A has control on the state of the environment. Therefore such a situation should be preferred.

What this example shows is that, even if there are situations which are optimal for both agents, maximizing empowerment from a selfish perspective leads to situations which are suboptimal. Therefore, selfish empowerment maximization in such an environment would lead to a competitive behaviour between the agents, because the agents have different preferred states of the environment.

5.3 The Well Problem

Now, consider the following situation. Two agents A and B are in front of a well. Only one agent at a time can be in the well. As in the previous example, the environment has three possible states (which both agents perceive): the well is either empty, occupied by A, or occupied by B. The key difference compared to the previous example is that once inside the well, an agent cannot get out. The only way for it to get outside is for the other agent to 'help him out'. Again, the agents have two possible actions, but their semantic depends on the state of the environment. Here are the possible situations:

- The well is empty. Agent A can either stay outside or move inside, therefore it has 1 bit of empowerment. The same applies to agent B, so it also has 1 bit of empowerment.
- Agent A is in the well. In this case agent A cannot do anything, and hence has zero empowerment. Agent B can choose between leaving A in the well, or getting him out. Therefore it has 1 bit of empowerment.
- Agent B is in the well. Agent A has 1 bit of empowerment, agent B has no empowerment.

It is clear that this example is the opposite of the bathroom problem. Whether one considers global empowerment or selfish empowerment, the best situation is for both agents to stay outside the well. Although, from the perspective of agent A for instance, if B is already in the well there is no incentive for A to get it out, because that would not change its empowerment.

This example is typically a collaborative environment. This is due to the fact that the preferred states of both agents are actually the same. Therefore, empowerment maximization should lead to both agents controlling the environment state to reach this situation.

5.4 Two Agents in a Line World

First let us consider a simple case with two agents on a discrete line (see Fig. 5.2). The state of the environment is completely defined by the absolute positions of agents A and B, represented by random variables X and Y respectively.

Agent A has a size of 1, i.e. it occupies one site on the line. It can choose actions in the set {left,right,stay}. Moving out of the line on one side brings the agent to the other side (periodic boundaries).

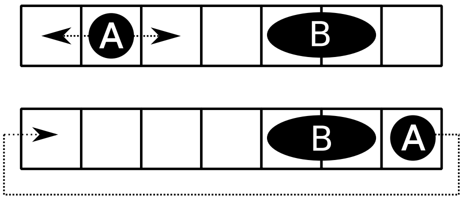

Figure 5.2: Two agents in the size 6 line world with periodic boundaries. Agent A has a size of 1, agent B has a size of 2. **Top:** Agent A can move to the left or to the right, or stay where it is. **Bottom:** Agent A collides with agent B, therefore it cannot go left. Because of the periodic boundaries, going to the right lands him on the left side of the line world.

Experiments in this section are performed in a size-6 line world . They study the effect of the following parameters on agent A's empowerment (see Fig. 5.3):

- The size of agent B: agent B can occupy zero (no collision), one or more tiles (consecutive). The larger agent B is the more constrained agent A's moves are.
- The sensor S used by agent A.
- The context M used by agent A.
- The number of empowerment steps considered, i.e. the length of action sequences (or source horizon).

Two main classes of sensors are studied:

- Absolute position sensors: these capture information about the absolute position of the agents:
  - X, Y: senses the full state of the world.
  - X: only agent A's absolute position is perceived.

$$\begin{tikzcd}
& M_t \arrow[r] & A_t \arrow[rr, dotted] \arrow[dr] & & S_{t+1} \\
X_t, Y_t \arrow[ur] \arrow[rrr] & & & X_{t+1}, Y_{t+1} \arrow[ur] &
\end{tikzcd}$$

Figure 5.3: General causal Bayesian graph for the line world experiments. Solid arrows: causal links. Dashed arrow: sensorimotor channel. The state of the channel X_t, Y_t (i.e. the absolute positions of both agents) is mapped to a context M_t. Action A_t is picked according to the state of the context (if there is a context). As a result the position of agent A changes, which is perceived through the sensor.

  - Y: only agent B's absolute position is perceived.
  - Nxx (with xx a number): noisy sensors that capture the full state of the world X, Y and add noise to it. The number specifies the amount of noise added to the sensor, for example $N30$ means that the sensor will return the actual value 70% of the time and a uniformly distributed random value 30% of the time (uniformly distributed over all allowed states of the world).

- Density sensors: these capture information about the presence of the other agent in the neighbourhood of the sensing agent:
  - D: directional sensor. The agent senses the presence of the other agent in one of the neighbouring sites. It is directional, meaning that agent on the left and on the right are distinguished.
  - T: totalistic sensor. The agent senses the total amount of agents in its neighbourhood, therefore it only distinguishes between not neighbouring, neighbouring on one side, neighbouring on both sides.

The different experiments performed investigate the following aspects:

- Empowerment in the motor channel that gives an absolute limit on that of any sensorimotor channel.
- Context-free and context-dependent empowerment for sensorimotor channels with absolute position sensors.
- Context-free and context-dependent empowerment for sensorimotor channels with density sensors.

5.4.1 Methodology

In all the following experiments on the line world, the initial state of the world is an equidistribution over all the valid states (i.e. states where the agents are not overlapping).

This means that, before agent A performs any action, the position of both agents is as random as possible.

Agent B is not allowed to perform any action. Thus, during agent A behaviour, the position of agent B is unchanged. By doing so, there can be no interference between the agents' actions (this aspect will be studied in chapter 6).

For each possible initial state, all the actions of agent A are executed and the resulting state is collected to generate the statistics of the motor channel. The mappings corresponding to the different sensors are then applied to these statistics at the sensor and/or context level in order to generate the statistics of the different sensorimotor channels. From these statistics, context-free and context-dependent empowerment are computed using standard Blahut-Arimoto algorithm.
For multiple N-steps empowerment, every N-steps sequence of action is executed and the last world state is collected (therefore this is open-loop non-interleaved empowerment with source horizon N).

5.4.2 General Bounds

The actuator of agent A is the same for all experiments, i.e. it has three different actions. This imposes an upper bound for 1-step empowerment of $\log_2(3) \simeq 1.58$ bits and of $N \times 1.58$ bits for N-step empowerment.

Because agent B does not move, only the location of agent A can be changed. Therefore for each initial state, only $6 - size_B$ locations are accessible. This gives another upper bound to empowerment in the motor channel of $\log_2(6 - size_B) \leq 2.58$ bits, whatever the number of steps considered. These bounds are depicted on Fig. 5.4.

Because of inequality 3.6, these bounds also apply to any sensorimotor channel. Moreover these upper bounds are valid for the maximal context case, and therefore they are also valid for any possible context.

5.4.3 Motor Channel

Because the motor channel gives upper bounds to all the possible sensorimotor channels, studying its properties gives us a reference that other sensorimotor channels can be compared to.

The causal Bayesian graph of this experiment is represented on Fig. 5.5. The quantity investigated is the N-step context-dependent empowerment $\mathfrak{E}(A_t^N \to X_{t+N}, Y_{t+N}|M_t)$.

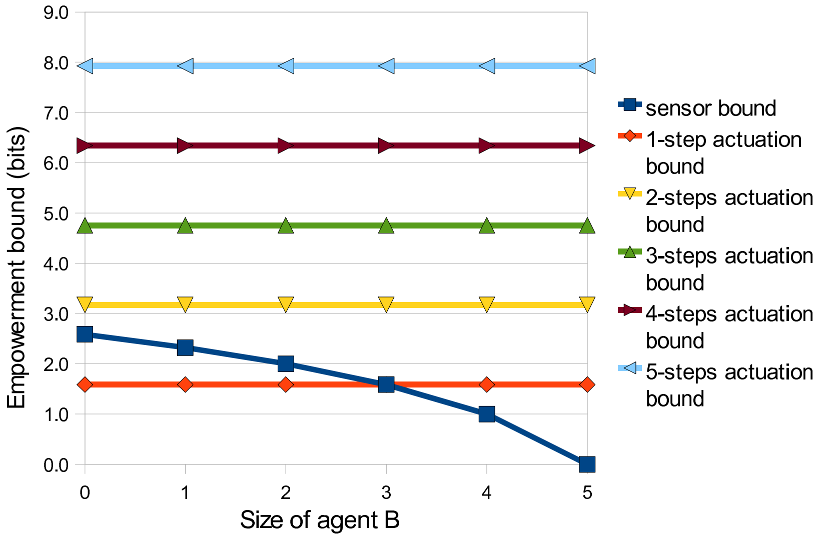

Figure 5.4: Upper bounds on empowerment for different number of steps and sizes of agent B.

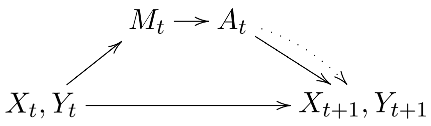

Figure 5.5: Causal Bayesian graph of the 1-step motor channel experiments. Solid arrows: causal links. Dashed arrow: motor channel.

The context M_t is taken among the following variables:

- X, Y: full state of the world. This is a maximal context, i.e. there does not exist any context that gives a higher empowerment.
- X: absolute position of agent A.
- Y: absolute position of agent B.
- Nxx: X, Y sensor with $xx\%$ noise.
- $\emptyset$: no context.

Results of this experiment for 1-step empowerment are shown on Fig. 5.6.

Let us first look at the maximum case, i.e. for context variable X_t, Y_t. As it captures the full state of the world, this variable is a maximal context, and therefore the corresponding empowerment is the maximum achievable. In the case where agent B has size 0,

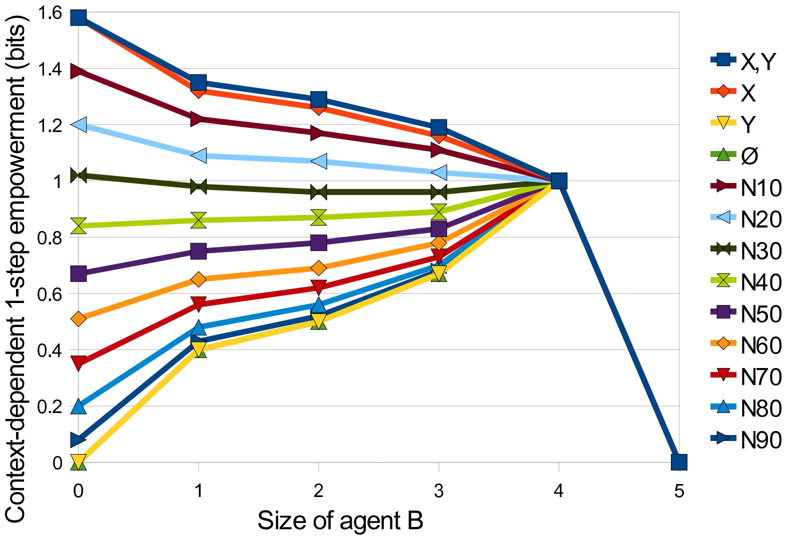

Figure 5.6: Context-dependent 1-step empowerment of the motor channel for different contexts in the line world. $\emptyset$ and Y values overlap.

i.e. it does not physically exist, the empowerment for each initial state is maximum, because the agent can reach three different destinations at each time-step. If a second agent is introduced then empowerment decreases, due to the fact that when agent B is present, there exist initial states for which the empowerment is not maximum. More precisely in such cases only two destinations can be reached: staying next to agent B or moving away from it. The more space agent B occupies, the more states have low-empowerment. In the worst case, size 5, there is only one site left for agent A, and therefore it cannot move to any other location and its empowerment is 0 regardless of the context (this can be seen from the upper bounds on Fig. 5.4). In such spatial environments, when the full-state of the environment is known, empowerment of the motor channel is maximized when interaction constraints are minimized.

The situation radically changes when the state of the environment is unknown or just partially known to agent A. This is illustrated in the minimum case, i.e. when there is no context (line $\emptyset$ on Fig. 5.6). One can see that the behaviour is quite the opposite of the maximal context case. With no context and no interaction, the empowerment is 0. However increasing the size of agent B increases the context-free empowerment. Until the point of size 5 where empowerment is 0 for the same reasons given in the previous paragraph. This phenomenon can be explained by the fact that the presence of agent B structures the environment by colliding with agent A, allowing some correlations between actions and future world states to be visible even when the initial state is unknown.

Unsurprisingly, the noisy contexts have empowerment values between maximal context and context-free cases. However it is interesting to note that the context Y_t, i.e. the position of agent B, has exactly the same values as the context-free case. What this means is that in the motor channel, knowing the location of the other agent does not bring any empowerment on its own. However if one looks at the X_t context, i.e. the position of agent A, empowerment is very close to that of the maximal-context case but not quite. The small difference comes from the use of agent B's position in the maximal context case. This can be interpreted in the following way. Knowledge about the position of agent B alone does not improve empowerment, however combining it with agent A's brings more empowerment that the location of A alone. Therefore these two variables have a synergetic effect: taken together they bring more than their independent contributions.

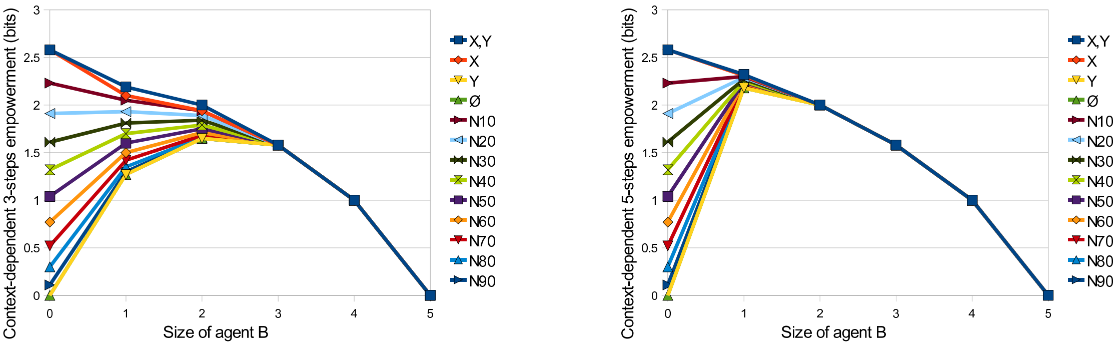

Figure 5.7: Context-dependent 3-steps (left) and 5-steps (right) empowerment of the motor channel for different contexts in the line world.

If more empowerment steps are considered, as illustrated in Fig. 5.7, empowerment increases for less constrained situations. To explain this one has to understand that when more steps are considered, more locations become accessible. In the 5-steps case all the sites of the size-6 line-world are reachable. Indeed, when considering the 5-steps empowerment in the motor channel with maximal context, values reached are that of the upper bound described before, i.e. $\log_2(6 - size_B)$.

However, if we look at the context-free case, one can see that when agent B is absent the empowerment is still 0, whereas it immediately gets close to the maximum value when agent B is introduced with size 1 (maximum values are actually reached when more steps are considered). How can this be explained? When no context is used in the size 0 case, i.e. agent B is absent, the channel looks totally random. Indeed whatever actions are performed, the end-state of the world will directly depend on the initial state. And because this initial state is uniformly distributed and unknown to the agent, the end state will also be uniformly distributed leading to 0 empowerment.

But the situation changes dramatically if agent B is introduced. What happens is that sequences of actions can be correlated with the end state even when the initial state is unknown. Consider for example the two following sequences: move left for N steps, and move right for N steps. The first sequence will lead to a subset of the world states which is all the states where agent A is just on the right of agent B. Similarly the second sequence will lead to the subset of all the states where agent A is neighbouring agent B on its left. Therefore these two sequences deterministically lead to two distinct subsets of the world states, bringing at least 1 bit of empowerment.

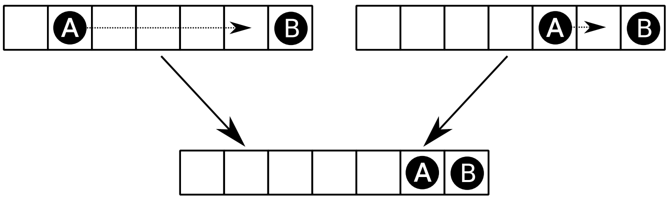

Figure 5.8: Gaining context-free empowerment through reduction of the environment state entropy. **Top:** two initial positions for agent A. By performing 5 steps to the right, whatever the initial position, the agent is sure to end up touching agent B on its left. **Bottom:** resulting position with high context-free empowerment. The entropy over the full state of the environment is reduced to those with relative position -1 (calculated using $X - Y$). From this state, all relative positions are deterministically reachable.

To understand how more empowerment can be reached let us consider the situation on Fig. 5.8. The key aspect here is a change of perspective. Instead of looking at the state of the environment as the combination of both agents' positions X, Y, one can refer to the relative position of the agents $X - Y$.

Originally, because we have defined the distribution over the states of the environment as being uniform (maximum entropy), the distribution over relative positions $X - Y$ is also uniform. However, by acting over a few steps, the agent is able to set the relative position to a specific value. Hence, the entropy of the environment's state is reduced to 0 and set to a known value. Thanks to this, the agent is then able to deterministically set the environment in any relative position it wants, i.e. it is in a highly empowered situation.

Therefore, the presence of agent B creates structure in the environment that can be exploited to increase the context-free empowerment of agent A. To do so, agent A has to reduce the entropy of the state of the environment through its behaviour.

5.4.4 Sensorimotor Channels in the Context-Free Case

Now we look at the context-free case but for different sensorimotor channels (corresponding Bayesian graph on Fig. 5.9. The quantity investigated is then $\mathfrak{E}(A_t^N \to S_{t+N})$ where S is taken among the following variables:

- X, Y: full state of the world.
- X: absolute position of agent A.
- Y: absolute position of agent B.
- Nxx: X, Y sensor with $xx\%$ noise.

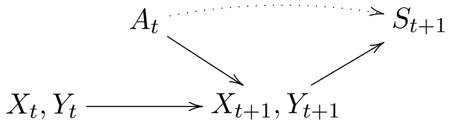

Figure 5.9: Causal Bayesian graph of the 1-step context-free sensorimotor channel experiments. Solid arrows: causal links. Dashed arrow: sensorimotor channel.

Results of this experiment are shown on Fig. 5.10.

Let us first have a look at the two extreme sizes. When agent B has size 0, and thus does not interact with agent A, the context-free empowerment is 0 and does not depend on whether the sensor captures full information or not. This can be explained in the following way. Agent A indeed has control on which location it ends up to, but this location also depends on where it started. As we assumed that the initial distribution is uniform, and because we are looking at context-free empowerment, there is no way for the agent to generate any correlation between its actions and the end location. In the other extreme case, size 5, the upper bound on the motor channel applies leading to 0 empowerment.

These two cases show that, in such an environment, context-free empowerment is positive only between the two extreme cases of maximum freedom (no interaction) and maximum constraint (no possible movement).

However, this effect is also a consequence of the uniform distribution over the agents'

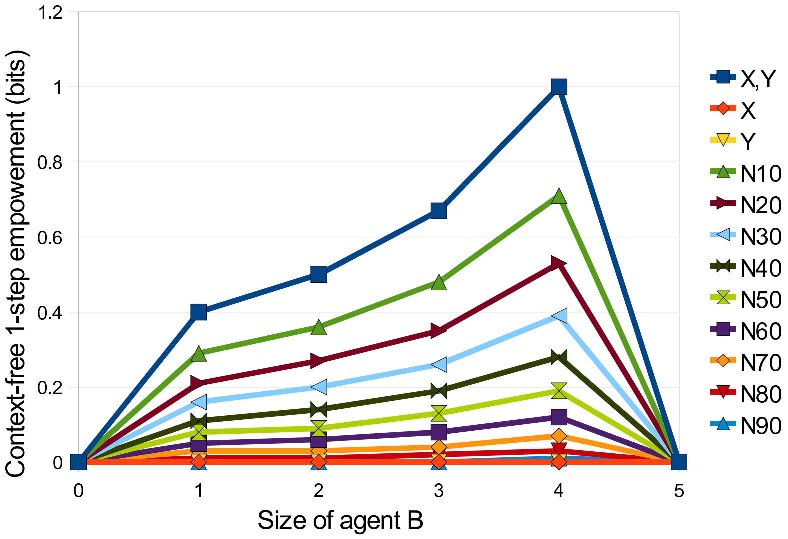

Figure 5.10: Context-free 1-step empowerment of agent A in the line world for different sensors and different sizes of agent B. X and Y values overlap.

positions. If the distribution was not uniform, context-free empowerment would be positive even when there is no agent B to collide with.

Let us now consider the intermediate sizes. First in the case of sensor X, Y (full state sensor) one can see that the empowerment increases with the size of agent B. To understand this phenomenon it is necessary to get into the details of the system. In the case of size 4 agent A has two adjacent tiles it can move to. If we consider the set of all valid world states, this set can be divided into 2 subsets: the subset $\mathcal{L}$ of states where agent B is immediately to the left of A, and the subset $\mathcal{R}$ where B is immediately to the right of A. In this situation, whatever the initial locations of both agents are, moving left leads to a state in the subset $\mathcal{L}$ while moving right ends up in the subset $\mathcal{R}$. The action *stay* can lead to either subset depending on the initial location, and therefore does not contribute to the context-free empowerment. Thus being able to choose deterministically between these two subsets brings 1 bit of empowerment to agent A. For lower sizes of B, there are more and more non-neighbouring states which make the outcome more dependent on the initial location, unknown in the context-free case, and therefore lowers the empowerment.

Unsurprisingly, adding noise to the sensor decreases the empowerment while keeping the same overall profile. However an interesting phenomenon occurs. The two lines corresponding to sensors X or Y on Fig. 5.10 have 0 empowerment whatever the size of agent B. This does not seem surprising at first sight for sensor Y because agent A has no control on this variable. However for sensor X, i.e. the position of agent A, it is rather peculiar. Moreover if we consider even very noisy sensor such as $N90$ the empowerment is greater than that of the X sensor without noise. How can this be explained?

To illustrate the phenomenon at work here let us look at a simple example. Imagine a XOR gate with two inputs A and B and an output Y. Consider that A is the action of agent A and that B is the location of agent B. If we take the same assumption as the experiment, i.e. that the original location of B is uniformly distributed, then the channel capacity from A to X is 0. This is due to the fact that whatever value is sent by A the outcome X will depend on what is the value of B. And because this value is $\{0, 1\}$ with probabilities $\{\frac{1}{2}, \frac{1}{2}\}$ the outcome will also be equidistributed. Therefore $C(A \rightarrow X) = 0$. However if we look at the quantity $C(A \rightarrow X|B)$ then we find that this capacity is 1 bit. Indeed A is able to inject one bit in the outcome, but to perceive the correlation one needs to also see the other input B. The XOR channel is synergetic.

In the line world the situation is similar. The agent A is able to actually inject information into its next location; however, for the correlation between his action and the next location to be visible, it needs to also perceive the location of agent B. There is a synergetic relationship between the action A_t, the next location X_{t+1} of agent A, and agent B's location Y_t (or indifferently Y_{t+1} because agent B does not move). Therefore sensing both locations X and Y even with a lot of noise is better than sensing only one of the two variables without noise.

From this experiment, two main principles can be extracted about context-free empowerment of two agents in a spatial environment such as the line world:

- context-free empowerment is maximized by trading-off freedom of agents and constraints due to interaction, with 0 empowerment at the extremes, and
- because of the synergetic nature of the channel, sensors have to capture information correlated with the location of both agents in order to have any empowerment.

These principles are also valid for N-steps empowerment, but the maximum empowerment appears at different points in the trade-off, more precisely the more steps are considered the closer it is from the constraint-free situation. Results for 3-steps and 5-steps empowerment are shown on Fig. 5.11.

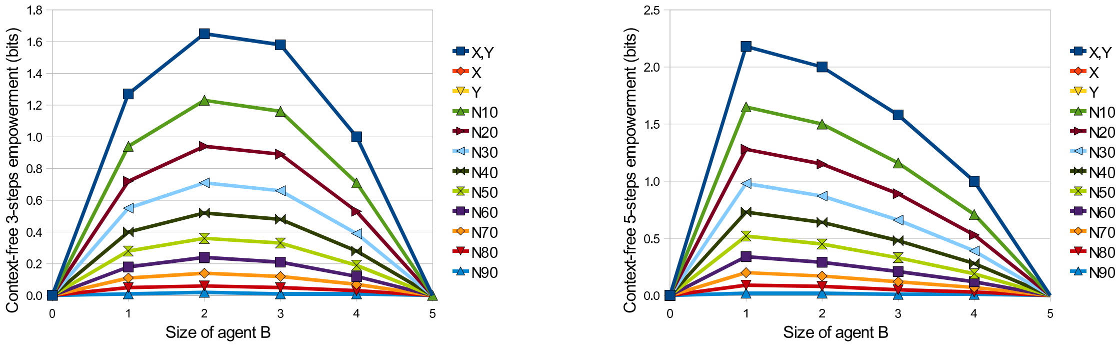

Figure 5.11: Context-free 3-steps (left) and 5-steps (right) empowerment for different sensorimotor channels of agent A in the line world.

This is explained by the fact that less sites are accessible to agent A when B occupies many of them, therefore less world states are accessible, limiting the capacity of the motor channel, and hence that of the sensorimotor channel. In the size 6 line world, after 5 actions all the accessible states can be reached.

5.4.5 Sensorimotor Channels in the Context-Dependent Case

In a previous experiment we have seen how knowledge about the state of the motor channel effects its empowerment. We now study the impact of context in the different sensorimotor channels described in the last experiment. The main idea is that we consider sensorimotor channels where the context used is the previous sensor reading (see Fig. 5.12). It means that the agent has some information about A's and/or B's locations prior to performing its action. As was shown before, adding a context can only increase empowerment, however it is not clear how significant this increase can be and how the nature of the sensor used may impact it.

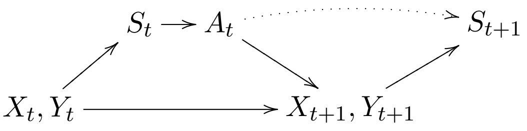

Figure 5.12: Causal Bayesian graph of the 1-step context-dependent sensorimotor channel experiments. Solid arrows: causal links. Dashed arrow: sensorimotor channel.

The quantity investigated is $\mathfrak{E}(A_t^N \to S_{t+N}|S_t)$ where S is taken among the following variables:

- X, Y: full state of the world.
- X: absolute position of agent A.
- Y: absolute position of agent B.
- Nxx: X, Y sensor with $xx\%$ noise.

Results of this experiment for 1-step empowerment are shown on Fig. 5.13.

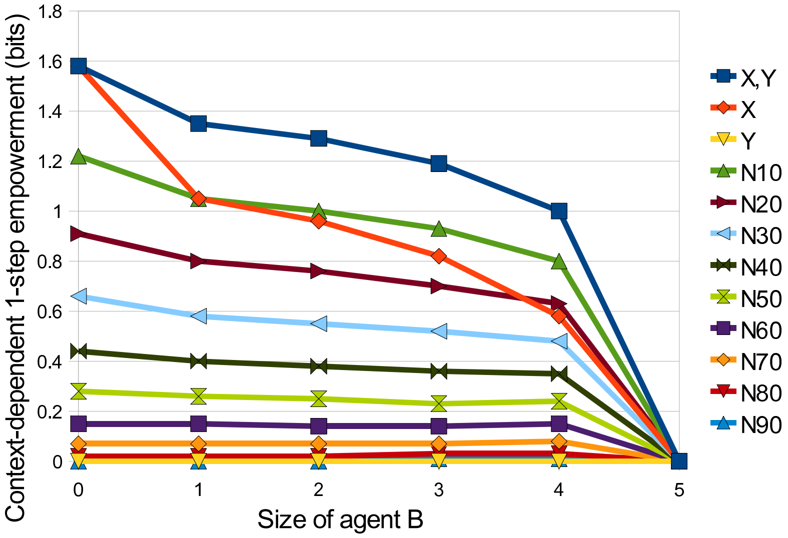

Figure 5.13: Context-dependent 1-step empowerment of agent A in the line world for different sensors and different sizes of agent B.

Results for size 0 interestingly differ from the context-free case. One can see that instead of having 0 empowerment, in this case we get maximum empowerment for sensor X, decreasing with noise, and 0 empowerment for sensor Y. What this means is that because agent A does not interact with B, knowledge about the location of B does not increase empowerment. However for intermediate sizes the synergetic nature of the channel is still visible. One can see for example that for size 3 it is better to have noisy sensor capturing information about both A and B than to have perfect perception of A's location.

It is important to notice at this point that maximum empowerment values are reached for different sizes of agent B depending on which sensor is used. This phenomenon differs from the context-free case where the empowerment profile is that same for each sensor, i.e. maximum empowerment are reached for the same sizes of agent B. In the context-dependent case each sensor has specific maximum points. For example the sensor $N60$

reaches maximum empowerment at both sizes 0, 1 and 4, while sensor X has its maximum only at size 0 and only at size 4 for sensor $N70$.

The empowerment profiles change slightly when the number of steps is increased, as shown on Fig. 5.14. For example the sensor $N20$ reaches maximum empowerment for size 0 in the 1-step case whereas this maximum is reached for size 1 when considering 5-steps.

Another important point to mention is that increasing the number of steps makes context-free and context-dependent empowerment profiles similar (compare Fig. 5.11 and 5.14). The reason for this is that agent A can use its extra initial steps to reduce the entropy of the environment's state enough so that it compensates for the lack of context (the same process that is described in Fig.5.8).

Interestingly, this is not true when agent B has size 0 and therefore does not interact with agent A. Because the line world has periodic boundaries and agent A cannot be 'blocked' by the other agent, there is no strategy that can reduce the entropy of the state of the environment.

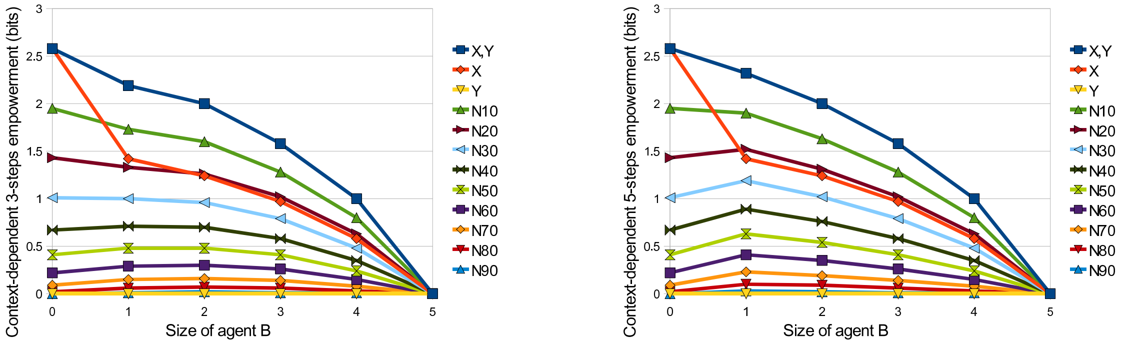

Figure 5.14: Context-dependent 3-steps (left) and 5-steps (right) empowerment for different sensorimotor channels of agent A in the line world.

5.4.6 Sensorimotor Channels with Density Sensors

The previous experiments only investigated absolute positions sensors. These can be seen as some kind of GPS-like sensors. However such sensors are not very biologically plausible. Living beings are generally able to perceive only local properties of their environments. One such property, in multi-agent systems, is the local density of agents.

Using the same methodology as the previous experiments, we investigate the empowerment properties of density-based sensors in the line world. We distinguish two different density sensors: the directional D and the totalistic T density sensors. The directional density sensor measures the density (or in the 2 agents case the presence) on the left and on the right of the agent. On the contrary the totalistic sensor only perceives if agents are present on the left and on the right but does not distinguish between the two directions. The totalistic sensor is then a reduced version of the directional one.

An important difference of these sensors compared to the absolute position ones is that they capture information about the environment-mediated relationship between the two agents. From our two agents perspective this means that when agent B has size 0, and therefore does not 'physically' exist, there is nothing for agent A to perceive.

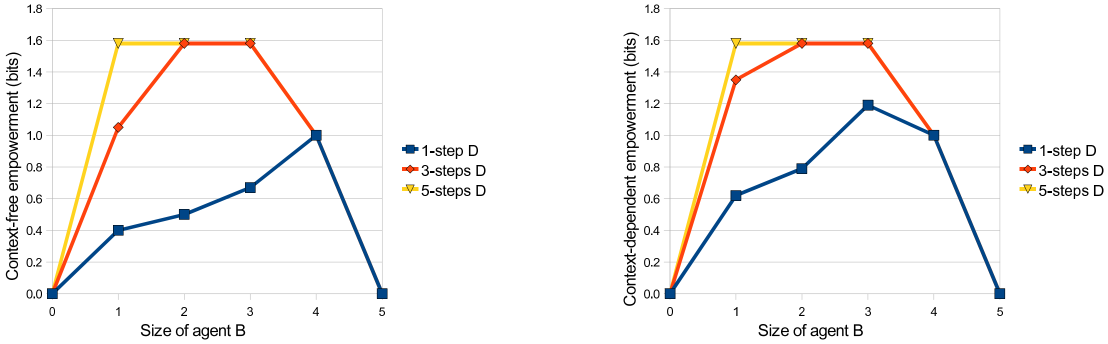

Figure 5.15: Context-free (left) and context-dependent (right) empowerment with directional density sensor.

Experimental results for the directional density sensor are shown on Fig. 5.15. Both figures show values for 1, 3 and 5-steps empowerment.

The context-free case does not look very different from the absolute sensors. Indeed the same trade-off between freedom and constraint is visible, and empowerment in the 1-step case is higher for more constrained situations. Also increasing the number of steps moves the maximum empowerment points towards the less constrained situations, actually reaching the maximum empowerment in the 4-steps case.

However empowerment profiles are very different in the context-dependent case. Indeed, as described above, the fact that the density sensor captures relational properties between the agents makes the presence of agent B necessary to have any empowerment. Therefore situations with high freedom, or low interactions between the agents, lead to low empowerment. This can be counterbalanced by considering more steps.

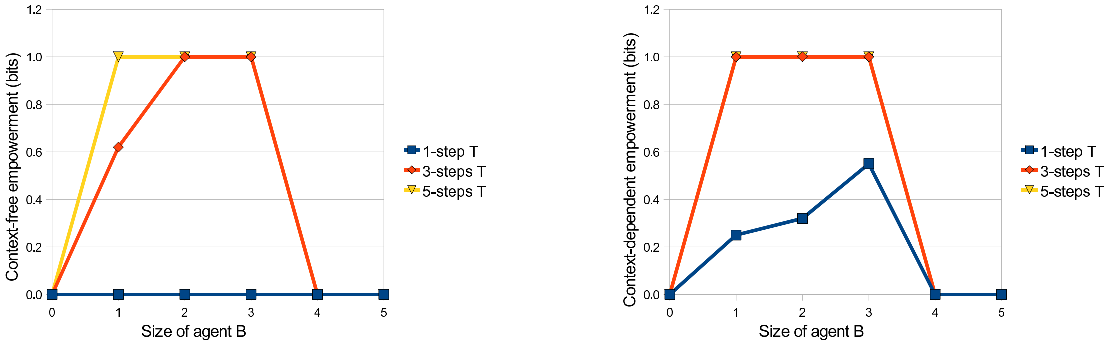

Figure 5.16: Context-free (left) and context-dependent (right) empowerment with totalistic density sensor.

The totalistic density sensor empowerment profiles are shown on Fig. 5.16.

As described above, the totalistic sensor is a reduced directional sensor, in the sense that the states ‘presence-left’ and ‘presence-right’ of the directional sensor are merged into one state ‘presence’. Therefore the empowerment profiles are quite similar between the two agents; however there are two main differences. The first difference is that in the size 4 case empowerment is 0. This is because when only two sites are accessible, the other agent is always present (on the left or on the right depending on the state), therefore the sensor cannot distinguish between the two states.

5.4.7 Synthesis of the Results

The set of experiments conducted in the line world allows to sketch some general principles about empowerment in spatial multi-agent systems:

- Context-free empowerment is maximized by trading-off freedom of agents and constraints due to interaction. Indeed, some freedom, in the sense that the agent is able to set the environment in different states (and perceive them), is needed to have any empowerment. On the other hand, too much freedom, for example when agent A does not collide with B, can lead to an environment too unstructured for the agent to have empowerment. However, this depends on many parameters, such as the initial distribution over the states of the environment, the number of steps considered, and the ability for the agent to reduce the entropy of the environment state.

- In the context-free case, empowerment profiles (and maximum empowerment situation) weakly depend on the sensor being used. Indeed, whether the sensor has noise, captures information about A, B, or both, profiles are very similar. Therefore, agents maximizing context-free empowerment should behave similarly whatever the sensor. On the other hand, increasing the number of steps has a strong impact on the empowerment profile, generally increasing the empowerment of situations with high freedom. Typically, empowerment-maximizing agents would be expected to move towards positions further away from other agents when the number of steps is increased.
- On the contrary, context-dependent empowerment profiles are specific to the sensor and the context being used but depend in a weaker way on the number of steps used. Contexts that are not very informative have properties rather close to context-free empowerment, i.e. better empowerment for moderately constrained situations moving towards free situations when the number of steps is increased. On the other hand, for contexts that are very informative, i.e. that capture most of the information about the state of the environment, most empowered situations occur when the agent is minimally constrained, and the number of steps has a very weak impact on the profile. This can be explained by the fact that when the context is informative, entropy reduction of the environment state are not useful to the agent because it already knows this state thanks to its context.

If we try to extrapolate these results to multiple agents having local sensors and interacting in a grid world, there should be an optimal density of agents for which context-free empowerment is maximized and that density would not depend on the specific sensor being used. Also, increasing the number of steps should lower this optimal density, because the agents would have better empowerment in less constrained situations, i.e. having less agents around.

On the other hand, in the case of context-dependent empowerment where the context is the sensor of the agent, the optimal densities should depend on the specific sensors used, more precisely, the more informative the sensor is, the smaller the optimal density should be. Also, assuming that local sensors can be generally considered as weakly informative, the same phenomenon as in the context-free case, i.e. decrease of optimal density, should occur when the number of steps is increased.

5.5 Density in Agent Swarms

In many collective systems such as ants and bacteria, chemical concentrations carry directly or indirectly information about the local density of agents. This information can be used to modulate behaviour, for example by triggering aggregation or the synthesis of specific proteins. Here we investigate the empowerment properties of swarms of agents with various density sensors and density of agents. One of the goals is to evaluate the trade-off between freedom and constraint identified previously and to check its validity in more realistic scenarios.

Experiment 5.5.1. ***Density in agents swarm***
Objective: *Identify the impact of sensors and density of agents on empowerment in a swarm of agents.*
Main results: *Empowerment is directly connected to the density of agents. It is maximized for an optimal trade-off between freedom of movement and constraints imposed by the presence of other agents. The trade-off point depends on the sensors used by the agent for context-dependent empowerment, but not for context-free empowerment.*

The setup is very similar to the previous experiment, but we use a larger environment with two dimensions and several agents. Also, sensors are not based on the absolute position of the agent, but on local sensing of the density of agents around the sensing agent. Similarly to the previous experiment the density sensors are divided in two classes (directional and totalistic) and different ranges are investigated (see Fig. 5.17 and 5.18). Agents occupy only one tile and the only constraint to their movement is that they cannot move to a tile where another agent already is. The parameter we will investigate is the global density of agents on the grid. The more agents there are, the higher the probability to interact with another agent, but the lower the freedom of an agent is.

According to the results of the previous experiment, our hypothesis is that there must exist a maximum in the empowerment for a given density that is the optimal trade-off between strength of interaction and freedom. This maximum should be the same for all sensors in the context-free case whereas for context-dependent empowerment the maximum should be specific to the sensors used.

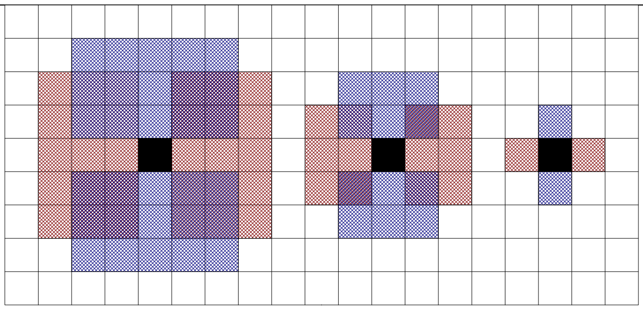

Figure 5.17: Range 3, 2 and 1 directional density sensors used in the experiment. Black squares represent sensing agents. Each sensor is divided into 4 parts (left, right, up and down) which independently estimate the density of agents in the hatched area they cover. These areas overlap for range 3 and 2 sensors. In the case of the range 3 sensor, the number of agents in each area is divided by 2 to fit the reading in 3 bits (this compression is a requirement for having a small enough state space over which empowerment quantities can be computed). The entire sensor reading is then stored on 12 bits (for example 3,2,5,6 agents leads to 1,1,2,3 after dividing by two, the final binary sensor reading is then 001001010011). For the range 2 sensors, the number of agents in each area is directly stored in 3 bits (therefore 12 bits in total). In the case of the range 1 sensor each area needs one bit, the total sensor reading is then stored in 4 bits.

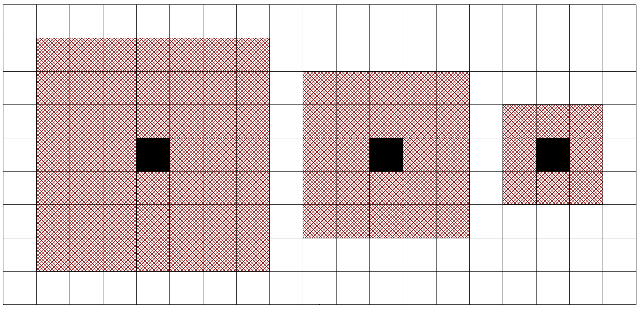

Figure 5.18: Range 3, 2 and 1 totalistic density sensors used in the experiment. Black squares represent sensing agents. Each sensor counts the total number of agents present in the hatched area. Range 3 and range 2 sensors use 6 bits. Range 1 sensor needs 4 bits.

5.5.1 Methodology

The environment is a 10 × 10 grid-world with wrapped-around boundaries. Each agent occupies one site on the grid and they are not allowed to overlap. For each density of agents and each sensor the procedure is the following. The world is initialised by distributing uniformly the agents over the space. Statistics are collected during 1000000 time-steps. At each time-step the following operations are performed:

- An agent is randomly selected.

- Its current sensory state s_t is read.
- A random action a_t is uniformly picked in the set {stay,left,right,up,down}.
- The action is performed and the next sensory state is read s_{t+1}.
- The triplet s_t, a_t, s_{t+1} is added to the statistics.

The statistics collected for all agents $p(s_t)$ and $p(s_{t+1}|a_t, s_t)$ are used to estimate 1-step context-free $\mathfrak{E}(A_t \rightarrow S_{t+1})$ and context-dependent empowerment $\mathfrak{E}(A_t \rightarrow S_{t+1}|S_t)$ using Blahut-Arimoto algorithm with stopping criterion $\epsilon = 0.00001$.

5.5.2 Results

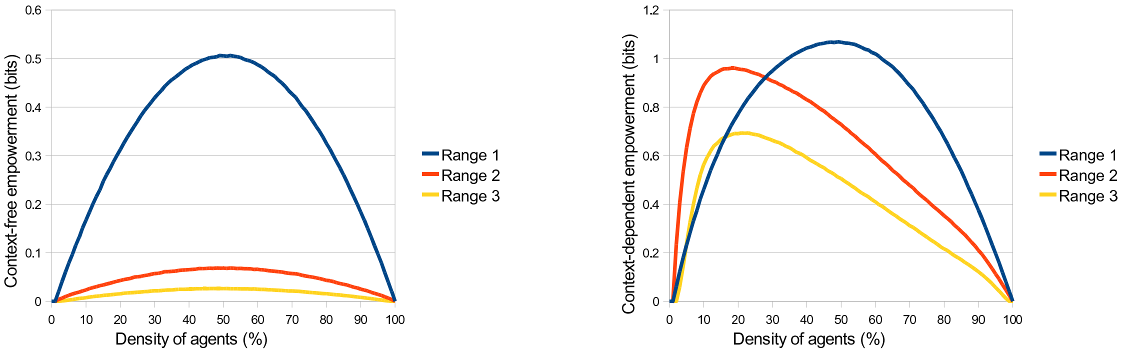

Figure 5.19: Context-free (left) and context-dependent (right) 1-step empowerment versus density of agents for the three directional density sensors depicted on Fig. 5.17.

Empowerment profiles for the directional sensor are shown on Fig. 5.19. The first thing to notice is that for each sensor range there exists a density that maximizes empowerment. At the two extreme densities, empowerment reaches minimal values, and it continuously increases when we go toward the optimal density. This observation supports our hypothesis that there exists an optimal trade-off between the strength of interaction, or the degree of constraint that agents endure, and the freedom they have. Too much freedom leads to a completely unstructured world where there is no information to gather or where the randomness of the environment is too high for the agents to make use of it. On the other side, when there is too much constraint, the agents are so restricted in their actions that they cannot have any control on their environment.

At this point a parallel can be made with the behaviour of physical systems such as water. At one extreme the system is so constrained that it results in a very ordered static structure, like an ice crystal. At the other extreme the system is completely free and it ends up in a completely random structure, or barely any structure at all like in a gas. In the middle there exists an area where the system can exhibit complex structures, such as vortices, which can exist only because there is enough order and disorder at the same time. However the parallel has to stop here because in our context agents do have autonomy, which distinguishes them from passive water molecules.

Pushing the analysis further, we can observe that in the case of context-free empowerment, the maximum values are reached for the same medium densities. Moreover whichever sensor range is used, the maximum is reached for the same density. Whereas in the case of context-dependent empowerment, changing the sensor has an effect on the position of the optimum density values. If we now consider the different densities as different environments, we can observe that the sensory apparatus an agent possesses defines a range of environmental conditions for which the agent has optimal control. This observation could speculatively be related to the concept of biological niches. In this context, the information theoretic perspective provides a natural principle for relating sensorimotor capacities to their associated niches. This can be used in two different ways, the first is that an agent with a given sensorimotor apparatus should move toward the niche where his abilities are optimal. The second way is to use this principle as a criterion for evolving sensors for a given environment (related results are described in (Klyubin, Polani and Nehaniv 2005)).

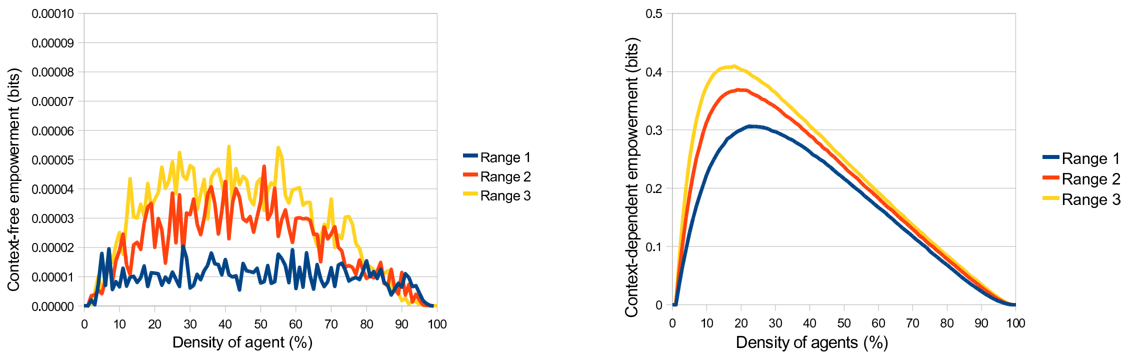

Figure 5.20: Context-free (left) and context-dependent (right) 1-step empowerment versus density of agents for the three totalistic density sensors depicted on Fig. 5.18.

Empowerment profiles of the totalistic sensors are described on Fig. 5.20. It is first important to notice that the empowerment values for the context-free case are below the stopping criterion of the Blahut-Arimoto algorithm. Other experiments have shown that when this criterion is lowered and the number of samples is increased then the context-free empowerment vanishes. Therefore it can be safely assumed that the context-free empowerment of the totalistic sensor is 0, whatever the range.

Results for the context-dependent empowerment also confirms the two hypothesis that maximum empowerment is reached at a trade-off between freedom and constraint and that the density for which the maximum is achieved depends on which sensors are available to the agent.

5.6 Spatial Organization in Agent Swarms

In the previous experiment the spatial distribution of the system was voluntarily kept unorganized by letting the agents perform random walks, leading to a more or less uniform distribution of the agents over the space. However, one of the most striking aspects of collective systems is their spatial organization. If we follow the hypothesis that agents are trying to maximize their empowerment, we should expect that they will eventually find out spatial organizations that obey this maximization principle. Therefore agent swarms should more naturally occur in some specific organizations, those that bring maximum empowerment.

We have conducted experiments in order to investigate this aspect. The general idea is to search the space of possible spatial organizations for those that maximize the empowerment of the agents. Experiments are conducted for different sensors, number of agents, and size of the environment. From the previous experiments we can hypothesize that the configurations reached will have local densities that approximate the optimal densities identified previously.

Experiment 5.6.1. ***Spatial organization in agents swarms***
Objective: *Identify spatial organizations that maximize the empowerment of the agent swarm.*
Main results: *Highly regular organizations are maximizing empowerment of the agents. These organizations are embodiment-specific.*

5.6.1 Methodology

The environment is a $N \times N$ grid world with wrapped-around boundaries. Actions of the agents are taken in the set {stay,left,right,up,down}. Directed and totalistic sensors of range 1 are used. For a given size of the world, sensor, and number of agents the search in the space of configurations is performed by using a simulated annealing algorithm that maximizes context-free empowerment $\mathfrak{E}(A_t \to S_{t+1})$. The initial solution is constructed by randomly distributing agents on the grid.

Empowerment is estimated by collecting the statistics of the perception-action loop of each agent through 'virtual' execution of their actions (i.e. each action a_t is performed and the resulting sensor reading s_{t+1} is collected and added to the statistics). All the statistics are aggregated together in order to estimate 1-step context-free empowerment. Stopping criterion for the Blahut-Arimoto algorithm is $\epsilon = 0.00001$.

A new solution is generated by altering the position of 1 agent in the following way: with probability $\frac{1}{2}$ the agent performs a local move, and with probability $\frac{1}{2}$ its coordinates are randomly changed to a non-occupied site without any locality considerations.

5.6.2 Results

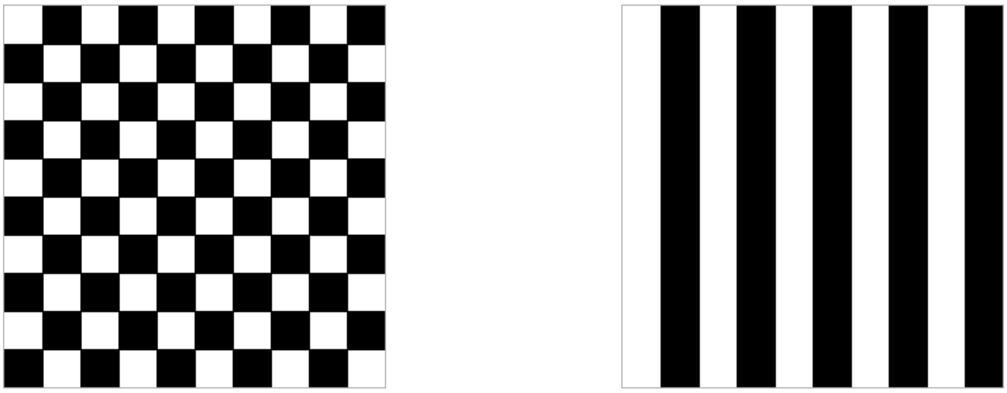

Figure 5.21: 1-step context-free maximum empowerment organizations for 50 agents in a 10×10 wraparound grid world with directional (left) and totalistic (right) range 1 sensors. Empowerment values are 2.32 bits for the directional sensor and 1 bit for the totalistic sensor. Simulated annealing parameters are $T_{start} = 0.02$, $T_{dec} = 0.00001$, $T_{stop} = 0.001$.

Figure 5.21 depicts the spatial organizations that maximize empowerment for two different sensors. To understand why these organizations are optimal one simply has to look at the perception-action loop of a single agent. Indeed, for both sensors, all the agents are surrounded by the exact same distribution of agents.

In the case of the directional sensor the organization obtained is a checker-board pattern. If we pick one agent and consider its potential actions we obtain the following 5 possibilities:

- the agent stays where it is: the sensor does not perceive any agent in any direction,
- the agent goes west: the sensor sees one agent north, one south, and one west,
- the agent goes east: the sensor sees one agent north, one south, and one east,
- the agent goes north: the sensor sees one agent west, one east, and one north,
- the agent goes south: the sensor sees one agent west, one east, and one south.

All these resulting situations are distinct from a sensory perspective. Also as the dynamic is deterministic, this means that each agent can reach 5 different sensory states by performing 1 action. The empowerment is therefore $\log_2(5) \simeq 2.32$ bits.

For the totalistic sensor the situation is different. Looking at the outcome of each actions we have the following possibilities:

- the agent stays where it is: the sensor counts 2 agents (north and south),
- the agent goes west: the sensor counts 5 agents (north-west, west, south-west, north-east and south-east),
- the agent goes east: the sensor counts 5 agents (north-east, east, south-east, north-west and south-west),
- the agent goes north: it collides and stays where it is, the sensor reads 2 agents,
- the agent goes south: same a north.

Therefore, out the the 5 actions, only 2 different sensory states can be reached, namely counting 2 agents or 5 agents. The empowerment is then $\log_2(2) = 1$ bit.

As a matter of fact, the checker-board pattern is also optimal for the totalistic sensor, however it seems that the search algorithm converges more easily to the line pattern. One reason to explain this is the following. In the case of the checker-board pattern, the two reachable sensory states of the totalistic sensor are counting 3 agents or 4 agents. Considering the search process, i.e. when the full pattern has not been completely reached yet, one can assume that the suboptimal positioning of some agents leads to a noisy version of the perception-action loop associated with the full pattern. If this noise is not uniform but instead has a Gaussian-like shape, then distinguishing between 3 and 4 agents is more difficult than distinguishing between 2 and 5 agents. Therefore in suboptimal situations the wider gap between sensor readings for the line pattern makes it a solution that is preferred to the checker-board pattern.

These results show that empowerment maximization at the global level leads to the formation of highly regular organizational patterns. Unsurprisingly, the regularities of such patterns crucially depends on the embodiment of the agents. However, because the maximization is done at the global level, it is not clear whether such organizations can emerge from the actual behaviour of the agents. This aspect is the object of the next experiment.

5.7 Organization-Inducing Behaviour

The previous section investigated the kind of spatial organizations that bring maximum empowerment to the agents. However, the agents were not actually behaving, instead, the space of possible organizations was directly searched for maximizing empowerment. This section takes a similar approach, but the searched space is that of agents' behaviours. The idea is to identify empowerment-maximizing organizations that can be induced by the behaviour rule.

Experiment 5.7.1. ***Empowerment Maximization through Organization-Inducing Behaviours***
Objective: *Identify behaviours and the resulting global organizations that maximize empowerment of the agents.*
Main results: *Maximum empowerment behaviours lead to the emergence of a single global organization with strong regularities.*

5.7.1 Methodology

The environment is a 100×100 grid with wrap-around boundaries on which 1000 agents are behaving. Actuators allow the agents to move to their immediate neighbourhood or stay at their place. Agents cannot occupy the same location. At each iteration of the simulation, one agent is selected randomly, and all its actions are 'virtually' executed to collected perception-action loop statistics $p(s_{t+1}|a_t, s_t)$ and $p(s_t)$. After the statistics have been collected, the agent picks an action according to its behaviour rule, or picks a random action with probability 0.01.

Two sensors are used in this experiment. The first one is the range 1 directed sensor shown on the right of Fig. 5.17. The second one is a variation of the range 3 directed sensor of Fig. 5.17 and will be referred to as the *range 3 partially directed sensor*[1]. It works in the following way:

- The number of agents in each 3×5 rectangular areas in the four directions is counted (see Fig. 5.17).
- The average over these four areas is computed and constitutes the 4 lower bits of the sensor reading.
- Four extra bits are dedicated to identify whether each area is above or below the average.

The agents behave deterministically and in a reactive way, i.e. their behaviour rule is defined as a mapping from the current sensor reading s_t to the current action a_t. Evaluation of a behaviour rule is performed in three steps:

- Starting from a uniform spatial distribution of the agents, 1000000 iterations are performed without collecting statistics. This gives enough time for the system to 'settle' and for the global structure to emerge.
- Agents behave asynchronously during 1000000 extra iterations, and perception-action loop statistics are collected (s_t, a_t, s_{t+1}).
- 1-step empowerment with sensor context is computed from the statistics, i.e. $\mathfrak{E}(A_t \rightarrow S_{t+1}|S_t)$.

The search is performed in the space of possible behaviours using a simulated annealing technique similar to experiment 5.6.1 where the quantity maximized is the empowerment described above.

5.7.2 Results

Figure 5.22 shows the organization obtained from the best solutions found. One can see that these organizations strongly depend on the embodiment used by the agents. In the

[1] Such a variation is needed because the volume of data to collect increases exponentially with the size of the state space. This new sensor is a lot more compact than the original range 3 directed sensor (8 bits instead of 12).

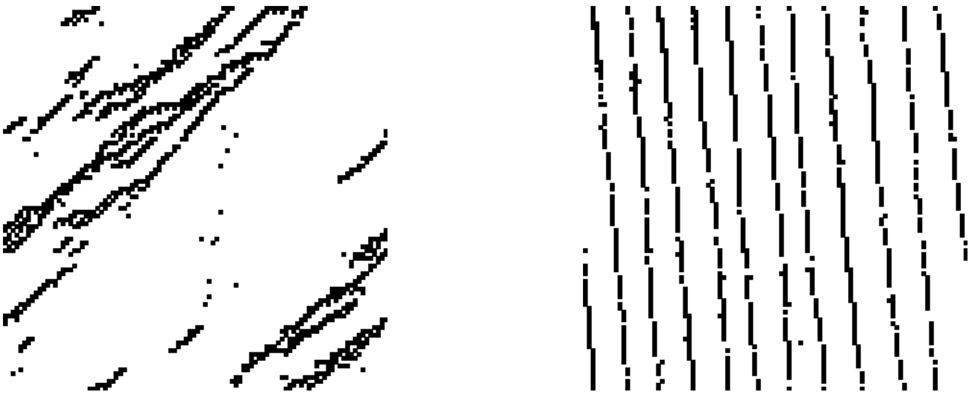

Figure 5.22: Best solutions found during the searches. **Left:** Result for range 1 directed sensor, $\mathfrak{C}(A_t \rightarrow S_{t+1}|S_t) \approx 1.20$ bits. **Right:** Result for range 3 partially directed sensor, $\mathfrak{C}(A_t \rightarrow S_{t+1}|S_t) \approx 1.38$ bits.

case of the range 1 directed sensor what is obtained is a tree-like structure, where lines of agents are branching out of each other. Such a structure allows the agent to have some freedom in their movements and provides some regularities in their environment. In the case of the range 3 partially directed sensor, maximum empowerment is obtained from a behaviour that generates a striped pattern. Interestingly, the organization generated is made of one single line that spans across the whole environment.

In both situations, the organization strongly increases the empowerment. As a point of comparison, the empowerment when the agents are behaving in a purely random way, and therefore when there is no global organization, has also been computed. For the range 1 directed sensor, the empowerment without organization is approximately 0.48 bits, and it reaches 1.20 bits when agents are using the behaviour rule identified. The change is even stronger for the range 3 partially directed sensor. Without organization, empowerment is 0.40 bits, whereas organization allows them to reach 1.38 bits.

In terms of behaviour rule, it is important to realise that in the case of the range 3 partially directed sensor, because of the very nature of this sensor, it is very difficult to obtain such a pattern using only information about the presence of agents inside the line. Instead, the emergence of this organization relies on the agents moving to situations where the horizontal densities are the same on both sides. Therefore, it is the horizontal spacing between the agents, and therefore between the lines, that globally creates this pattern.

5.8 Spatial Organization from Selfish Empowerment Maximization

In the previous experiment, it was shown that behaviour evolved for global empowerment maximization could lead to highly organized systems where all the agents have a high empowerment. However, if we assume that single agents try to selfishly maximize their own empowerment, it is not clear how such a principle would drive a system made of many such agents. This question is addressed in this section. The general idea is to derive local empowerment-maximizing behaviour rules from perception-action loop statistics and to look at the global organizations that are generated by such agents.

Experiment 5.8.1. ***Collective behaviour driven by local empowerment maximization***
Objective: *Study the global organizations obtained from selfish empowerment maximization at the agent level.*
Main results: *Several complex organization are obtained as the consequence of the 'coevolution' between the agents' behaviour and the resulting global organization. However, this does not generally result in highly empowered agents.*

5.8.1 Methodology

The environment is a 100 × 100 grid world with wrapped-around boundaries and 1000 agents. Actions are taken in the set {stay,left,right,up,down}. The sensor used is the range 3 partially directed sensor described in the previous experiment. The agents are equipped with an 'empowerment map' that associates every sensory state with an empowerment value. This empowerment map defines the behaviour of the agent.

The simulation is divided into epochs, during which the following steps are performed 10000000 times:

- An agent is randomly picked.
- Its perception-action loop statistics are collected by 'virtually' executing all its actions and updating the channel distribution $p(s_{t+1}|a_t, s_t)$ accordingly (this distribution is common to all agents, and it is reset at each epoch.
- The empowerment of the outcome s_{t+1} of each action is retrieved from the agent's empowerment map.

- The actions that lead to maximum empowerment are kept as candidates, and one of them is randomly picked.

At the first iteration, the empowerment map is zero everywhere, therefore the agents have no preference for any particular state. This leads the agents to perform random walks in the grid. At the end of each epoch the empowerment for each sensory state is computed, i.e. $\mathfrak{E}(A_t \to S_{t+1}|s_t)$, and the empowerment map of each agent is updated with the new empowerment values.

The general behaviour obtained from this rule is that each agent tries to make a move to get to locally more empowered situations, e.g. they are performing greedy empowerment maximization. Their perception of which situations have maximum empowerment is based on global statistics of the perception-action loop collected during the previous epoch.

5.8.2 Results

Some of the organizations resulting from this experiment are shown on Fig. 5.23. One can see that, using this simple empowerment maximization rule, a wide range of different global organizations emerge. These range from dense clustering of the agents to evenly spaced distributions over the whole space, going through organizations consisting of different density regions.

Note, however, that the empowerment resulting from the organizations is not as high as those obtained by directly evolving behaviour rules for maximum empowerment, as was done in the previous experiment. Indeed, the empowerment of the agents in the organizations of Fig. 5.23 ranges from 0.53 to 0.60 bits. This is far below the maximum empowerment achieved for the same sensor in the previous experiment, i.e. 1.38 bits. This is actually only slightly above the baseline empowerment when no organization is present, i.e. 0.40 bits.

Therefore, in such a scenario, selfish empowerment maximization does not generally lead to global organizations providing high empowerment to the agents. This can be explained by the fact that when one agent moves to a more empowered situation, it might at the same time reduce the empowerment of surrounding agents. This is typically the case when dense structures are obtained. In such a situation, agents are trying to get close to other agents while preserving some freedom of movement, however, other agents doing the same may actually push the previous ones in situations where they can barely move at all, hence reducing their empowerment.

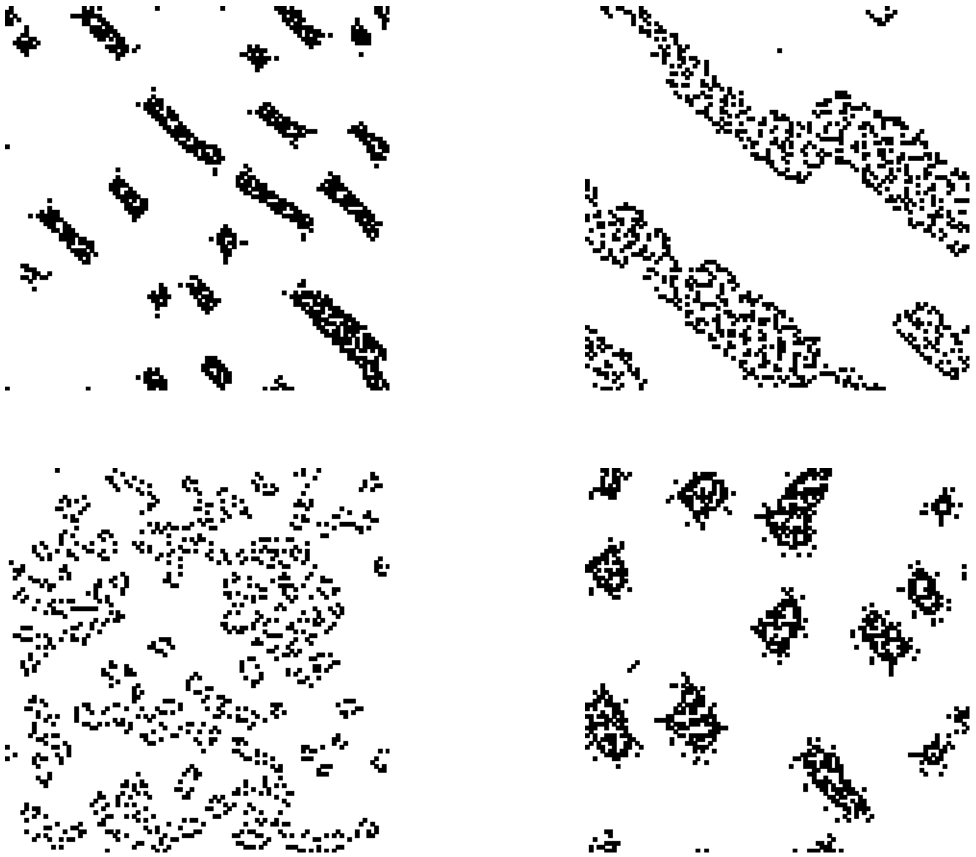

Figure 5.23: Some of the organizations obtained from selfish local empowerment maximization. Black squares are agents, white squares are empty tiles.

Nevertheless, it is interesting to see that this scenario creates a mechanism of 'coevolution' between the behaviour of the agents and the resulting organizations. Indeed, the behaviour of the agents induces structure in the environment and, because this structure creates specific regularities in the perception-action loop, they induce new empowerment maximizing behaviour rules in return, which creates new structure, and so on... It is this coevolution mechanism that has the ability to generate a wide range of complex organizations.

5.9 Summary

In this chapter I investigated empowerment in the context of multi-agent systems interacting in a spatial world. Interactions between agents were limited by forcing them to act asynchronously, removing the effect of simultaneous actions (these will be treated separately in the next chapter).

In the first section, a minimal environment with two agents, the size 6 line world with periodic boundaries, was investigated. Such a simple model is useful in order to identify fundamental properties of the agents that may impact their empowerment. The state of this environment is defined by the absolute positions of the two agents A and B. Only the empowerment of agent A was investigated. Agent B was not allowed to act, being merely an obstacle for the other agent. However, its size and positions were varied.

The empowerment of agent A has been studied in several different situations. The following parameters have been investigated:

- The size of agent B, ranging from size 0 (in which case agents do not collide) to size 5 (in this case agent A cannot move).
- The sensor used by agent A: either the full state of the channel, the position of A, the position of B, the full state with added noise, or density sensors.
- The context used by the agent: no context or the sensor reading.
- The number of empowerment steps considered (source horizon).

One of the key results of these experiments is that agent A generally has maximum empowerment at a trade-off between maximum freedom (agent B of size 0) and maximum constraint (agent B of size 5). The reason for this is that the agent needs some freedom to move, but it also needs to have structure in the environment which can be provided by the presence of agent B (they collide). When agent A does not know the state of the environment (assuming that this state is a uniform distribution over possible starting positions of both agents), i.e. in the context-free case, it was found that the empowerment profile (its general shape), is the same regardless of the sensor being used.

On the other hand, in the context-dependent case, the empowerment profile is specific to the sensor being used by the agent. Also, the more informative the context is (in terms of how much information it captures about the state of the environment), the more the empowerment profile deviates from the context-free case, favouring situation of high freedom.

In a second set of experiments, I investigated the empowerment of multiple agents interacting on a 2-dimensional grid world with periodic boundaries. All the agents perform random walks on the grid, and collide with each other. Two kinds of local sensors were studied: directional and totalistic sensors, with different ranges.

Experimental results confirmed the hypothesis from the first section: there is an optimal density of agents, i.e. a trade-off between freedom and constraints, for which empowerment is maximized. Moreover, this maximum density does not depend on the sensor used in the context-free case, whereas it is specific to the sensor when it is used as a context. Put another way, for a given sensorimotor apparatus, there are specific conditions of the environment for which context-dependent empowerment is maximized. It also means that, from an evolutionary perspective, agents are likely to have sensors and actuators that are 'tuned' to their usual environmental conditions.

The next set of experiments set out to identify the impact of spatial organization on the agents' empowerment. For this purpose, the same grid world was used, and the space of possible spatial organizations was searched in order to find those that bring maximum empowerment to the agents. Results of these searches showed that empowerment is maximum when highly regular organizations are formed, e.g. checkerboard or line patterns, depending on the sensors available to the agents.

Similar results, although not as regular as the previous ones, were found when the space of behaviours was searched instead of the space of organizations. Therefore, empowerment can be maximized through the emergence of a global organization resulting from the agents' behaviour.

In the last experiment, the behaviour of many agents selfishly maximizing their own empowerment was investigated. By doing so, a process of 'coevolution' between the behaviour of the agents and the generated organizations was induced, leading to a wide range of complex structures emerging at the global level. However, this mechanism was not successful in providing a high empowerment to the agents. Indeed, even though agents tried to maximize their own empowerment, in some cases the resulting global behaviour led to suboptimal situations for many agents, for instance many of them could be stuck in a very dense cluster, hence preventing them from moving, leading to a low empowerment.

Chapter 6

Interferences and Coordination Between Agents

In many situations, multiple agents acting together in the same environment may *interfere* with each other. For example, one agent can try to 'pull' in one direction while the other agent 'pulls' in the other direction. What happens in such a case depends on the environment, but from the information-theoretic perspective of embodied agents, such an environment can also be seen as a communication channel.

This chapter investigates these situations using concepts from network information theory, and mainly the multiple-access channel. However, the perspective presented in this thesis differs from the standard literature. Indeed our focus is on the control abilities and empowerment of agents operating with such channels.

Minimal models of two agents interacting in a shared environment are investigated in order to uncover the fundamental mechanisms and links between information-theoretic quantities of interest. Studying these minimal models allows us to identify properties that may prove useful in understanding more complex situations.

6.1 Information-Theoretic Picture

To understand what interferences are and how they impact on information-theoretic quantities one needs to look at the perception-action loop of two agents sharing a common environment (see Fig. 6.1). As one can see, the perception-action loops are intertwined.

The results is that the information flowing from the actions 'collide' in the environment R before flowing into the next sensor states.

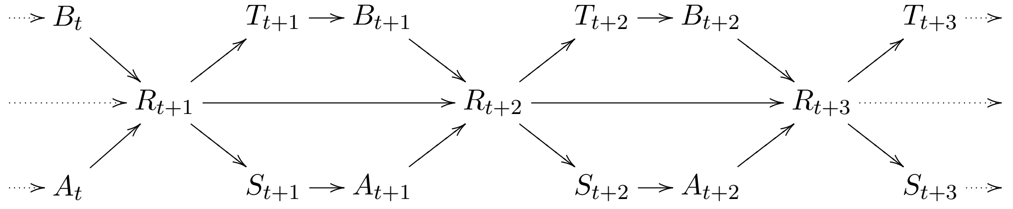

Figure 6.1: Typical perception-action loop of two agents sharing the same environment unrolled over three time-steps. The agents have no memory beyond their immediate sensor readings. R stands for the state of the environment. The first agent's sensor and action variables are respectively S and A, those of the second agent are denoted by T and B.

In order to get a clearer picture of these interferences, it is useful to consider a minimal model of the perception-action loop. This can be seen as a first step in understanding interactions between agents from an information-theoretic perspective. Indeed, by looking at minimal models in which interferences appear, it is possible to identify key aspects of this phenomenon and generalize to, or at least get an intuition of, more complex models.

For this purpose, we can first notice that the bottleneck occurs when the two actions are transmitted to the environment R. The resulting sensor readings are just noisy copies of R. Therefore, in order to simplify the model, one can remove the sensors and assume that the agents are able to directly sense the state of the environment. Secondly, because we are mainly interested in the effect of the two actions 'colliding' in the environment during one timestep, we can simplify the model by considering that the channel has no memory, and therefore that the agents are not allowed to pick their actions according to the previous state of the environment. This leads us to a minimalistic perception-action loop with only three variables and no time dependence: actions A and B and the resulting state of the environment R as depicted on Fig. 6.2.

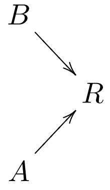

Figure 6.2: Minimalistic model of the perception-action loop for two agents. Agents and the environment have no memory. The agents perceive the full state of the environment R.

In this minimal model, the environment is fully described by the conditional probability distribution $p(r|a, b)$ and the agent policies by $p(a)$ and $p(b)$. In the network information-theory literature, this model is referred to as a 2-users *multiple-access channel* (MAC). One may study MACs with more than two users. However, considering only two of them is sufficient to identify the properties of these channels that are relevant to this thesis. A discrete MAC can be more precisely specified by adding the number of symbols for each input alphabet and for the output alphabet. For example a $(2, 2; 4)$-MAC has two symbols for the first user, two symbols for the second one, and four symbols for the output.

6.2 Interferences

Considering the communication channel that goes from A to R, the policy of agent A has an impact on the amount of information that is transmitted through the channel of A, but it has no effect on the capacity of the channel itself and therefore the behaviour of agent A does not change its empowerment. However, because the action picked by B can be seen as the state of the channel, knowledge about this action can potentially increase empowerment. The impact that one agent has onto the sensorimotor channel of the other agent is qualified as an *interference.*

Because of these interferences, the policy of one agent may modify the capacity of the actuation channel of the other agent. These are a consequence of the probabilistic structure of the channel studied. The next section investigates the space of possible channels in order to identify those for which interferences exist. Moreover, if we consider that the actions can be correlated, one can ask the question: How does the coordination between the two agents impact their empowerment? This question is also investigated in the following sections.

6.3 Capacity of Multiple-Access Channels

In a standard MAC, the input distributions are assumed to be independent. One of the main quantities that has been investigated in the literature is the amount of information that can be transmitted in the channel $A, B \to R$ under these assumptions. This is referred to as the *total capacity* or *sum capacity* of the MAC (the first term will be used throughout the thesis). Computing this quantity for an arbitrary MAC has long been an open problem. It is only recently that an algorithm was proposed in (Rezaeian and Grant 2004).

In this section we first detail the algorithm used to compute the total capacity of a MAC and discuss its relationship with the Blahut-Arimoto algorithm. Using the presented algorithm, the space of two-users MACs is numerically investigated in order to identify limits on capacity increase through coordination. Different channels are presented in order to illustrate some properties of MACs. In the last part we study simple scenarios involving embodied agents to show how these properties impact empowerment.

6.3.1 Computation of the Total Capacity

In order to understand how the total capacity of a MAC is computed, it is easier to first look at the simplest case, i.e. a channel with only one user. The standard Blahut-Arimoto algorithm (Blahut 1972, Arimoto 1972) for finding a capacity achieving distribution of a single-user channel $p(y|x)$ can be described as follows. The goal is to find a distribution $p^*(x)$ such that $I_{p^*(x)}(X;Y) = C$ where C is the capacity of the channel. The Blahut-Arimoto algorithm uses the following steps (following the notation of (Rezaeian and Grant 2004)):

- Initialise $p^0(x)$.
- Iterate until convergence $\forall x \in \mathcal{X} : p^{r+1}(x) = p^r(x) \frac{exp\big(I^r(x;Y)\big)}{\sum_{x'} exp\big(I^r(x';Y)\big)}$ where the information function is defined as $I^r(x;Y) = D_{KL}\big[p^r(Y|x) \| p^r(Y)\big]$.

The general idea is that the probabilities of each symbol are updated proportionally to their individual contribution to the total mutual information.

Two remarks have to be made about the initial distribution $p^0(x)$. In the case where the initial distribution is uniform, we are guaranteed to converge to the unique maximum entropy capacity achieving distribution. However if different initial distributions are used, the algorithm may converge to other capacity achieving distributions that do not necessarily have maximum entropy. A special case is when a symbol x has probability $p^0(x) = 0$; because its probability cannot move away from 0 this situation is equivalent to finding the capacity of a reduced channel in which x has been removed from the input alphabet.

In the single-user channel the computation of the capacity is a convex problem, for this reason the Kuhn-Tucker condition[1] is sufficient to achieve capacity (Boyd and Vanden-

[1] The Kuhn-Tucker conditions are a generalization of Lagrange multipliers to optimization problems with inequality constraints.

berghe 2004). However computing the total capacity of a MAC is not a convex problem. The work of (Watanabe and Kamoi 2002) shows that for any MAC either the Kuhn-Tucker condition is sufficient or the MAC can be decomposed into sub-MACs for which the condition is sufficient.

In (Rezaeian and Grant 2004) the authors translate this distinction using the concept of a *regular* MAC. A MAC is regular if the size of the input alphabet for each user is smaller than or equal to the size of the output alphabet. If a MAC is not regular, then it can be decomposed in a set of regular sub-MACs by removing letters from the input alphabet. The capacity of each regular sub-MAC can be computed and the total capacity of the MAC is the maximum capacity achieved among all sub-MACs.

We now describe the algorithm for finding the total capacity and an input distribution for a regular MAC that was introduced in (Rezaeian and Grant 2004). It has the same structure as the standard Blahut-Arimoto which is a special case of this algorithm. Let us consider that we have a M-users MAC described by the conditional probability distribution $p(y|x_1, ..., x_M)$. The algorithm works as follows:

- Initialise $\forall m \in 1, ..., M : p_m^0(x)$.
- Iterate until convergence $\forall m \in 1, ..., M, \forall x_m \in \mathcal{X}_m$:

  $$p_m^{r+1}(x_m) = p_m^r(x_m) \frac{\exp\big(I_m(x_m; Y)\big)}{\sum_{x'_m} \exp\big(I_m(x'_m; Y)\big)} \tag{6.1}$$

  where the information function is defined as

  $$I_m(x_m; Y) = \sum_{x_{\overline{m}}} p(x_{\overline{m}}) I(x_m, x_{\overline{m}}; Y) \tag{6.2}$$

  and $x_{\overline{m}}$ denotes the combination of inputs from all users except user m. When computing $p_m^{r+1}(x_m)$ the information function is calculated according to $p_n^{r+1}(x_n)$ for all $n \leq m$ and to $p_n^r(x_n)$ for all $n > m$.

One can see that in the case $M = 1$ the information function reduces to $I(x_m; Y)$ and the algorithm is equivalent to the Blahut-Arimoto algorithm.

As for the single-user case there may be several capacity-achieving distributions, however one difference is that there may exist multiple maximum entropy capacity-achieving distributions. Considerations about initial distributions are the same as for the Blahut-Arimoto algorithm, so initialising with uniform distributions leads to a maximum entropy capacity-achieving distribution, but there is a practical consideration that needs to be mentioned. There exist channels for which the algorithm will get stuck on the initial uniform distribution. The reason for this is that there are multiple maximum entropy capacity-achieving distributions and they are all equally distant from the uniform distribution. Therefore the algorithm cannot select between the solutions because all gradients are equal. This problem can be avoided by introducing noise into the initial distribution.

6.3.2 The Space of (2,2;4) MACs

We denote the total capacity by $C_\perp$, where $\perp$ symbolises the independence between the input distribution of the agents. For a two users MAC, the total capacity is defined as

$$C_\perp = \max_{p(a),p(b)} I(A,B;R). \tag{6.3}$$

A total capacity-achieving input distribution will be denoted by $p^*_\perp(a,b)$ which, because of the independence, can also be written as $p^*_\perp(a)p^*_\perp(b)$.

If the independence constraint on input distributions is relaxed, basically allowing correlations between the input distributions, then the MAC can be considered as a standard channel with one user. The capacity of this channel will be referred to as the *joint capacity* and denoted by C. It is then defined as

$$C = \max_{p(a,b)} I(A,B;R). \tag{6.4}$$

A joint capacity achieving input distribution will be denoted by $p^*(a,b)$.

It is straightforward to see that $C \geq C_\perp$ for the simple reason that the search space of input distributions for $C_\perp$ is included in that of C. However one can ask the question: How much capacity can be gained if correlations between input distributions are allowed? Therefore we are interested in the capacity gain from correlation which is defined as

$$\Delta C = C - C_\perp. \tag{6.5}$$

The following experiment tries to answer this question.

Experiment 6.3.1. ***Numerical investigation of the capacity of regular (2,2;4) MACs***

Objective: *Identify how the joint capacity, the total capacity and their difference are related.*

Main results: *Numerical results lead to the conjecture that allowing correlations can at most double the capacity. A related conjecture was proposed by (Thomas 1987).*

The space of regular (2,2;4) MAC is discretized and sampled. The discretization is done by considering only the channels for which the conditional probability distribution can be expressed as $p(r|a,b) = \frac{n}{N}$ where $\{n \in \mathbb{N}; 0 \leq n \leq N\}$ and N is fixed. The discretization parameter N is different for each experiment. For each channel, the joint capacity C and the total capacity $C_\perp$ are computed.

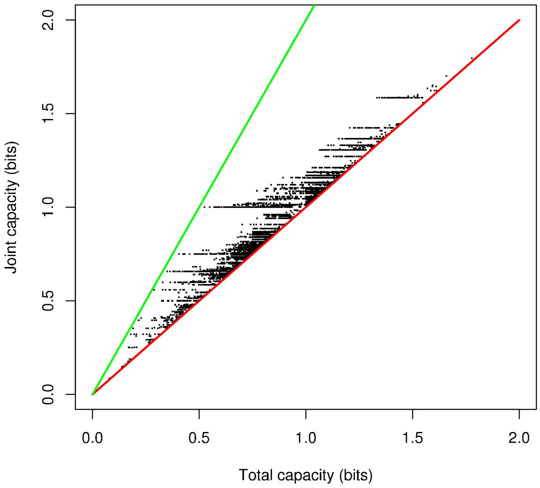

Figure 6.3: Numerical exploration of joint and total capacity in the space of (2,2;4)-MACs. Joint capacity C versus total capacity $C_\perp$. The red and green lines respectively have equations $y = x$ and $y = 2x$. Discretization parameter for the sampling process is $N = 4$.

Results of this experiment are shown on Fig. 6.3. From the joint capacity versus total capacity graphs one can see that indeed $C_\perp \leq C$, but more interestingly it seems that the joint capacity cannot be more than twice the total capacity of the MAC. Different samplings techniques were also used to search these spaces without being able to find any channel were this would be violated. Therefore we can conjecture that for any regular discrete memoryless (2,2;4)-MAC the joint capacity is at most twice the total capacity.

A related conjecture is presented in (Thomas 1987) which shows that the total capacity of any MAC with white Gaussian noise can be at most doubled when the channel is used with feedback (even for more than two users). This might sound as if it has nothing to do with coordination at the inputs, but it actually has. Feedback in that case allows the users to have a common source of information which they can use to correlate their messages. The author also shows that for a general MAC with k users the total capacity with feedback is bounded from above by kC where C is the total capacity without feedback. The paper concludes by conjecturing that the capacity doubling upper limit might hold for the general MAC with k users.

Unfortunately, I could not prove the capacity doubling conjecture. At this stage, one remark can be made about the capacity achieving distributions. Consider a joint capacity achieving distribution $p^*(a,b)$ and its marginalisation into independent distributions $p^*(a)p^*(b)$. One can ask whether the marginalised distribution is also a total capacity-achieving distribution. It turns out that this is typically not the case. However, there is a class of channels for which it is true. Such channels generally have a symmetric structure, meaning that the conditional probabilities of the channel for each user are the same up to a permutation of the symbols. Because of the mathematical simplification that follows from this property, proving the conjecture for this class of channels might be an achievable first step.

6.4 Input Correlations and Capacity Gain

In the previous subsection we looked at how much more capacity can be achieved if the inputs are allowed to be correlated. We now investigate the link between the amount of correlation that is used to achieve joint capacity and the capacity gain. The same space of channels as in the previous experiment is sampled but the measured quantities are the capacity gain ΔC and the amount of correlation in the joint capacity achieving distribution $I_{p^*(a,b)}(A;B)$.

Experiment 6.4.1. ***Numerical investigation of the capacity gain from correlation in regular (2,2;4) MACs***
Objective: *Identify how the amount correlation is related to the capacity gain.*
Main results: *Numerical results show that the capacity gain is not bounded from above by the amount of correlation.*

The space of regular (2,2;4)-MACs is sampled similarly to the methodology introduced in the previous experiment. Results are shown on Fig. 6.4.

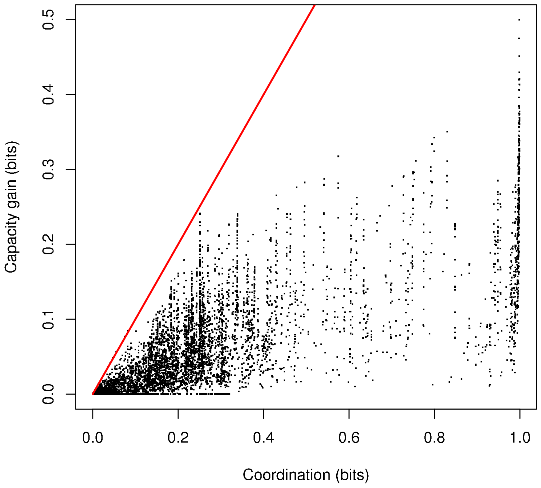

Figure 6.4: Numerical exploration of the space of (2,2;4)-MACs. Capacity gain ΔC versus coordination $I_{p^*(a,b)}(A;B)$. The red line has equation $y = x$. Discretization parameter is $N = 4$. There are some points just above the $y = x$ line, for example at coordinates $(0.0817, 0.08496)$ (see text for details).

It seems at first sight quite intuitive to expect that the amount of correlation between inputs is acting as an upper bound on the capacity gain. This is indeed the case for many of the channels that were sampled as can be seen on Fig. 6.4. However, there are a few of them for which this is not true. These are almost invisible on the graphs because the increase is very low, but they do exist at least for the (2,2;3) and (2,2;4)-MAC.

One such channel is the deterministic binary erasure channel with two sources. This channel and its capacity achieving distributions are described by the following table:

$\mathbf{p(r\|a,b)}$	r	0	1	2	$\mathbf{p^*(a,b)}$	$\mathbf{p^*_\perp(a,b)}$
a	b					
0	0	1	0	0	$\frac{1}{3}$	$\frac{1}{4}$
0	1	0	1	0	$\frac{1}{6}$	$\frac{1}{4}$
1	0	0	1	0	$\frac{1}{6}$	$\frac{1}{4}$
1	1	0	0	1	$\frac{1}{3}$	$\frac{1}{4}$

The total capacity of this channel is $C_\perp = 1.5$ and its joint capacity is $C = \log_2(3) \approx 1.584963$ bits. The joint capacity achieving distribution has coordination $I_C(A;B) \approx$

0.0817. Therefore the capacity gain $\Delta C \approx 0.08496$ is higher than the coordination.

To summarise, this experiment shows that in some channels, the amount of information that can be sent through the channel is increased if the sources are correlated. If we take the perspective of embodied agents, because the capacity is strongly linked to the controllability of the system, this means that several agents can improve their control on the environment if they coordinate their actions.

6.5 Classes of Multiple-Access Channels

The previous section looked at global properties of the space of multiple-access channels. We now focus on some specific channels of interest and study them more thoroughly. The goal is to find the properties of channels that favour coordination. The second question we want to answer is how much does the coordination effect the amount of information that is transmitted through the channel. The last experiment showed that in many channels, the amount of coordination to reach joint capacity was a limiting factor on the capacity gain. However this might not be the case for input distributions which do not achieve capacity.

In this section we introduce three different classes of multiple-access channels:

- non-interfering MACs,
- weakly-interfering MACs,
- and strongly-interfering MACs.

Each class is detailed in the following subsections along with their properties. Typical MACs representative of each class are numerically investigated according to the following methodology. For each of them we identify the *capacity region* of the MAC when used without coordination of the sources. A typical capacity region is represented in Fig. 6.5. According to (Cover and Thomas 2006) p. 526, the capacity region is defined as the closure of the convex hull (R_1, R_2) satisfying

$$\begin{aligned} R_1 &\leq I(A; R|B), \\ R_2 &\leq I(B; R|A), \\ R_1 + R_2 &\leq I(A, B; R). \end{aligned} \tag{6.6}$$

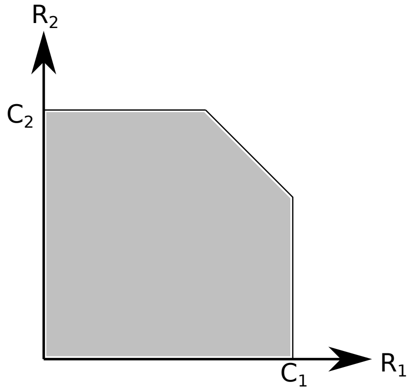

Figure 6.5: Example of a capacity region in a two-users MAC. C_1 and C_2 are the maximum capacities of each user. Grey area is the feasible region.

This capacity region makes the assumption that the input of all the senders are known in order to evaluate the amount of sent information. In the model presented, this is only the case if the receiver's perspective is taken.

A second capacity region is looked at which we refer to as the senders' perspective capacity region. For this one simply assumes that the senders do not know what the others are sending. It is defined as the convex hull (R_1, R_2) satisfying

$$\begin{aligned} R_1 &\leq I(A;R), \qquad (6.7)\\ R_2 &\leq I(B;R),\\ R_1 + R_2 &\leq I(A,B;R). \end{aligned}$$

In order to identify these regions, the space of independent input distributions is discretized and sampled according to the methodology described in Experiment 6.3.1.

Similarly the space of joint distributions is also sampled. For each of them the joint rate $I_{p(a,b)}(A,B;R)$ and the total rate $I_{p(a)p(b)}(A,B;R)$ are computed and compared. The last value we look at is the amount of coordination $I(A;B)$ which is compared to the rate gain $I_{p(a,b)}(A,B;R) - I_{p(a)(b)}(A,B;R)$ (which can be negative).

6.5.1 Synergy in the MAC

A key quantity to understand information rates in the MAC is the *synergy*. Synergy colloquially refers to the case where the whole is more than the sum of its parts. It has been used for instance in neural networks as a way to measure the amount of information that a set of neurons encodes about a stimulus that is not contained in any individual neuron (Brenner, Strong, Koberle, Bialek and Van Steveninck 2000). It is also referred to as *co-information* (Bell 2003).

In the context of the 2-users MAC presented above, synergy is a property of both inputs and the output of the channel. It can be seen as the amount of information that is jointly transmitted through the channel but that none of the inputs individually transmits.

Let us consider the total amount of information transmitted through the MAC, i.e. $I(A, B; R)$. This quantity can be decomposed as follows:

$$I(A, B; R) = I(A; R) + I(B; R) + I(A; B|R) - I(A; B). \tag{6.8}$$

The first two terms $I(A; R)$ and $I(B; R)$ are the individual contributions of each input. The remaining part is the synergy. In the three variables case, synergy can be expressed in three different ways:

$$\begin{aligned} Syn(A; B; R) &= I(A; B|R) - I(A; B) \qquad (6.9) \\ &= I(A; R|B) - I(A; R) \\ &= I(B; R|A) - I(B; R) \end{aligned}$$

As one can see from these expressions, because mutual information is a positive quantity, the synergy can be either positive or negative.

Let us first consider the case of positive synergy. If the inputs of the channel are independent, then $I(A; B) = 0$ and therefore the synergy can be expressed as

$$Syn(A; B; R) = I(A; B|R) \tag{6.10}$$

which is indeed a non-negative quantity.

Consider the case of a MAC which copies binary inputs into independent degrees of freedom of the output. R is then a compound variable containing both A and B. In this case, the total amount of information transmitted is simply the sum of the independent contributions; the synergy is equal to 0.

Let us now consider the case of a MAC where the output is a XOR of the inputs. Assuming that the inputs are uniformly distributed and independent, the total amount of information transmitted through the MAC is $I(A, B; R) = 1$ bit. Where does this information come from? If we look at the independent contributions, one can see that these are actually 0. Indeed, when B is unknown, A and R appear uncorrelated (and the same is true for B and R when A is unknown).

Now if we look at the synergy $I(A; B|R)$ in the case $R = 0$, because of the XOR gate it means that either $A = B = 0$ or $A = B = 1$. Therefore, if we know the output R and one of the inputs, the other input can be deduced. The same reasoning goes for the case $R = 1$. This means that, when the output is known, the inputs are correlated. Hence, the XOR MAC with uniform inputs has one bit of synergy. More information is sent through the channel than the individual contributions of each input.

When does synergy become negative? As one can see from the expressions above, this can be the case only if the inputs of the channel are correlated. Consider the first MAC example from above. When the inputs are independent and uniformly distributed, each input contributes 1 bit to the output, therefore the total amount of information transmitted is 2 bits.

Now if the inputs are correlated, for example if B is a copy of A, then each of them still independently contributes 1 bit to the output, but because they are correlated, it is actually the same bit that they are transmitting. Therefore the total amount of information transmitted through the MAC is 1 bit. In this case, the synergy is:

$$Syn(A; B; R) = I(A; B|R) - I(A; B) = 0 - 1 = -1. \tag{6.11}$$

In this situation, negative synergy can be seen as a measure of the amount of redundant information. Less information is transmitted through the channel than the sum of the independent contributions, because these contributions are redundant.

6.5.2 Non-Interfering MACs

This first class of channels is the simplest one. It can be characterised by the fact that each source of the MAC sends information into independent degrees of freedom at the output. It is quite obvious that in such channels the amount of information independently sent by one user is unaffected by that of the other user. Put differently: any rate is reachable by any user independently of the other users. As a consequence, introducing correlations between sources can only decrease the total amount of information sent, for the simple reason that correlations reduce the degrees of freedom available to the users, and therefore reduce the rate at which they can send information through the channel.

The most prototypical channel of this class is a composition of two binary symmetric channels. It is referred to as the binary symmetric (2,2;4)-MAC and described by the following conditional probability distribution:

$p(r\mid a,b)$	r	00	01	10	11
a	b				
0	0	1	0	0	0
0	1	0	1	0	0
1	0	0	0	1	0
1	1	0	0	0	1

Experiment 6.5.1. ***Numerical investigation of the binary symmetric (2,2;4)-MAC***

Objective: *Identify the relationship between amount of coordination and amount of information sent.*

Main results: *Coordination reduces the amount of information sent through the channel. The relation between amount of coordination and rate loss is linear.*

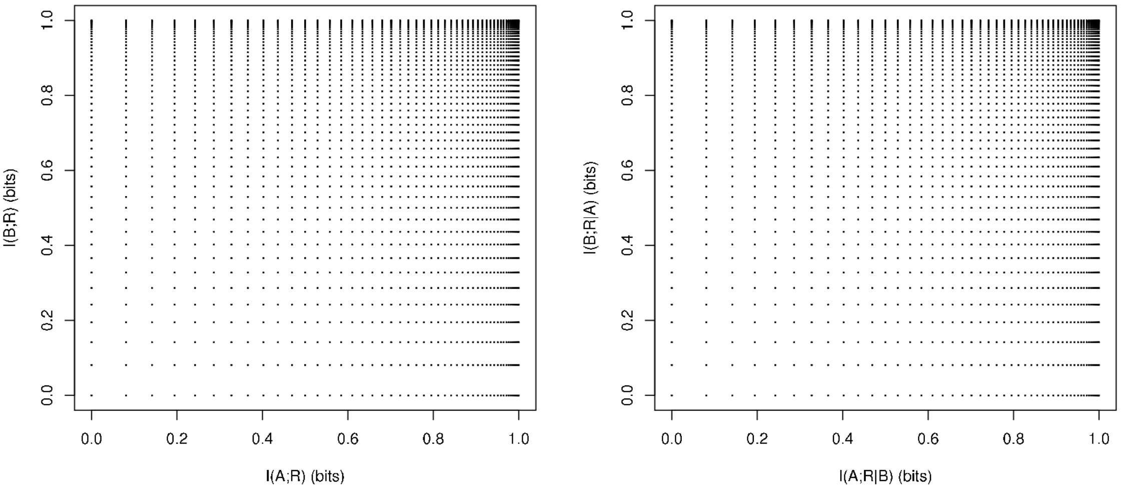

Figure 6.6: Numerical exploration of the space of input distributions for the binary symmetric (2,2;4)-MAC. Left: capacity region from the sender's perspective. Right: capacity region from the receiver perspective. Discretization parameter is $N = 100$.

Results of this experiment are shown on Fig. 6.7. From these results one can see

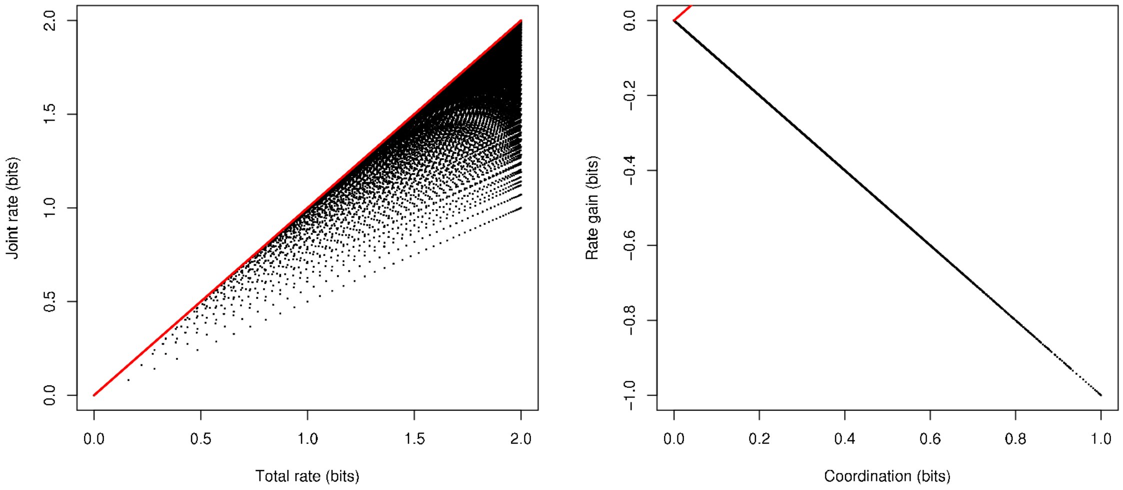

Figure 6.7: Numerical exploration of the space of input distributions for the binary symmetric (2,2;4)-MAC. Left: amount of information sent through the channel versus that of the factorised input distribution. The red line has equation $y = x$. Right: rate gain (or loss) g versus coordination. Because both senders transmit into independent degrees of freedom, coordination directly decreases the amount of transmitted information. The red line has equation $y = x$, the data points are on the line $y = -x$. Discretization parameter is $N = 100$.

that coordination does not improve the amount of sent information. Indeed, whatever the amount of information is transmitted by a joint input distribution, the corresponding marginalised input distribution transmits at least as much information. On the rate gain versus coordination graph, one can actually see that coordination is inversely correlated with the rate gain. Put another way, the amount of information used for coordination is directly transformed into a loss of transmitted information.

These results are not really surprising. The binary symmetric MAC is a compound of two channels which are independent from each other. Sources in this channel do not interfere with each other. This class of channels have the property that the capacity gain from coordination $C - C_\perp$ is equal to 0, meaning that the same information rates can be achieved with or without coordination of the sources. However the converse is not necessarily true, the next channel is an example of such a situation.

6.5.3 Weakly-Interfering MACs

This class of channel is characterised by the following properties:

- they do not belong to the non-interfering class, i.e. the rate achievable by one user depends on the rates achieved by the other users,

- the total capacity and the joint capacity are the same, i.e. there exist independent inputs distributions that transmit as much information as any joint input distribution.

A good example of such a channel is the XOR (2,2;2)-MAC. Basically the output is computed as a XOR of the two inputs. It has the following conditional probability distribution:

$p(r\|a,b)$	r	0	1
a	b		
0	0	1	0
0	1	0	1
1	0	0	1
1	1	1	0

Experiment 6.5.2. ***Numerical investigation of the XOR (2,2;2)-MAC***
Objective: *Identify the relationship between amount of coordination and amount of information sent.*
Main results: *Coordination can improve the amount of information transmitted through the channel. However capacity can be achieved without any coordination. The relationship between amount of coordination and information gain is non-linear.*

As one can see from the results on Fig. 6.9, the XOR (2,2;2)-MAC channel behaves quite differently from the binary symmetric one. A similarity is that for many input distributions, the corresponding marginalised distribution actually transmits at least as much information. However there exist some joint input distributions that are able to send more than their marginalised counterparts (e.g. all the points above the red line on Fig. 6.9). Such distributions do not appear anymore when considering only capacity-achieving situations. This is directly connected to the fact that the capacity gain from coordination ΔC is 0. Therefore such channels cannot be identified directly using this property.

Although similarly to Experiment 6.3.1 about the space of MACs, we are tempted to hypothesise that the information gain is bounded from above by the coordination or linearly related to it, but one can see from Fig. 6.9 that this is not the case. The coordination is not acting as a bound, nor is it linearly related to the information gain (or loss). Indeed the relationship is actually non-linear, and no simple functional relation could be identified at this point.

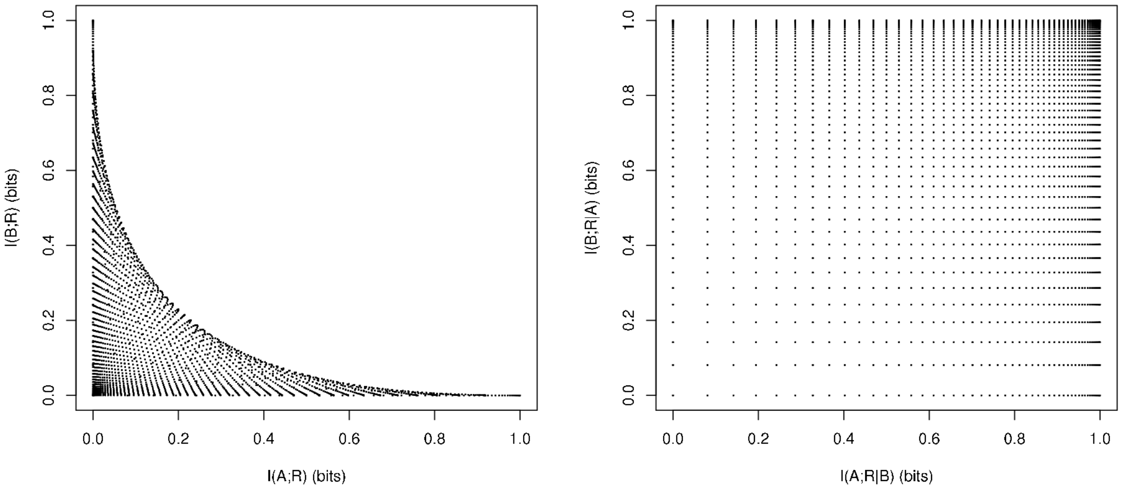

Figure 6.8: Numerical exploration of the space of input distributions for the XOR (2,2;2)-MAC. Left: capacity region from the sender's perspective. Right: capacity region from the receiver perspective. Discretization parameter is $N = 100$.

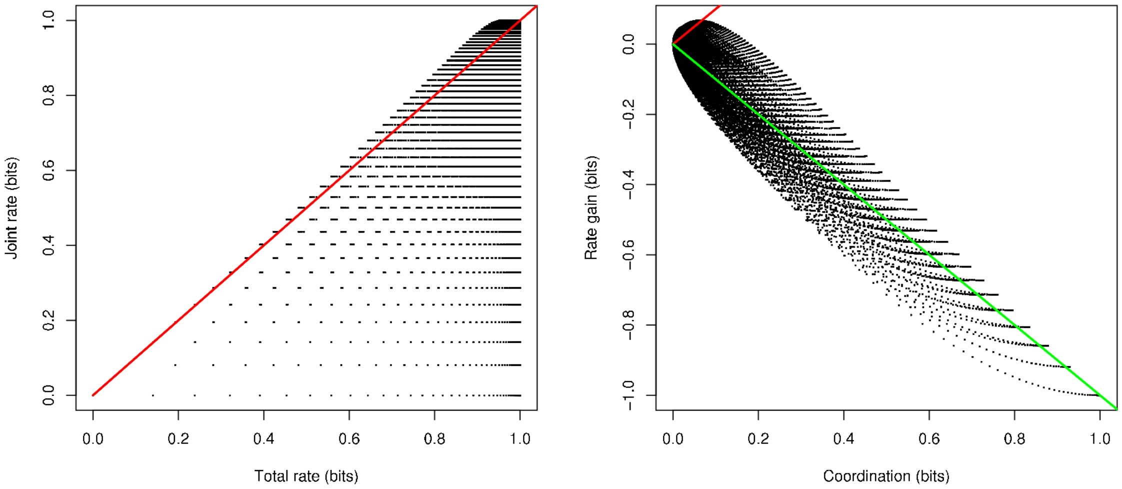

Figure 6.9: Numerical exploration of the space of input distributions for the XOR (2,2;2)-MAC. Left: amount of information sent through the channel versus that of the factorised input distribution. The red line has equation $y = x$. Right: rate gain (or loss) versus coordination. The red line has equation $y = x$, green line has equation $y = -x$. Discretization parameter is $N = 100$.

As explained before, the XOR MAC cannot be identified as an interfering channel by looking only at the capacity gain ΔC. This quantity would make us think that we are dealing with a compound of independent channels as with the binary symmetric MAC (see Experiment 6.5.1).

One way to identify this channel as an interfering one is to look at its synergy. For this purpose we first need to choose an input distribution for the channel. One such distribution is the total capacity-achieving distribution[2] $p^*_\perp(a,b)$ of the channel.

As has been explained previously, synergy measures the amount of information jointly transmitted through the channel that is not independently transmitted by any of the inputs. With independent inputs, having no synergy means that the inputs are transmitted to independent degrees of freedom of the output. On the other hand, a positive synergy means that the inputs interfere with each other in the channel.

We define the total synergy of the channel as the synergy the total capacity-achieving distribution mentioned before:

$$Syn_{total}\big(p(r|a,b)\big) = I_{p^*_\perp(a,b)}(A;B|R). \tag{6.12}$$

In the case of the binary symmetric MAC, the total capacity-achieving distribution is a uniform one. The corresponding total synergy is 0, meaning that the inputs end up in different degrees of freedom in the output, and therefore are not interfering. For the XOR MAC, uniform input distributions are also achieving total capacity, but the resulting total synergy is 1 bit, meaning the the inputs are interfering. However, it is not clear at this point whether this quantity is a definite criterion for distinguishing such channels.

6.5.4 Strongly-Interfering MACs

This class of channel is characterised by the fact the the joint capacity and the total capacity are different. This means that in order to achieve the maximum possible rate allowed by the channel the users *have* to coordinate their inputs.

A first example of such a channel is the binary erasure (2,2;3)-MAC. This channel has been identified in Experiment 6.4.1. It has the property that the capacity gain from cooperation $C - C_\perp$ is higher than the amount of coordination for reaching joint capacity. The binary erasure channel can be seen in the following way. If both inputs are the same, the output takes one of two distinct values depending on which input was sent. If the two input values differ, the output is a specific error symbol that is not related to the corresponding inputs. It is defined by the following conditional distribution:

[2] If several such distributions exist, then the one with maximum entropy $H(A,B)$ should be chosen.

$p(r\mid a,b)$	r	0	e	1
a	b			
0	0	1	0	0
0	1	0	1	0
1	0	0	1	0
1	1	0	0	1

Experiment 6.5.3. *Numerical investigation of the Binary Erasure (2,2;3)-MAC*

Objective: *Identify the relationship between amount of coordination and amount of information sent.*

Main results: *Coordination can improve the amount of information transmitted through the channel, and it is required for achieving joint capacity. Similarly to the XOR MAC the amount of coordination and the amount of transmitted information are related in a non-linear way.*

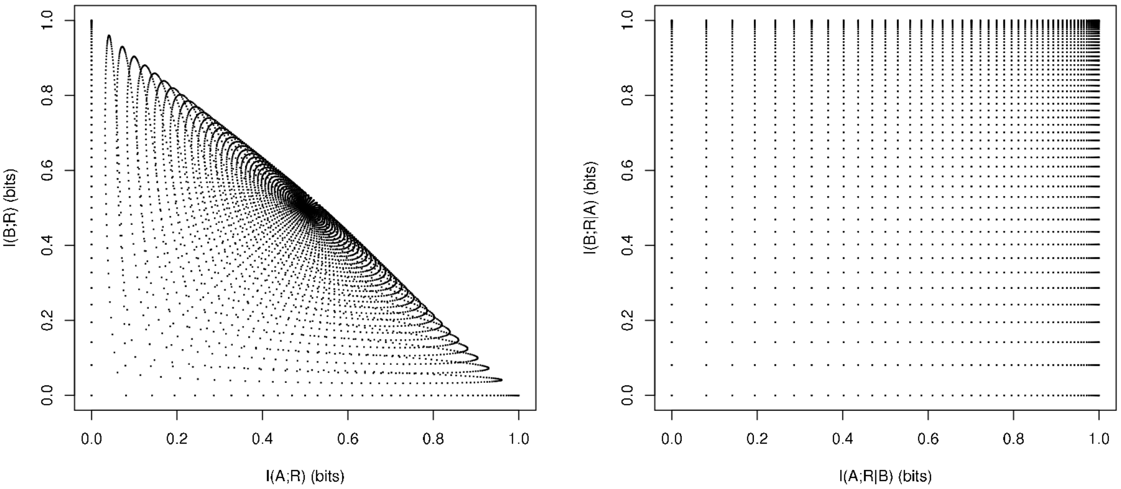

Figure 6.10: Numerical exploration of the space of input distributions for the binary erasure (2,2;3)-MAC. Left: capacity region from the sender's perspective. Right: capacity region from the receiver perspective. Discretization parameter is $N = 100$.

As one can see from the results on Fig. 6.11, similarly the XOR MAC, coordination can increase the amount of information sent through the channel. But a main difference is that joint capacity is achieved only with some coordination. This class of channel can be directly identified by the property that $C - C_{\perp} > 0$.

Another similarity with the XOR MAC is that the amount of coordination is not linearly correlated with the information gain, nor is it acting as a bound. An interesting

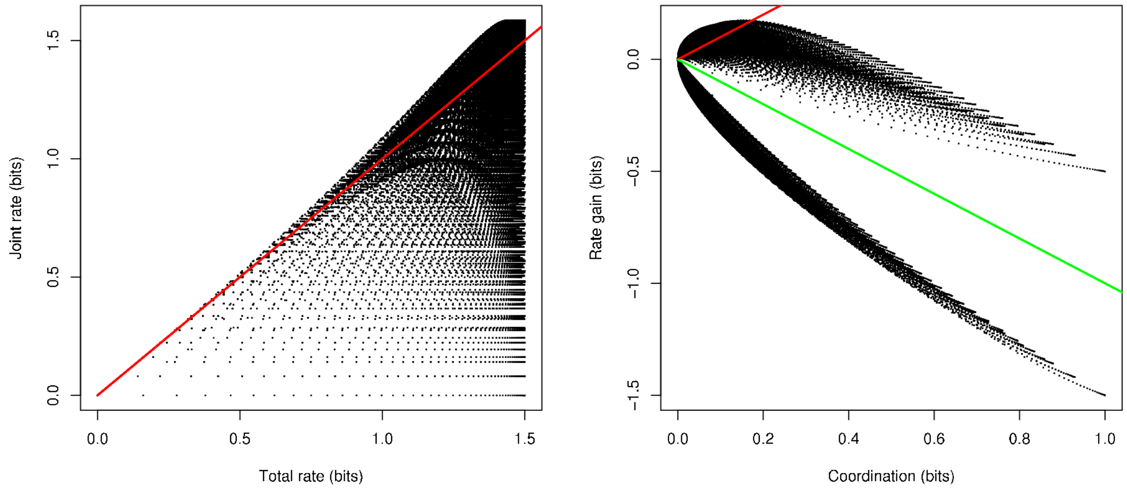

Figure 6.11: Numerical exploration of the space of input distributions for the binary erasure (2,2;3)-MAC. Left: amount of information sent through the channel versus that of the factorised input distribution. The red line has equation $y = x$. Right: information gain (or loss) g versus coordination. The red line has equation $y = x$, the green one has equation $y = -x$. Discretization parameter is $N = 50$.

point to notice is that when the coordination goes above 0.5 bits then that information gain is necessarily negative. This can be explained by the fact that going above 0.5 bits of coordination removes degrees of freedom in the input that are necessary to achieve total capacity. Indeed the maximum entropy for the joint input is 2 bits, and the total capacity is $C_{\perp} = 1.5$ bits. Therefore, if inputs are correlated by 0.5 bits, this leaves only 1.5 bits of entropy for the joint input. Further increasing the correlation would then reduce the available entropy, and consequently, the achievable capacity.

A second example of a strongly-interfering channel is a variant of the binary erasure channel which we refer to as the totalistic channel. The idea is that instead of having an error symbol, the output is a uniform distribution over the set of all allowed symbols. Basically if all the inputs are the same then the corresponding output symbol is emitted, otherwise the output is a uniform distribution. The totalistic (2,2;2)-MAC is described by the following conditional probability distribution:

$p(r\|a,b)$	r	0	1
a	b		
0	0	1	0
0	1	$\frac{1}{2}$	$\frac{1}{2}$
1	0	$\frac{1}{2}$	$\frac{1}{2}$
1	1	0	1

Experiment 6.5.4. ***Numerical investigation of the totalistic (2,2;2)-MAC***
Objective: *Identify the relationship between amount of coordination and amount of information sent.*
Main results: *Coordination can improve the amount of information transmitted through the channel, and it is required for achieving joint capacity. Similarly to the XOR MAC the amount of coordination and the amount of transmitted information are related in a non-linear way.*

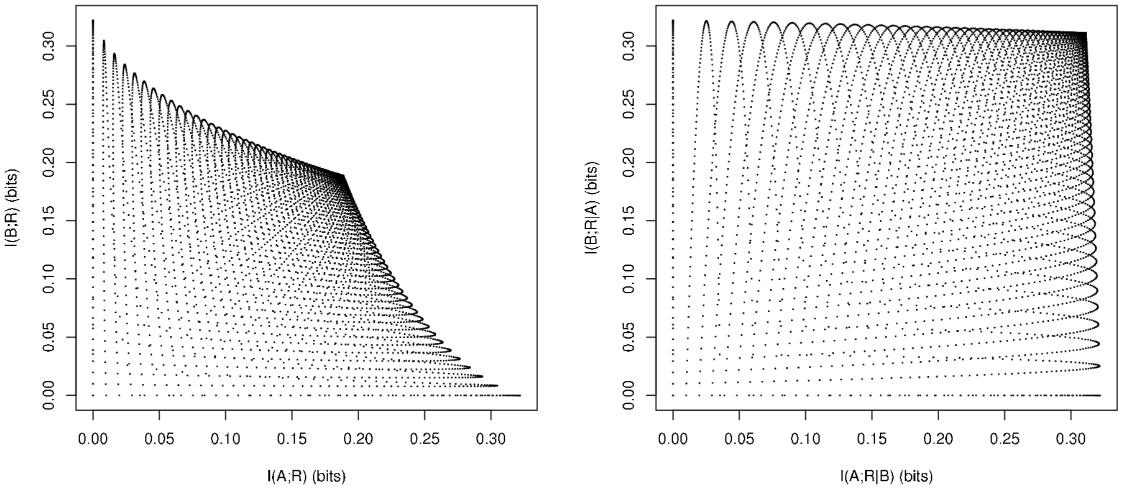

Figure 6.12: Numerical exploration of the space of input distributions for the totalistic (2,2;2)-MAC. Left: capacity region from the sender's perspective. Right: capacity region from the receiver perspective. Discretization parameter is $N = 100$.

Results are shown on Fig. 6.13. On the left graph one can see that correlated inputs can transmit more information than independent ones. In this case the maximum information gain is reached for the joint capacity, where correlation between inputs is maximal. The joint capacity for this channel is twice the total capacity, which is the theoretical maximum according to the conjecture presented earlier.

Again for this channel the amount of coordination is not linearly related to the information gain, nor is it a bound on it. An interesting aspect clearly visible in this channel is

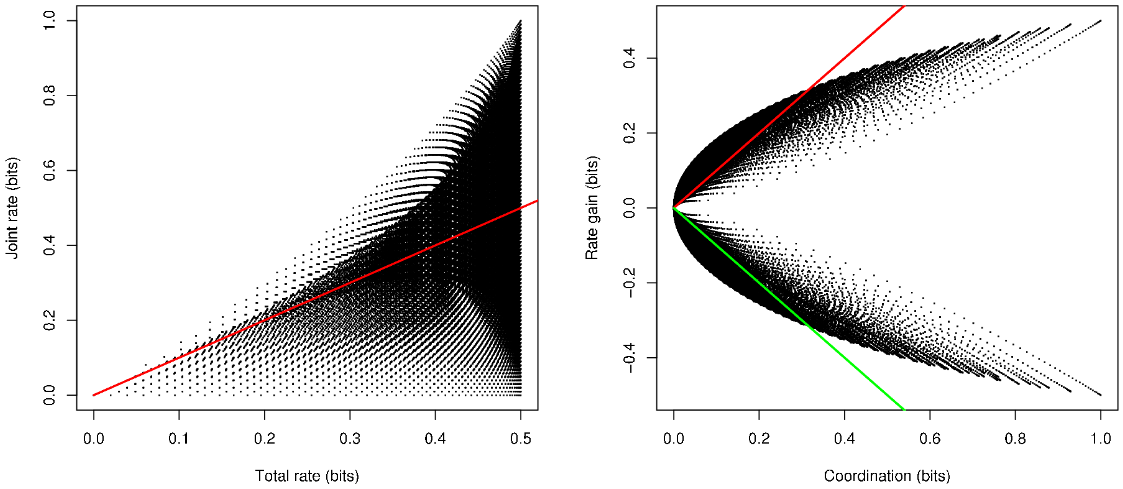

Figure 6.13: Numerical exploration of the space of input distributions for the totalistic (2,2;2)-MAC. Left: amount of information sent through the channel versus that of the factorised input distribution. The red line has equation $y = x$. Right: information gain (or loss) g versus coordination. The red line has equation $y = x$. Discretization parameter is $N = 50$.

the symmetry between information gain and information loss. Indeed for a given amount of correlation between inputs, the information gain can be either positive or negative, hinting at the fact that such correlation also has some kind of valence.

6.5.5 Synthesis of the Results

Different channels have been studied in the previous experiments which have specific properties. This allows us to propose a rough classification of two-users MACs based on some general quantities:

- *Non-interfering channels*: these are the most simple as they can be directly decomposed in two independent sub-channels. This is the case of the binary symmetric (2,2;4)-MAC presented in Experiment 6.5.1. They can be identified by the fact that their synergy $I(A; B|R)$ for a uniform input distribution is equal to 0. They also have the property that the channel capacity gain from coordination is equal to 0. And indeed, as we have seen in Experiment 6.5.1, coordinating the sources reduces the amount of information sent through the channel. This is due to the fact that correlation necessarily removes degrees of freedom for one or both sources. And as all the information contained in the sources is transmitted through the channel, this necessarily reduces the total amount of information transferred.

- *Weakly interfering channels*: such channels cannot be decomposed anymore into independent sub-channels, however the capacity can be achieved without any coordination of the sources. Put another way the capacity gain by coordination in these channels is equal to 0. Nevertheless, below capacity there are situations where coordination can improve the rate of information transmission compared to the corresponding independent sources. A typical example is the XOR MAC presented in Experiment 6.5.2. These can be identified by the fact that they have a non-zero capacity, and the synergy for a uniform distribution of inputs is also non-zero.
- *Strongly interfering channels*: typical examples of such channels are those presented in experiments 6.5.4 and 6.5.3. For these channels, joint capacity can only be achieved if the sources are coordinated. Therefore their capacity gain from coordination is non-zero.

More generally we have seen that the amount of coordination between the sources and the gain in transmitted information are related in a complex non-linear way. For examples there are cases for which adding a small amount of coordination can lead to a high increase in transmitted information. We also conjectured that the joint capacity can at most be twice the total capacity, and that this gain is not always bounded by the amount of coordination needed to achieve joint capacity.

6.6 Multiple-Access Channels of Embodied Agents

In the previous section, we investigated an idealistic model of the perception-action loop of two agents interacting with each other. This first step allowed us to distinguish different kinds of MACs and to understand the impact of coordination between agents.

However, compared to the full model of the intertwined perception-action loops, some of the simplifying assumptions seem quite unrealistic. This section investigates a more realistic model.

Also, the previous section used the terms coordination and correlation interchangeably for the channel sources. Although the correct technical term is correlation, the word coordination better reflects the fact that these sources are actually actions performed by different agents.

In this section we look at channels which are more directly inspired from the intertwined perception-action loops. This means that the two following assumptions are used:

- The agents have no direct access to the other agent's actions and the quantities considered are specific to each agent. This means that instead of looking at the overall information transmitted through the channel, we are now interested into how much each agent can send into the channel. Also, it is assumed that agents do not have direct access to the simultaneous action of the other agents.
- Coordination between agents does not come from 'nowhere', but has to be explicitly taken into account by causal mechanisms. In the previous section, arbitrary joint distributions of the agents' actions were investigated. We now assume that the agents are allowed to pick their actions according to a common source of information, but not according to each other's actions. Coordination occurs only indirectly. A crucial difference from the previous model is also that the shared information is not counted as part of the agent's transmitted information.

6.6.1 Common Information Source

We extend the minimal model of intertwined perception-action loops of Fig. 6.2 by introducing a new variable C which is a common source of information that the agents can use. In some way this variable represents the common past of the agents from which they can decide on what to do in the present. This model is shown on figure 6.14.

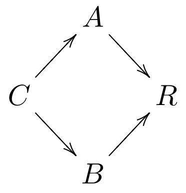

Figure 6.14: Channel with two agents and a common source of information.

The common information source is assumed to be uniform, providing the maximum possible amount of information. The two agents policies are described by the conditional probability distributions $p(a|c)$ and $p(b|c)$. The output of the motor channel does not depend on C and is fully described by $p(r|a, b)$. Therefore the common information source has two different but interrelated effects:

- It allows agents to coordinate their actions, leading to a potential coordination $I(A; B)$, but the shared information is not counted as information transmitted by the agents.
- The variable C acts as a context for both agents. In the case where one agent picks its actions according to C (and has an impact on the channel output), then the channel

appears to the other agents has having side-information, some of which is accessible through the context, potentially increasing its rate of information transmission.

In this setup, the overall amount of information sent through the channel can be decomposed in the following way:

$$I(A, B; R) = I(A; R|C) + I(B; R|C) + I(C; R) + I(A; B|C, R). \tag{6.13}$$

The two first terms can be interpreted as the amount of information that each agents sends independently of the other. The third term is the amount of information that the agents transmit together from the common source to the output of the channel. The last term can be understood as the synergy in this channel. More precisely it measures how much synergy there is between the output of the channel and the actions of the agents when all the correlations coming from the common source are discarded.

The previous section has shown that this total amount of information can be increased by using coordination in weakly and strongly interfering channels. However what is not known yet is whether this increase also applies to the information that each agent sends through the channel independently of the other agent, i.e. $I(A; R|C)$ and $I(B; R|C)$. In order to answer this question we numerically investigate the capacity regions of embodied agents with a common source of information.

The general methodology is to sample the space of agent policies $p(a|c)$ and $p(b|c)$ using the same discretization as in Exp. 6.3.1. The number of symbols for C is fixed to the maximum number of input symbols, and its distribution is assumed to be uniform (hence providing the maximum amount of information). The capacity region for policies independent from C and policies that may depend on C are computed and compared. The difference between the two regions shows situations for which coordination improves the accessible rates. For these situations, we also try to identify which components of the overall information sent are changed in order to understand the mechanism underlying the rate increase.

The following subsections investigate the three interfering channels that have been presented as examples of the different classes, i.e. the XOR, the binary erasure, and the totalistic MAC.

6.6.2 The XOR (2,2;2)-MAC

Experiment 6.6.1. *Numerical investigation of the XOR (2,2;2)-MAC with common information source*
Objective: *Identify situations where coordination from a common source increases the amount of information transmitted by each agent.*
Main results: *Coordination extends the capacity region of the channel, allowing both agents to transmit at equal rates while achieving channel capacity. Information transmitted from the common source to the output of the channel and synergy play no role in achieving the capacity.*

As one can see from the results on Fig. 6.15, coordination through the source allows the agents to access the centre part of the capacity region, where agents can at most reach a rate of 0.5 bits at the same time.

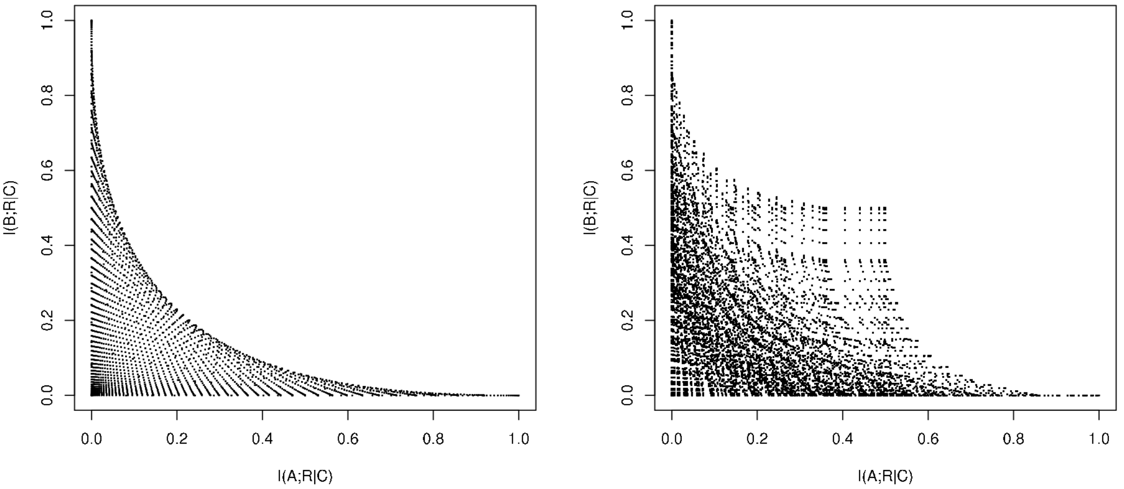

Figure 6.15: Numerical exploration of the space of input distributions conditioned on a common information source for the XOR (2,2;2)-MAC. **Left:** capacity region for policies independent of C. Two maxima are visible, for 0 coordination, and for $I(A;B) \approx 0.12$. Discretization parameter is $N = 100$. **Right:** capacity region for policies that can depend on C. Discretization parameter is $N = 20$.

Such a rate can be easily understood in terms of time sharing. Basically each time the common source sends a zero, agent A sends a fixed symbol and agent B can send information through the channel with a rate of one bit per step. When the common source sends a one, the opposite behaviour occurs, i.e. B sends a fixed symbol and A transmits

information. Because the common source follows a uniform distribution, A can send information only half of the time and B the other half.
All the region accessible only with coordination can be understood in terms of this time-sharing mechanism and suboptimal versions of it.

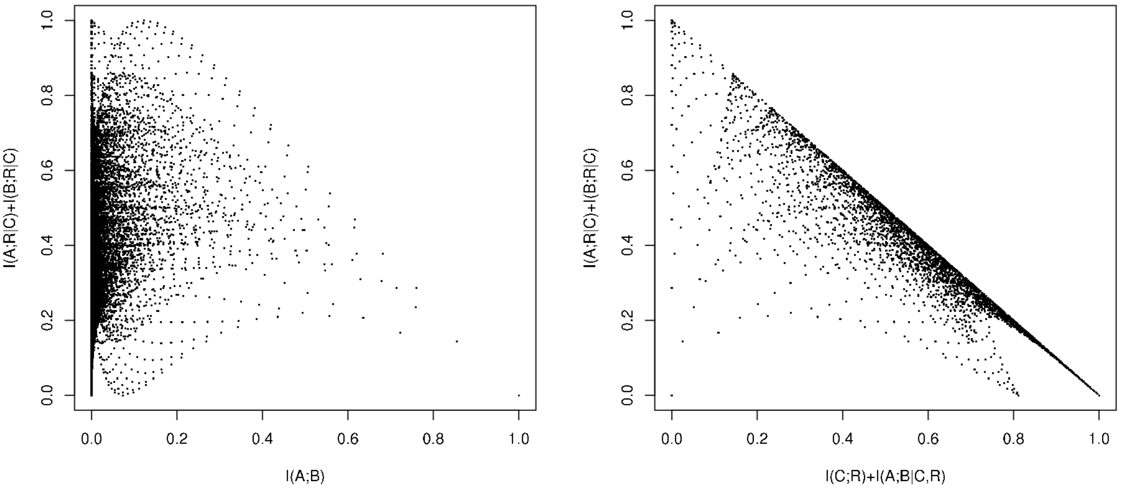

Figure 6.16: Numerical exploration of the space of input distributions conditioned on a common information source for the XOR (2,2;2)-MAC. Left: sum of independently transmitted information versus coordination. Right: sum of independently transmitted information versus remaining part of the overall transmitted information. Discretization parameter is $N = 20$.

If we now consider the coordination between the two agents $I(A;B)$ and the sum of independently transmitted information (Fig. 6.16 left) one can see that there are two levels of coordination for which the maximum is reached:

- When there is no coordination: it is the case for example when one agent sends no information and the other agent sends at maximum rate.
- For low coordination $I(A;B) \approx 0.12$: this is the time-sharing situation mentioned above.

6.6.3 The Binary Erasure (2,2;3)-MAC

Experiment 6.6.2. ***Numerical investigation of the binary erasure (2,2;3)-MAC with common information source***

Objective: *Identify situations where coordination from a common source increases the amount of information transmitted by each agent.*

Main results: *Coordination does not extend the capacity region. Transmitting at maximum rate simultaneously for both agents creates synergy in the channel.*

As depicted on Fig. 6.17, and unlike in the XOR MAC, coordination does not extend the capacity region of the individual agents. This may sound surprising because the capacity of the binary erasure MAC is increased through coordination. Hence it means that the increase of transmitted information is not due to the sources independently sending more information, but to either the synergy being increased or the common information being transmitted.

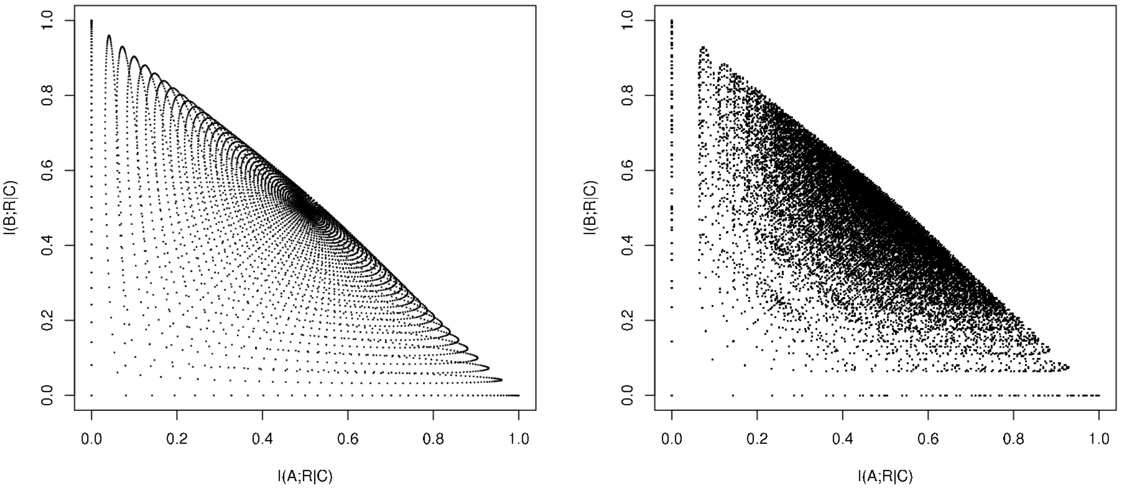

Figure 6.17: Numerical exploration of the space of input distributions conditioned on a common information source for the binary erasure (2,2;3)-MAC. Left: capacity region for policies independent of C. Discretization parameter is $N = 100$. Right: capacity region for policies that can depend on C. Discretization parameter is $N = 20$.

Differently from the XOR MAC, in this channel the maximum amount of transmitted information by both agents (i.e. the sum of their independent contributions) is reached only at the centre of the capacity region, i.e. when both agents are sending information at the same rate (which is actually $I(A;R|C) = I(B;R|C) \approx 0.52$ bits). This maximum can be reached with or without coordination.

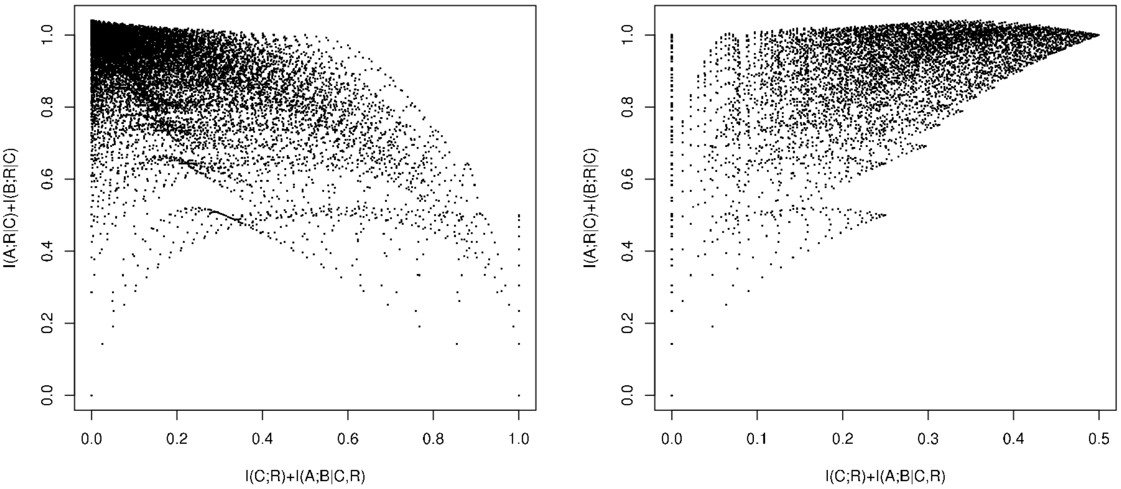

Figure 6.18: Numerical exploration of the space of input distributions conditioned on a common information source for the binary erasure (2,2;3)-MAC. Left: sum of independently transmitted information versus information transmitted from the common source to the output. Right: sum of independently transmitted information versus synergy. Discretization parameter is $N = 20$.

The coordination versus sum of independently transmitted information graph (not depicted here) shows that, similarly to the XOR MAC, the maximum can be reached either without coordination or with a low amount of coordination.

A more interesting result is shown on the two graphs of Fig. 6.18. These show the relation between the sum of independently transmitted information with both the synergy and the amount of information transmitted from the common source. As can be seen on the graphs, the maximum is reached when no information is transmitted from the source. This makes sense because, as the source information is not counted in the agents' own rates, transmitting it reduces the bandwidth available for sending the agent's own information. Another interesting aspect is that the more information is sent through the channel, the higher the synergy. This means that when agents try to send high amounts of information, more interferences start to appear in the channel because they have to 'squeeze' their information into the available degrees of freedom of the output.

6.6.4 The Totalistic (2,2;2)-MAC

Experiment 6.6.3. ***Numerical investigation of the totalistic (2,2;2)-MAC with common information source***
Objective: *Identify situations where coordination from a common source increases the amount of information transmitted by each agent.*
Main results: *Coordination extends the capacity region of this channel, but the maximum sum-rate point is still achievable without coordination.*

Results of this experiment are shown on Fig. 6.19. One can see that the capacity region is extended when coordination is allowed, however the maximum sum rate is the same as in the non-coordinated case (when both agents send at $I(A;R|C) = I(B;R|C) \approx 0.19$ bits).

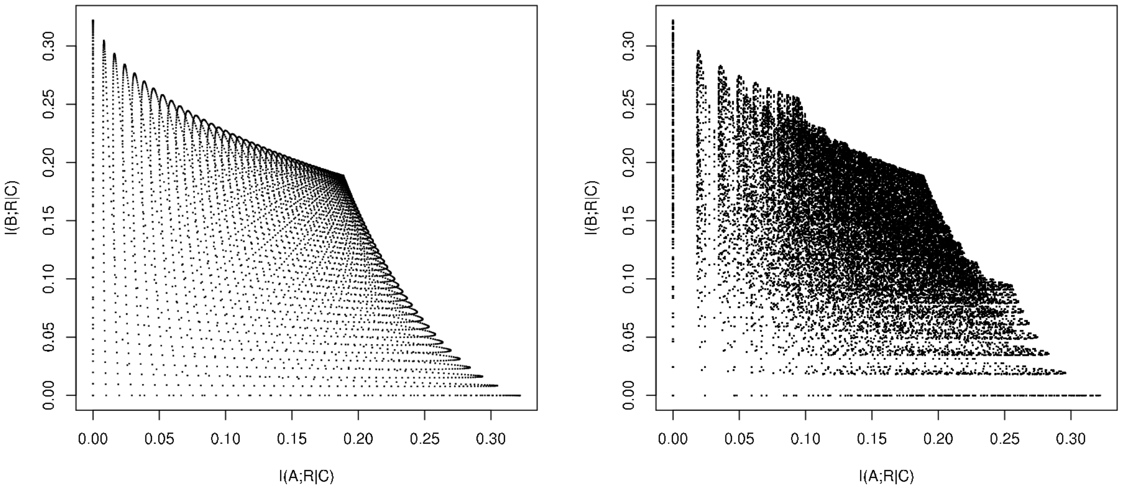

Figure 6.19: Numerical exploration of the space of input distributions conditioned on a common information source for the totalistic (2,2;2)-MAC. Left: capacity region for policies independent of C. Discretization parameter is $N = 100$. Right: capacity region for policies that can depend on C. Discretization parameter is $N = 20$.

Considering the sum of transmitted informations versus coordination graph (Fig. 6.20 left) one can see that the maximum sum rate is reachable only when there is no coordination between the actions. However a low level of coordination allows to reach suboptimal rate pairs that are not accessible to non-cooperating agents.

On the second graph (Fig. 6.20 right), one can see that the maximum sum rate is achievable only when having either synergy or information transmitted from the common source.

Graphs not presented here show that the maximum is reached when there is no in-

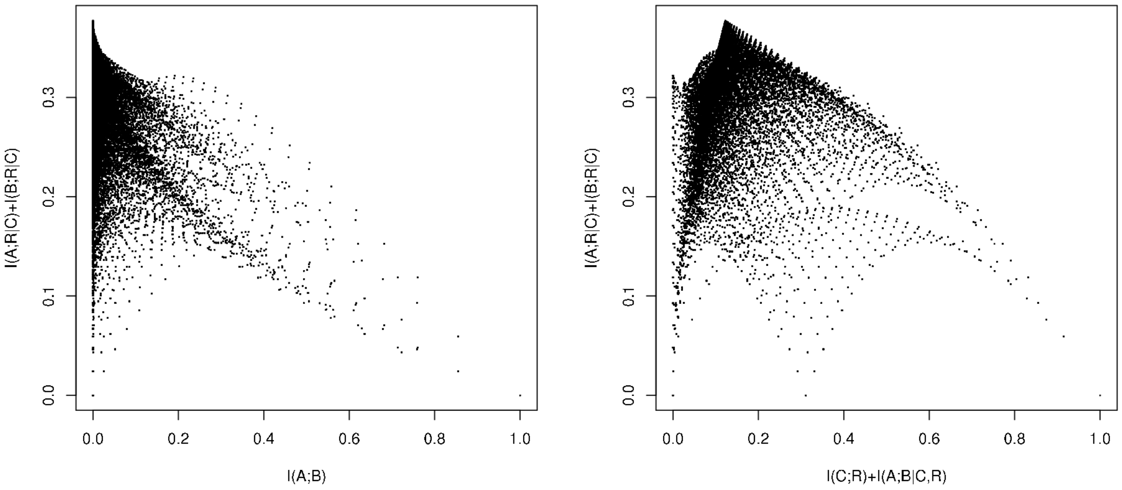

Figure 6.20: Numerical exploration of the space of input distributions conditioned on a common information source for the totalistic (2,2;2)-MAC. Left: sum of independently transmitted information versus coordination. Right: sum of independently transmitted information versus remaining part of the overall transmitted information. Discretization parameter is $N = 20$.

formation transmitted from the common source but when maximum synergy is reached. This confirms the interpretation from the binary erase MAC that transmitting information from the common source reduces the bandwidth available for the agents' actions, and that increased transmission rates of the agents generate interferences in the channel.

6.6.5 Synthesis of the Results

This set of experiments showed that even though coordination can increase the overall amount of information transmittable through a MAC, this does not mean necessarily that it will lead to an increase of the amount of information independently transmitted by agents relying on a common information source (which is the actual situation that embodied agents are facing with intertwined perception-action loops).

Moreover, it seems that the maximum amount of information sent independently by each agent cannot be more than what they are maximally able to send when they are not coordinated. At the same time, it appears that the sum-rate in the non-coordinated case is an upper-limit on the sum-rate when coordination is allowed. This can be explained by the fact that the non-coordinated sum-rate $I(A, B; R)$ and sum-rate for agents coordinating through common source $I(A, B; R|C)$ indeed have the same maxima. However, in some situations (which are below the maximum rates), new rate pairs are accessible to the agents if they coordinated their actions. These cases can be seen as the extension of the capacity region when coordination is allowed, and can be understood through the time-sharing mechanism described above.

A last remark can be made about what are the contributions to the capacity increase with coordination. As one can see, it is not the increase of the independent information flows that contribute to the capacity increase; indeed, their maximum sum-rate is the same as in the non-coordinated situation. Instead the contributing parts are the synergy and the information sent from the common source to the output of the channel, with the latter apparently being the highest contributor. These are therefore not useful for single agents, however if instead of considering two distinct agents one sees this model as two actuators of the same agent, it means that coordinating actuators does increase the capacity of the channel, because in this case both the synergy part and the common part are included in the transmitted information.

6.7 Empowerment Maximization in MACs with Memory

The previous sections investigated the effect of coordination between two agents in memoryless multiple-access channels. It was shown that, although coordination can increase the empowerment of both agents taken together, their individual empowerment could not be directly increased by coordination. However, if one considers channels that have memory, this might not be valid anymore. Indeed, coordination could be used to improve control over the state of the channel and therefore allow the agents to reach states of higher empowerment.

In this section, we look at such channels. The two first channels studied are minimal models that instantiate competitive and collaborative scenarios similar to the bathroom and the well problems described in the previous sections. The main difference is that agents are allowed to perform actions at the same time, leading to potential interferences. Using tools from game theory, namely Nash equilibria, we identify the behaviours that should arise if both agents try to maximize their own empowerment. A third example is then presented, the mountain-climbing scenario, that is intended to be less abstract that the two others. Because of its complexity, game theoretic tools cannot be directly used. Instead, empowerment maximizing behaviours are identified using simulated annealing.

6.7.1 Competitive Scenario

The MAC used in this scenario has two possible states for which the empowerment of each agent is different. The corresponding causal Bayesian graph is depicted on Fig. 6.21.

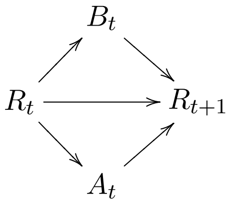

Figure 6.21: Causal Bayesian graph in the competitive scenario. The environment has memory. Both agents perceive the full state of the environment and are allowed to pick their actions according to this state.

We will call r_a the state in which agent A has a high empowerment and B has a low one. State r_b is exactly symmetric to a. Both agents perceive the exact state of the environment and are allowed to pick their actions accordingly. Actions are defined as either 0 or 1. The MAC is described by two conditional distributions. The distribution when the environment is in state r_a is the following:

$p(r_{t+1}\|a, b, R_t = r_a)$	r_{t+1}	r_a	r_b
a	b		
0	0	0.3	0.7
0	1	0.2	0.8
1	0	0.7	0.3
1	1	0.8	0.2

When the channel is in state r_b, the conditional distribution is the following:

$\mathbf{p(r_{t+1}\|a, b, R_t = r_b)}$	r_{t+1}	r_a	r_b
a	b		
0	0	0.2	0.8
0	1	0.8	0.2
1	0	0.3	0.7
1	1	0.7	0.3

As one can see from the distributions, when the channel is in state r_a then agent A has almost all the control on the future state, B's actions can only slightly interfere with the outcome of the channel. The converse applies to state r_b.

We now assume that both agents try to maximize their own empowerment. If agent B has a fixed policy $p(b_t|r_t)$, then agent A should pick a policy $p(a_t|r_t)$ that maximize its own empowerment $\mathfrak{E}(A_t \rightarrow R_{t+1}|R_t)$. Similarly, if the policy of agent A is fixed, then B should pick a policy that maximizes empowerment. If we assume that this process is repeated for a long time, one can wonder whether stable solutions can be found.

In order to answer this question, we use the concept of *Nash equilibrium.* A Nash equilibrium is basically a point in the space of policies for which no agent can improve its situation by unilateraly changing its policy. Such a point can be different from the global equilibrium. Let us consider for example the prisoner's dilemma (Axelrod and Hamilton 1981). The problem is described as follows. Two suspects A and B are arrested for a crime and interrogated separately by the police. Each player has two possible strategies: either it remains silent, or it accuses the other player. Depending on their strategies, the prisoners may serve some time in jail. The outcomes are the following. If both players remain silent they both serve 1 year in jail. On the other hand, if each prisoner charges the other one, then they both serve 5 years. However, if one of the prisoners remains silent while the other accuses him, then the accused one gets a 10 years sentence while the accusing one is immediately released. These outcomes are summarized in the following payoff matrix:

	B stays silent	B accuses A
A stays silent	$-1, -1$	$-10, 0$
A accuses B	$0, -10$	$-5, -5$

In the payoff matrix, each cell contains the payoff for A on the left and for B on the right. The general goal of the game is to maximize the payoff (this is why years in jail are negative). One can see that there seems to be a 'global maximum' to the payoff matrix, when both agents remain silent. However, in this specific situation, any of the two prisoners would be better off accusing the other one. Indeed, instead of serving 1 year in jail, he would be directly released. Suppose that B accuses A. In this case, instead of serving 10 years in jail, A would also be better off accusing B, resulting in a 5 year sentence for both of them.

The concept of Nash equilibrium captures this aspect of the game. The 'global maximum' payoff $-1, -1$ is not an equilibrium point because each agent has an interest in unilaterally changing its strategy. Similarly, the asymmetric situations with payoffs $-10, 0$ and $0, -10$ are not equilibrium points; the accused agent is better off changing its strategy to accusing. Therefore, we are left with the sub-optimal situation $-5, -5$ from which no agent can unilaterally improve its payoff (even though they could both improve their payoffs by simultaneously changing their strategies). Hence, the $-5, -5$ situation is a Nash equilibrium of this game.

Now coming back to our problem of interest, the following methodology is used in order to identify Nash equilibria:

- We assume an uniform initial distribution $p(r_0)$ over the states of the channel.
- The space of possible pairs of policies $p(a_t|r_t)$ and $p(b_t|r_t)$ is sampled by discretizing this space similarly to experiment 6.3.1.
- For each pair, the stationary distribution over the states $p(r_t)$ is computed according to the technique described in appendix C.
- The resulting empowerment $\mathfrak{E}(A_t \to R_{t+1}|R_t)$ and $\mathfrak{E}(B_t \to R_{t+1}|R_t)$ are computed and used to construct the payoff matrix.
- The coordination between agents $I(A_t; B_t)$ is computed.
- Nash equilibria are identified from the payoff matrix.

Experiment 6.7.1. ***Empowerment maximization in a competitive scenario***
Objective: *Identify Nash equilibria in the space of possible policies for two agents in an empowerment-competitive scenario.*
Main results: *Although many good solutions are available, the only equilibrium point is strongly sub-optimal and involves no coordination.*

As one can see on Fig. 6.22, the competitive nature of this scenario is characterized by the shape of the accessible empowerment pairs region. In order for one agent to get high empowerment, the empowerment of the other agent has to be reduced. The Nash equilibrium identified brings the same empowerment to both agents, i.e. approximately 0.06 bits. Interestingly, this point is very far from the maximum possible value, which is approximately 0.15 bits of empowerment for each agent. One can also see on Fig. 6.23 that the equilibrium point does not involve coordination.

These results seem to indicate three different properties of competitive scenarios:

- Empowerment maximizing agents seem to converge towards a Nash equilibrium where the empowerment is ‘shared’ between the agents.
- Solutions obtained are far from the maximum possible values. This is consistent with the results obtained with multiple agents in experiments 5.7.1 and 5.8.1. Indeed, it

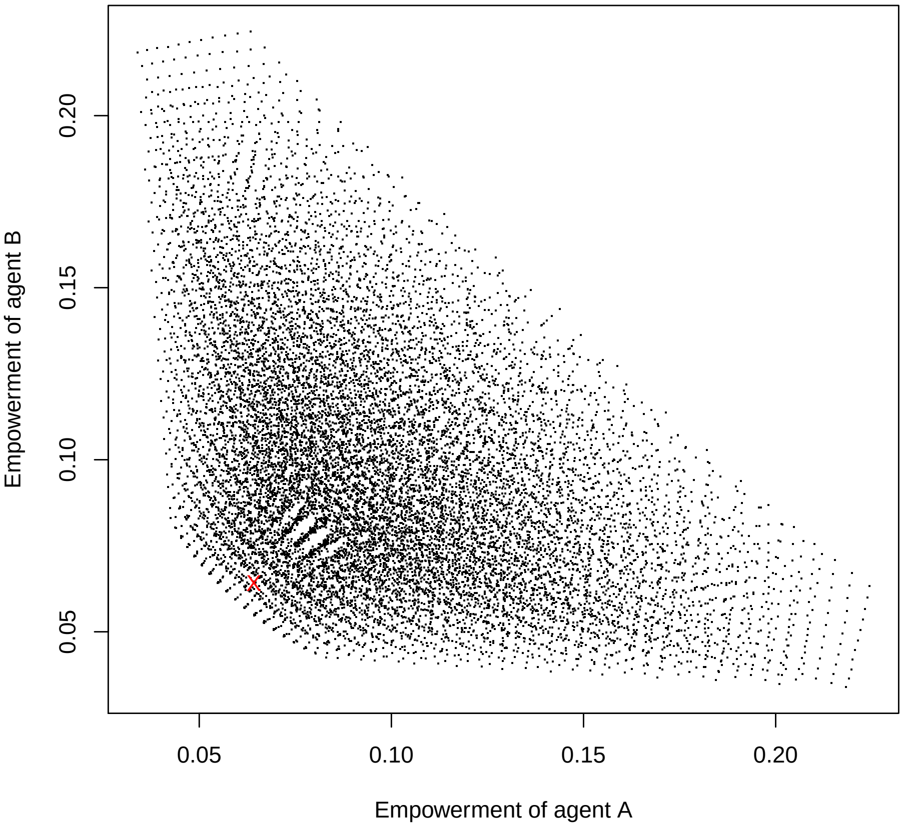

Figure 6.22: Space of accessible empowerment pairs in the competitive scenario. Nash equilibrium is represented as a red cross on the bottom left of the graph.

was shown that values obtained from agents selfishly maximizing their own empowerment were far below that of agents whose behaviour was evolved so as to maximize global empowerment.

- Competitive environments seem to lead to solutions in which the agents are not coordinating their actions.

From a game-theoretic perspective, the competitive scenario is close to a zero-sum game. Indeed, the empowerment depends mostly on the distribution over the states of the channel $p(r_t)$, therefore each agent tries to push this distributions towards its preferred state, at the expense of the other agent. As a result it is likely that any such competitive

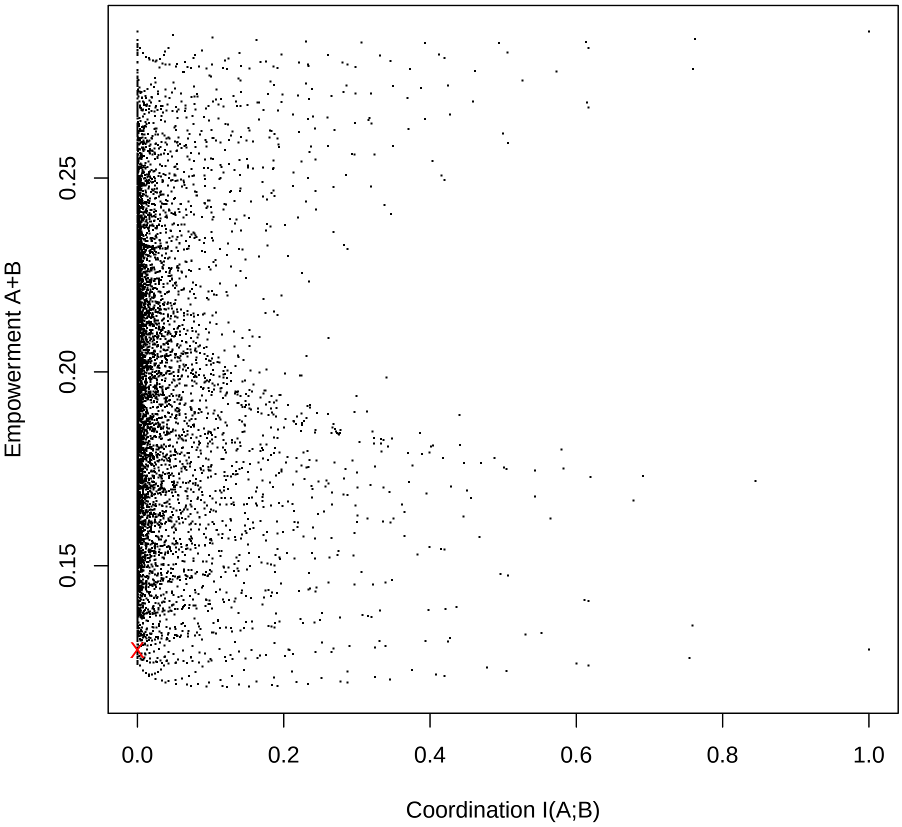

Figure 6.23: Amount of coordination $I(A_t;B_t)$ versus sum of both agents' empowerment in the competitive scenario. Nash equilibrium is represented as a red cross on the bottom left of the graph.

scenario would result have an equilibrium with shared empowerment.

On the other hand, results about coordination between the agents may not be an actual consequence of competitive scenarios but an artifact of the specific channel under consideration. Even though it apparently makes sense from a game-theoretic perspective that agents do not coordinate their actions, it is actually not the case here.

In standard game theory coordinating actions means that at least one agent has some knowledge about what the other agent's action is going to be, therefore making it possible to exploit this information to its advantage. A typical example is the famous rock-paper-scissors game. If your moves can be predicted, the opponent can very easily beat you by picking its moves according to what you are going to play. Therefore behaving in an unpredictable way is a good strategy.

However this actually applies to what is referred to as *mixed-strategies* (i.e. where each agent behaves according to a distribution over the pure strategies which are the elementary moves). Indeed, it is knowledge about which strategy is actually employed by the agent among those of its distribution that can be exploited by the agent. In our setup, we are only looking at pure strategies, but as these strategies have a probabilistic nature, they can generate coordination between actions *without* coordination between strategies. Therefore the standard game theoretic reasoning does not apply here.

The Nash equilibrium found in this scenario involves the following strategies:

- In state r_a, agent A tries to stay in this state by performing action 1. Meanwhile, agent B tries to increase the odds in favour of moving to state r_b by performing action 0.
- In state r_b, the opposite happens. Agent B tries to stay in state r_b by performing action 0 while agent A tries to change the odds in favour of moving back to state r_a and for this it performs action 1.

At the end of the day, agent A always performs action 1 while agent B always selects action 0. From an information-theoretic perspective this means that their actions are not coordinated. However, the channel could be changed in such a way that its structure is preserved but the actions picked in state r_b would have to be different (i.e. agent A performing action 0 and agent B performing action 1). In this case, actions would then be coordinated.

6.7.2 Collaborative Scenario

Similarly to the previous experiment, the MAC used in this scenario has two states. However, one of these states gives a high empowerment to both agents, and the other state gives a low empowerment to both agents. We will refer to these states as r_{high} and r_{low} respectively. Again, both agents perceive the exact state of the environment and are allowed to pick their actions accordingly in the set $\{0; 1\}$. The conditional distribution of the MAC in the state r_{high} is the following:

$\mathbf{p(r_{t+1}\|a,b,R_t = r_{high})}$	r_{t+1}	r_{high}	r_{low}
a	b		
0	0	0.8	0.2
0	1	0.0	1.0
1	0	0.0	1.0
1	1	0.5	0.5

When the channel is in state r_{low}, the conditional distribution is the following:

$\mathbf{p(r_{t+1}\|a,b,R_t = r_{low})}$	r_{t+1}	r_{high}	r_{low}
a	b		
0	0	0.5	0.5
0	1	0.5	0.5
1	0	0.5	0.5
1	1	0.8	0.2

We use the same methodology as in the previous experiment.

Experiment 6.7.2. ***Empowerment maximization in a collaborative scenario***
Objective: *Identify Nash equilibria in the space of possible policies for two agents in an empowerment-collaborative scenario.*
Main results: *Best solutions are reaching the maximum possible empowerment for both agents, and involve partial coordination between the agents.*

The space of accessible empowerment pairs depicted on Fig. 6.24 shows very clearly the collaborative nature of this scenario. Highest empowerment values for one agent are correlated with high empowerment values for the other agent. Four different Nash equilibria have been identified in this setup. For all of them both agents have the same empowerment. By looking at Fig. 6.25, one can see that two of these equilibria (empowerment of 0.23 and 0.44 bits per agent) require no coordination between the agents. The equilibrium point with lowest empowerment (0.16 bits) uses maximum coordination. On the other hand, the best solution (0.50 bits of empowerment per agent) requires a relatively high level of coordination (0.72 bits).

These results indicate the following properties of collaborative scenarios:

- Similarly as in the competitive scenario, empowerment maximizing agents seem to converge towards solutions where the empowerment is ‘shared’ between the agents.
- Best solutions reach the maximum possible empowerment for both agents.

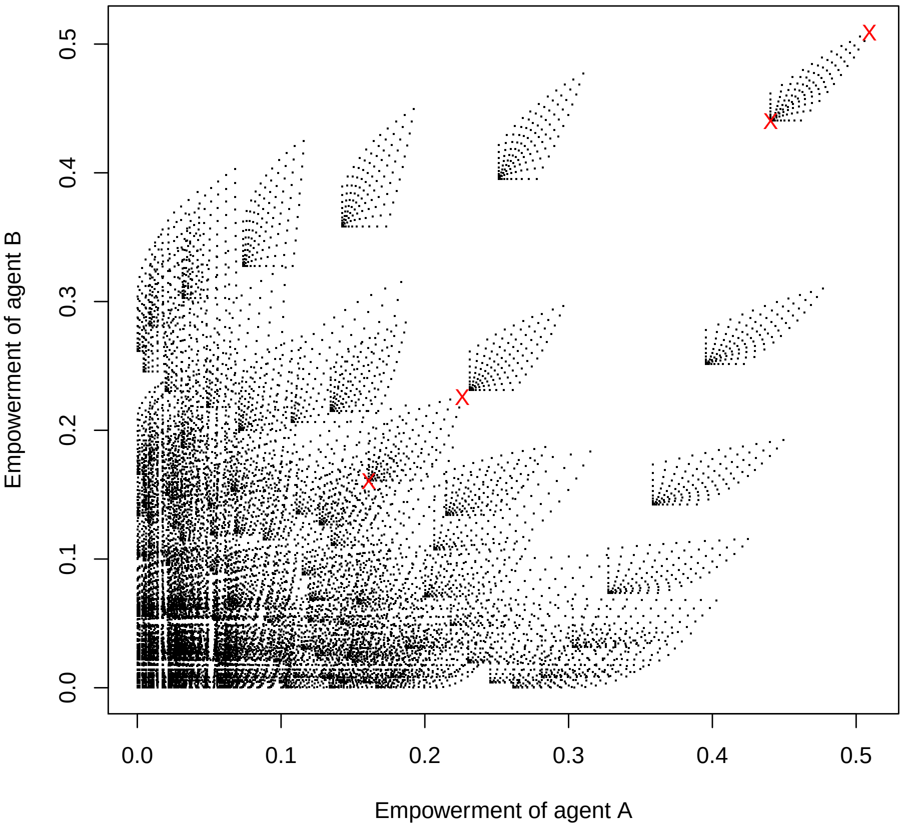

Figure 6.24: Space of accessible empowerment pairs in the collaborative scenario. Nash equilibria are represented as red crosses.

- Reaching maximum empowerment requires a high amount of coordination. However slightly suboptimal solutions can be reached without any coordination.

Similarly to the competitive scenario, it turns out that the amount of coordination is actually a consequence of the specific channel studied. Indeed, the optimal strategy is the following:

- In state r_{low} both agents perform action 1 in order to move towards state r_{high}.
- In state r_{high} both agents perform action 0 in order to stay in this state.

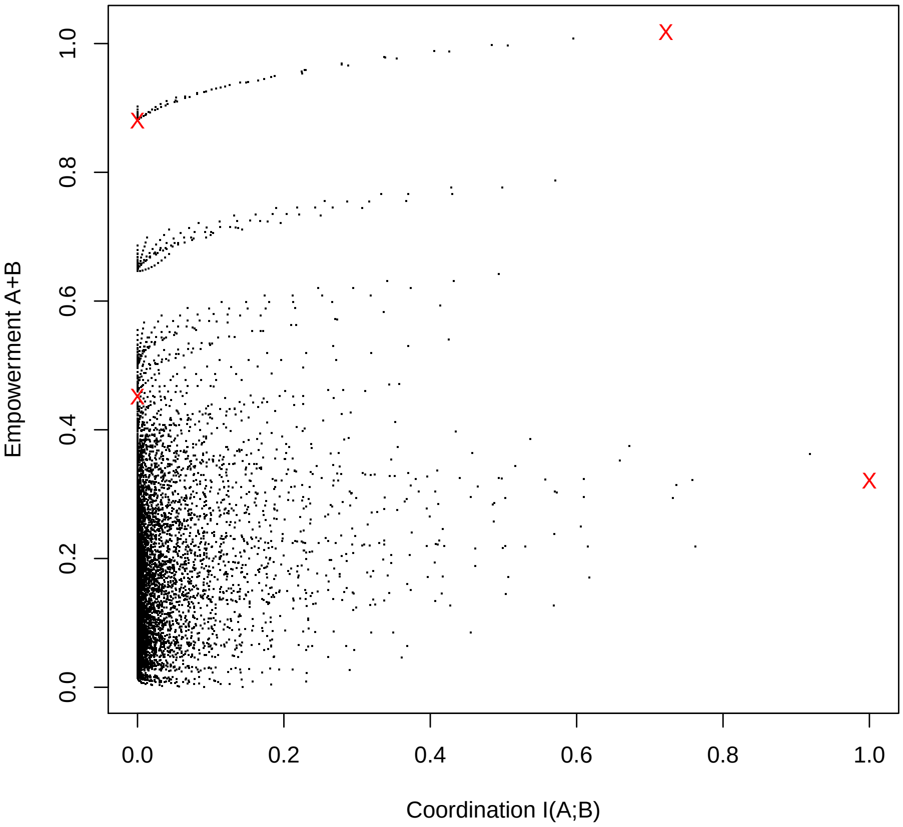

Figure 6.25: Amount of coordination $I(A_t; B_t)$ versus sum of both agents' empowerment in the collaborative scenario. Nash equilibria are represented as red crosses.

In the end, when agent A performs action 0 agent B also does, and the same is true for action 1. If the channel had been slightly different (but structurally similar), for instance if the action pair $1, 1$ had been the optimal action to stay in state r_{high} then no coordination would have been needed in order to achieve maximum empowerment.

However, the ability to reach optimal empowerment would be preserved. This property is therefore likely to be a general consequence of collaborative scenarios.

6.7.3 The Mountain Climbing Scenario

This section introduces a more realistic scenario, in which two agents are climbing on a mountain. The key aspects of this scenario are the following:

- The mountain is a multiple-access channel with memory.
- The state of the channel is defined by the combined vertical positions of both agents, which can range from 0 (ground) to 5 (top). For instance, the state $(0, 0)$ means that both agents are on the ground, while state $(5, 0)$ means that agent A is at the top of the mountain and agent B on the ground.
- The agents have 8 different actions: staying at the current level, climb up one level, climb down 1 to 5 levels (the lowest reachable being the ground state 0), or help the other agent to climb up.

Because being at the top allows the agent to directly climb down to any level, this makes it a state of high empowerment (decreasing towards the bottom). The interesting part is that the probabilities make it very difficult for one agent to get to the top on its own. Helping each other out makes it easier, and therefore it is expected that agents trying to maximize their empowerment should collaborate to climb the mountain. Moreover, helping the other agent can only be done when the helping agent is just above the agent being helped. Therefore, the optimal strategy to get to the top is to climb one step, then help the other agent into reaching the same level, then climbing another step, and so on. The goal of this section is to identify whether empowerment maximizing principles would lead such agents to find the optimal strategy. Unfortunately, because the space of possible strategies is relatively large, using the game-theoretic approach presented in the previous sections is out of the question. Instead, the space of strategies is searched using a simulated annealing algorithm.

Similarly to the previous experiments, the methodology used for evaluating the efficiency of a candidate solution is the following:

- We assume an uniform initial distribution $p(r_0)$ over the states of the channel.
- The stationary distribution over the states $p(r_t)$ is computed according to the technique described in appendix C.
- The resulting empowerments $\mathfrak{E}(A_t \to R_{t+1}|R_t)$ and $\mathfrak{E}(B_t \to R_{t+1}|R_t)$ are computed.
- The coordination between agents $I(A_t; B_t)$ is computed.

The following probabilities were used for the channel:

- Probability of falling down one level when trying to stay at the current one: 0.2.
- When climbing without help: probability of success: 0.1, probability of falling down one level: 0.3.
- When climbing with help: probability of success: 0.7, probability of falling down one level: 0.3.

Experiment 6.7.3. *Empowerment maximization in the mountain-climbing scenario*
Objective: *Identify whether selfish or global empowerment maximization can lead to coordinated behaviour in a complex scenario.*
Main results: *Selfish maximization of empowerment leads to suboptimal solutions without coordination. Maximizing the combined empowerment of both agents leads to optimal solutions and coordinated behaviour.*

The parameters used for the simulated annealing were $T_{start} = 0.001$, $T_{stop} = 0.00001$, $T_{dec} = 0.000001$. Two different searches were performed: a selfish search, where each agent's policy is improved alternatively for improving its own empowerment, and a global search during which the two policies are evolved at the same time for maximizing the sum of both agent's empowerment. The baseline empowerment for a random policy is approximately 0.05 bits per agent.

The selfish search led to a cyclic evolution between various solutions. For example, two typical solutions found with this search were:

- A suboptimal 0.66 bits of empowerment per agent with 0 bits of coordination.
- A better but still suboptimal solution of 1.10 bits of empowerment per agent with 0.34 bits of coordination.

On the other hand, simultaneous maximization of both agents' empowerment converged most of the time to a solution of 1.55 bits of empowerment per agent with a coordination of 1.38 bits.

One can see from the results that selfish maximization in this scenario led to various suboptimal solutions using limited amount of coordination. Comparatively, global maximization converges to highly empowered solutions that make heavy use of coordination. Although the environment is collaborative, the selfish maximization did not manage to induce coordination. One tentative explanation for this is that because the number of states of the channel is larger than in the previous examples, 1-step empowerment is not able to 'see' far enough into the future to anticipate the need for helping out the other agent. Using a larger horizon could be a way of solving this issue.

6.7.4 Synthesis of the Results

This section investigated various scenarios involving two agents maximizing their own empowerment in an environment with memory:

- A competitive scenario where each agent tries to put the environment in a different states.
- A collaborative scenario where both agents are aiming for the same environment state.
- A more complex collaborative scenario where both agents have to work together in order to reach maximally empowered states.

The concept of Nash equilibrium from game theory was used in order to identify stable solutions in the space of possible agents' behaviours. In the case of competitive scenarios, it was shown that the game-theoretic approach was close to a zero-sum game, each agent increasing its empowerment at the expense of the other agent. In such a scenario, selfish empowerment maximization leads to suboptimal solutions in which the empowerment of both agents if relatively low compared to other possible solutions. A similar effect was already identified in Chapter 5 for spatial scenarios (which are competitive). Indeed, agents striving at maximizing their own empowerment did not generally manage to get to globally high empowerment states.

Experiments on the competitive scenario indicate that equilibrium points are reached without using coordination between the agents' actions. However, further analysis showed that this might not be generalizable to other competitive scenarios. Indeed, simple changes in the environment studied, while keeping its competitive nature, leads to solutions with coordination.

In the case of the collaborative scenario, it was shown that equilibrium points include optimal solutions where both agents have high empowerment. From a game-theoretic perspective, these situations are non-zero sum games. Indeed, best solutions imply that both agents try to put the environment into a state of high empowerment which benefits both of them.

Even though coordination was used in this scenario to reach optimal solutions, further analysis showed that this may not be generalized to other collaborative scenarios.

The last scenario studied, the mountain-climbing experiment, is a partly collaborative scenario in which agents actually need to work together in order to reach high empowered states. Because the state space of this scenario is relatively large compared to the two previous ones, a game-theoretic analysis was not possible. Instead a search was performed in the space of possible behaviours in order to find solutions that either selfishly maximize the agents' empowerment, or that maximize the combined empowerment of both agents.

Surprisingly, even though the environment is collaborative, selfish empowerment maximization did not generally lead to situations of high empowerment. It is not clear whether this is due to the nature of the search process (which contrarily to the game-theoretic analysis does not exhaustively explore the space of behaviours), or to the fact that, having a larger number of states, the agents should 'look further' into the future by maximizing empowerment with larger horizons.

However, global maximization of empowerment successfully led the agents to maximum empowerment situations. In this case coordination between agents was used in order to reach these states, but, as mentioned before for the collaborative scenario, it is not clear whether this can be generalized to other situations.

6.8 Summary

In this chapter we set out to investigate the effect of interferences between agents, i.e. the simultaneous execution of agents' actions. For this purpose, and in order to remove any other effect, a simplified model of the perception-action loops of two agents interacting in an environment has been proposed. This model has already been introduced in the field of network information theory and is referred to as the multiple-access channel (MAC). However, in this thesis, our perspective differs from standard network information theory. Indeed, our focus is on the control abilities of agents sharing an environment, i.e. their empowerment. Nevertheless, recently introduced tools for dealing with such channels, e.g. the algorithm for computing the total capacity (Rezaeian and Grant 2004), proved useful in this investigation.

This simplified model has been used as a first step in order to identify key properties of interferences and to study the effect of coordination between agents' actions on the capacity of MACs. One such effect is that, in some channels, the capacity can be increased by coordinating actions. Moreover, I formulated the conjecture that this capacity can at most be doubled when coordination is introduced. This conjecture is very similar to the capacity-doubling conjecture in MAC used with feedback that has been proposed in (Thomas 1987).

Various MACs have been numerically investigated in order to identify properties of these channels that make coordination between agents useful. More precisely, the following classes of channels have been identified:

- Non-interfering MACs: channels for which the capacity can be achieved without coordination of the agents and in which no interference occurs. This can be understood as a channel in which each agent acts on independent degrees of freedom of the output. In such channels, the behaviour of each agent has no impact on the empowerment of the other agents.
- Weakly interfering MACs: channels for which the capacity can be achieved without coordination, but in which actions interfere. Therefore, the behaviour of the agents have an impact on the empowerment of the other agents. Moreover, in such channels, knowing what actions the other agent is performing (e.g. through a context) can increase empowerment.
- Strongly interfering MACs: channels for which the capacity can only be achieved using coordination. These channels also share the properties of weakly-interfering channels.

In the second part of the chapter, the MAC model has been extended in order to better reflect the actual intertwined perception-action loops. Specifically, a common source of information (akin to the past sensorimotor experience of the agents) was introduced. The rationale behind this is that coordination in the actual perception-action loop has to be causally accounted for through the dependence of actions on past sensorimotor or context variables. Moreover, in the case of embodied agents, the information contained in the context variables are not counted in the empowerment of the agents (even though they may change its value). Explicitly modelling the common information source allows to properly account for these aspects.

Three different channels (one weakly and two strongly interfering) have been numerically studied using the common source model. Contrarily to the previous model, it was shown that coordination does not increase the empowerment of the agents. However, it allows them to reach new rate pairs that are inaccessible to non-coordinated agents. These new rate pairs become accessible through time-sharing mechanisms.

The last part of the chapter investigates models of MACs with memory using tools from game theory. More specifically, Nash equilibria are identified as points in the space of behaviour policies where empowerment-maximizing agents should converge. The agents are able to change their own empowerment by altering the distribution over the states of the environment through their own behaviour. Also, their behaviour has an impact on the empowerment of the other agents.

Two simple scenarios are investigated: a competitive one in which the two agents have different maximum empowerment states, and a collaborative one in which both agents share the same empowerment maximizing states.

Nash equilibria found in the competitive scenario show that agents competing for empowerment end up in a state of shared low empowerment (compared to other possible solutions). This confirms results found from spatial scenarios studied in Chapter 5, i.e. selfish empowerment maximization leads to suboptimal empowerment in competitive situations. Coordination between agents was not used in the equilibrium of the studied scenario, however it turns out that this cannot be generalized to other competitive situations.

Results from empowerment maximization in the collaborative scenario led to Nash equilibria in which agents shared a high empowerment. This property is likely to be generalizable to other similar scenarios. Optimal solutions found in our experiment used coordination between the agents' actions. However, similarly to the competitive case, this appears to be a property of the specific environment being studied, and therefore it may not be generalizable to other collaborative scenarios.

In the last experiment, the mountain climbing scenario, I investigated a more complex collaborative model in which the environment has a larger memory. Because of this, game-theoretic tools could not be used, and a stochastic search in the space of possible behaviours was performed. Results showed that agents selfishly maximizing their own empowerment did not manage to reach optimal solutions. It is not clear whether this result is due to the search process or to the 'short-sightedness' of 1-step empowerment. Nevertheless, optimal solutions could be found when the agents maximized their combined empowerments. The optimal solution made use of coordination between the agents. Coordination in this case is used to improve the control that the agents have over the state of the environment, allowing them put it into a state of high empowerment. However it is not clear whether this can be generalized to other scenarios.

From a design perspective, if one wants to create a system in which agents have high control abilities, i.e. high empowerment, then results from this chapter indicate that the following guidelines should be followed:

- The environment and/or the sensorimotor apparatus of the agents should be designed so as to limit interferences between their actions.
- The system should be designed so that high empowerment situations coincide for as many agents as possible. The idea is to avoid competitive situations and seek collaborative ones.
- Coordination between agents should be introduced if it can be used to improve the control that the agents have over the state of the environment.
- Depending on the memory that can be stored in the environment, proper temporal horizons may have to be identified, or agents should aim at maximizing a combination of their own empowerment and that of others.
- If the environment has to be competitive, agents should not selfishly maximize their empowerment, instead they should try to maximize a combination of their own empowerment and that of others.

Chapter 7

Conclusions and Future Work

In this thesis, we set out to identify general principles that may underlie the complexity appearing in the behaviour of living beings and in their collective organizations. For this purpose, the framework of causal Bayesian networks was used in order to model the perception-action loop of embodied agents. In this model, information has been identified as a crucial 'currency' that any agent has to take into account. The identification of agent-centered communication channels and their capacities allows us to make explicit some general limits on what can be achieved by a given agent.

One of these quantities, *empowerment* (Klyubin et al. 2008, Klyubin 2007), i.e. the capacity of the communication channel that goes from the actions of an agent to its future sensor states through the environment, has been shown to capture fundamental limits on how much perceivable control an agent has onto its future. Moreover, this quantity has several interesting properties, such as being agent-centric, local, task-independent and semantic free.

7.1 Contributions of the Thesis

The work presented in this thesis brought the following contributions by order of importance:

Extending the perception-action loop framework to multiple agents: A minimal model of the perception-action loop of two agents interacting in a common environment was presented. This model has then been connected to network information theory and the formalism of the *multiple-access channel*, allowing us to use recently introduced algorithms that can compute the capacity of such channels (Watanabe and Kamoi 2002, Rezaeian and Grant 2004), and therefore study the empowerment of multiple agents in such situations. The use of minimal models was introduced as a first step that allows us to identify general properties of multiple-access channels that can be used to understand more complex situations.

Identifying drives towards organization and coordination: It was shown that empowerment maximization in multi-agent systems can lead to competitive or collaborative situations. Global organization emerges as the result of such behaviours in a spatial environment. The multiple-access channel model allowed us to study the effect of coordination between agents. It was shown that such coordination can increase their empowerment for a class of channels. Empowerment maximization in various scenarios induces coordinated behaviour.

Identifying theoretical bounds on empowerment gain: An agent that has access to information about the state of the environment can increase its empowerment from two perspectives: (i) better distinguishing the effect of its actions, and (ii) picking actions according to the current state of the environment. These two aspects have been distinguished, and some theoretical bounds on their respective empowerment gain were identified.

Introducing new algorithms for context extraction: Information about the state of the environment is only available to the agent through its sensorimotor history. Extracting this information has been done in (Klyubin 2007) using evolutionary algorithms. This technique is improved upon by using one of the theoretical bounds previously identified. Two iterative algorithms inspired from the information bottleneck principle (Tishby et al. 1999, Slonim 2002) have been proposed that efficiently perform the same task.

Extending the perception-action loop framework to include feedback: The causal Bayesian model of the perception-action loop and the associated quantities introduced in (Klyubin 2007) were modified in order to include feedback mechanisms using the formulation of *directed information* (Massey 1990) and *feedback capacity* (Yang et al. 2005, Tatikonda and Mitter 2009).

Introducing heuristics for on-board model-acquisition: An accurate probabilistic model of the perception-action loop is needed to make use of the information theoretic tools presented in this thesis. A first heuristic was presented that allows an agent to acquire such a model with a minimum number of samples and that adapts to changes in the environment or embodiment. A second set of heuristics was introduced that help identify causal relationships that may span over long time delays, reducing the computational cost by not having to process all the intermediate time-steps.

7.2 Empowerment Maximization for Single Agents

Two main mechanisms for empowerment maximization have been studied in this thesis:

- manipulating the environment into preferred states: this amounts to controlling the environment in order to change it toward situations where the agent has a better control;
- acquiring information (context) about the state of the environment: knowledge about this state allows the agent to improve its control abilities and to distinguish the effect of its actions from the apparent noise induced by unknown hidden variables.

Because empowerment gives values to states of the environment, an agent that strives at maximizing its empowerment would then try to control the state of the environment to bring it to those that have high empowerment. Maximizing empowerment by changing this state can be seen metaphorically as navigating inside an abstract state space.

However, for an agent to be able to navigate in this state space, it first has to discover these states. Section 4.3 focused on this context-extraction problem. The ability for an agent to construct a context variable that identifies the state of the environment from its sensorimotor data is crucial to maximize its empowerment. When such context information is accessible, empowerment is increased by two factors:

- it allows the agent to distinguish the effect of its own actions from what appears as noise coming from hidden variables, and
- it allows the agent to increase its control on the environment by selecting actions according to the current state.

It has been demonstrated in this thesis that the amount of information that the context captures about the state of the environment limits the empowerment gain due to the first factor. Moreover, the amount of synergy between the context, actions, and future sensor states acts as a lower bound on the capacity increase. It was shown that context extraction can be done in a computationally efficient way by evolving a context-automaton that maximizes the lower bound, instead of directly maximizing empowerment as was done in (Klyubin 2007)). Moreover, two iterative algorithms have been proposed in Chapter

4 that allow us to compress the sensorimotor history into a context variable in order to maximize empowerment under limits on the amount of information captured.

These algorithms have also been used in order to study limits on empowerment gain using context when the amount of context information is constrained. Experimental results in a grid world scenario indicate that the overall empowerment gain is bounded from above by the amount of captured information. Moreover, they also indicate that there is an almost linear relationship between these two quantities.

Coming back to our original question about the behavioural complexity of agents. The aforementioned principles allow us to partially answer it. The complexity of the inferred model and the resulting behaviour are in direct relation to the complexity of the causal structure of the environment. However, under constraints on information capture and processing, agents have to limit the complexity of the inferred model. Put another way, they have to trade-off the empowerment gain of the inferred model (which is related to its complexity) and its informational cost.

This thesis also proposed other tools that can help agents to maximize empowerment. The first one is a heuristic that allows an agent to adaptively sample its perception-action loop in order to construct a reliable model of it while minimizing the number of samples. This heuristic is able to deal with changing environments and focus the exploration of the agent onto the aspects that have changed while ignoring the others. A second tool was proposed in order to deal with causal relationships that span over long time delays. Such situations are difficult to treat with the full model of the perception-action loop because of the combinatorial explosion that they may lead to.

7.2.1 Future Work

Further research could be conducted on mixing the two empowerment maximization mechanisms, i.e. control of the state and context extraction. For instance, an agent starting its life in an environment has no knowledge about its causal structure. After some time performing 'motor-babbling' in this environment and collecting statistics about its perception action-loop, the context-extraction algorithms can be used to infer the causal structure and states of the environment. By doing so, the agent is then equipped with a model of how this state is impacted by its actions. It then becomes possible to use empowerment to 'value' these inferred states, and to devise a behaviour policy that makes use of the model in order to reach states of high empowerment. The agent would then spontaneously identify and make use of the causal structure of the environment to maximize its control.

On top of this, if several agents are interacting in the same environment, the same principles should push them towards inferring models of each other's internal states (since those are part of the state of the environment). Then, equipped with a model of the environment and the other agents, an agent maximizing its empowerment would try to control the other agents in order to put itself in situations of high control. If all the agents are going through the same kind of process, one should expect a 'co-evolution' between each other's behaviours. Their internal models should then reflect this 'ecological' complexity (i.e. the behaviour of other agents), and because of this co-evolution, this may result in highly complex behaviours.

More theoretical work is needed to understand how the two empowerment maximization mechanisms are related, and what limits apply on each of them and their combination. Also, it would be interesting to compare this approach to similar ones such as (Still 2009). This last approach differs mainly by the fact that predictability, instead of empowerment, is maximized. It is possible to conjecture that Still's approach would capture aspects of the environment that are not necessarily relevant to what the agent can do in this environment, whereas empowerment maximization ignores any causal structure that has no relevance to the outcome of the agent's actions. Applying the two approaches in similar scenarios would allow one to compare the resulting models. The proposed conjecture could then be verified by studying different scenarios, with some of them implying irrelevant (from the agent's perspective) causal structure in the environment.

7.3 Empowerment Maximization for Multiple Agents

This thesis studied the effect of multiple agents interacting in the same environment. It was shown that two mechanisms of interaction can be identified:

- shared control of the environment state, and
- interferences between actions performed at the same time.

These two mechanisms have been investigated separately in chapter 5 and 6. In order to study the shared control, asynchronous models have been used. This property prevents agents from performing actions at the same time, and therefore avoids interferences.

Two simple examples, the bathroom and the well problem, have been described in terms of empowerment. Because of their specific structure, the value that different agents associate with the states of the environment lead to either competitive or collaborative situations. The bathroom problem is typical of competitive situations. In this example, two agents are competing to get to the highest empowerment situation. When one agent reaches this high empowerment situation, the other one has no control at all. And, although there is a situation which is optimal for both agents, selfish empowerment maximization should lead to a competitive behaviour.

In the well example, the situation is inverted. Basically, both agents 'agree' on the empowerment of the different states. More precisely, the state of highest empowerment is the same for both agents. Therefore, even in a selfish empowerment maximization context, both agents will 'work together' towards putting the environment in this preferred state.

Various spatial scenarios have also been studied. The main idea behind them is to have agents able to move in the space and perceiving each other's presence. It was shown that, in such scenarios, a general empowerment maximizing principle could be identified. It relies on the agents' ability to get to situations where they are close enough to each other while retaining sufficient freedom of movement. The principle behind this trade-off is that, in the studied environments, the only thing to perceive is the other agents, therefore being close enough to each other is necessary for the sensors to acquire any information at all. On the other hand, if the agents are too close to each other, they are prevented from moving, making their empowerment drop significantly. It was shown in Sec. 5.5 that in scenarios where several agents are randomly moving on a grid, there exist empowerment-optimal densities which are specific to the sensors used by the agents.

This thesis also studied the impact of the spatial organization of agents on their empowerment. It has been demonstrated in Section 5.6 that empowerment is maximized for specific organizations, which depend on the embodiment of the agents. For instance, directional sensors and density sensors lead respectively to checker-board or line patterns in a grid world scenario. The reason behind this is that the spatial organization provides strong regularities in the perception-action loops of the agents which would not exist in less organized systems. It was also shown that the behaviour of the agents could be selected for generating such empowerment-maximizing organizations from simple local rules.

The ability for selfish empowerment-maximizing agents to generate such organizations was also investigated. Experiments of Section 5.8 showed that, even though various complex organizations could emerge from the local behaviour, empowerment was not generally strongly increased. This can be explained by the prevalence of competitive situations in this particular scenario. This might not be the case in other kind of environments; more research should be conducted in order to identify scenarios where cooperation would dominate.

The phenomenon of *interferences* was studied separately, leading to one of the major contributions of this thesis. By reducing the perception-action loop of multiple agents to a minimum, it was possible to connect it to an information-theoretic model called the *multiple-access channel* (MAC). This thesis uses recently developed algorithms for computing the capacity of such channels in order to study their properties.

It was demonstrated in Sec. 6.4 that the overall capacity of such channels can be significantly increased if the agents coordinate their actions. Different classes of channels are identified, that basically differentiate between channels where coordination does not bring anything, increased suboptimal points of the capacity region, or increased the overall capacity. Moreover, channels have been identified for which the capacity gain can be higher than the amount of coordination between the agents' actions.

Several experiments have been conducted in Section 6.6 for various scenarios of two embodied agents acting in a memoryless MAC. These indicate that, even though the overall capacity of the channel can be increased by coordination, the amount of information that each agent sends independently into the channel cannot be increased. Instead, most the overall amount of information transmitted comes from the common source of information that allows the coordination. However, the shared control ability of both agents considered together is increased, which can lead to an increase in empowerment in environments that have memory.

Other experiments have been conducted in Section 6.7 in order to identify in a more systematic way the situations that would be reached by selfish empowerment-maximizing agents. Using tools from game theory, minimal competitive and collaborative scenarios have been investigated. Results indicate that empowerment maximizing agents should generally get to suboptimal situation in competitive environments and exhibit no coordination, whereas in collaborative scenarios agents should be able to reach optimal solutions and exhibit coordinated behaviour. Similarly, a more complex mountain-climbing scenario is investigated using simulated annealing. Agents selfishly maximizing empowerment in this scenario led to various solutions, either optimal or suboptimal, and did not necessarily use coordination. However, empowerment maximization at the global level led to the emergence of coordinated behaviour generally leading to optimal solutions.

Some designing principles could be identified for engineering multi-agents systems in which agents have maximum control over the environment. One of them is to design the embodiment of the agents and the environment so that the underlying channels are minimally interfering, allowing the agents to control independent degrees of freedom of the environment. In the case where interferences cannot be avoided, allowing the agents to coordinate their action is necessary in order to maximize their joint control on the environment, allowing them to guide the environment into states for which they have high empowerment.

Also, designing the agents and the environment so as to maximize collaborative situations over competitive ones is a very useful approach. Indeed, as was shown using game-theoretic tools, selfish empowerment maximization in collaborative environments generally leads to maximally empowered situations (to the extent that the appropriate horizons are used) whereas competitive scenarios lead to suboptimal situations.

7.3.1 Future Work

Further theoretical results about the MAC channel would be interesting to obtain. Specifically, it would be useful to identify precise bounds on how much empowerment can be increased through the use of coordination. Proving the capacity doubling limit through coordination that I conjectured in Section 6.3 for two or more agents would also be a useful result. Moreover, it may help resolving a related conjecture involving feedback that was introduced in (Thomas 1987). One of the most striking features of coordination which is still unresolved is that for some channels more capacity can be gained than the amount of coordination between agents.

In most of the experiments performed, only 1-step empowerment was considered. Further research should be conducted in order to study scenarios with longer horizons to which the current results may not be generalized. One could also consider looking at situations where the agents use a dedicated coordination channel that does not have to go through the environment, this would help identify upper bounds on what coordination can bring.

Also, more research is needed in order to get a better understanding of the interplay between shared control and interferences. Moreover, the interference phenomenon revealed very complex relationships between the various informational quantities of interest. A potential strategy of investigation would be first to identify how the control abilities of two agents differ from that of a single agent using an approach similar to (Touchette and Lloyd 2004), and then to study the impact of coordination between the agents, whether the common information comes from the state of the controlled process or from a separate source. It would also be of interest to study MACs with more than two agents, as the informational relations in such a case may differ from the two agents case.

Appendix A

List of Symbols

X, Y	Random variables
$\mathcal{X}, \mathcal{Y}$	Event spaces
x, y	Events of random variables
X_t, X_{t+1}	Random variables at specific timesteps
$X_t^n = (X_t, X_{t+1}..., X_{t+n-1})$	Sequence of random variables over time
$p(x), p(y)$	Probability distributions
$E[X]$	Expected value of X
$\lvert\mathcal{X}\rvert$	Cardinality of $\mathcal{X}$
$H(X)$	Entropy of a random variable
$H(X, Y)$	Joint entropy
$H(X\lvert Y)$	Conditional entropy
$I(X; Y)$	Mutual information
$D_{KL}(p(x)\lvert\lvert q(x))$	Kullback-Leibler divergence
$C(X \rightarrow Y)$	Capacity of the channel $p(y\lvert x)$
$\Phi(X \rightarrow Y)$	Information flow
$\widehat{X}$	Intervened node
$I(X \rightarrow Y)$	Directed information
$\mathfrak{E}(A_t \rightarrow S_{t+1})$	Empowerment
$\mathfrak{E}(A_t \rightarrow S_{t+1}\lvert S_t)$	Empowerment with context

Appendix B

Information Theory

Information theory (Shannon 1948) is a mathematical framework that quantifies properties of probability distributions. We refer the reader to (Cover and Thomas 2006) for a complete introduction to the field. One of the most important quantity is the *entropy* of a probability distribution. Consider a random variable X for which each event x can take a value in the set $\mathcal{X}$. The probability of one event x is written as $p(X = x)$. For the sake of simplification we will use the notation $p(x)$ when there is no ambiguity about the random variable considered. The entropy of this random variable is defined as

$$H(X) := -\sum_{x\in\mathcal{X}} p(x)\log_2 p(x). \tag{B.1}$$

This value reflects the uncertainty about the outcome of the random variable.

The *conditional entropy* of X given Y is defined as:

$$H(X|Y) := -\sum_{x\in\mathcal{X},y\in\mathcal{Y}} p(x,y)\log_2 p(x|y). \tag{B.2}$$

This quantity measures the amount of uncertainty about X when Y is known.

The *joint entropy* of X and Y is defined as:

$$H(X,Y) := -\sum_{x\in\mathcal{X},y\in\mathcal{Y}} p(x,y)\log_2 p(x,y). \tag{B.3}$$

Another important quantity of information theory is the *mutual information* between two random variables. This value measures the mutual dependance of the two variables:

$$I(X;Y) := \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} p(x,y) \log_2 \frac{p(x,y)}{p(x)\,p(y)} \tag{B.4}$$

where X and Y are two random variables, $p(x,y)$ is the joint probability distribution of X and Y, and $p(x)$ and $p(y)$ are the marginal probability distribution functions of X and Y respectively.

Because the mutual information measures the amount of information that any of the two variables provides about the other, it is connected to the entropies through:

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) && \text{(B.5)} \\ &= H(Y) - H(Y|X) && \text{(B.6)} \\ &= H(X) + H(Y) - H(X,Y) && \text{(B.7)} \end{aligned}$$

Appendix C

Computation of Stationary Distributions

Consider a conditional probability distribution $p(r_{t+1}|r_t)$ that describes the time evolution of a dynamical system. We are asking the following question: given an initial distribution $p(r_0)$, what is the expected stationary distribution $p(r_t)$ when $t \to \infty$?

The standard way of computing it is to consider the conditional distribution as a matrix P_{ij} where i is the source state and j is the destination state. Define $P^{(1)}$ as being equal to P_{ij}, one can compute the probability of transition after k steps by calculating $P^{(k)} = P_{ij}^k$. For a sufficiently high k, $P^{(k)}$ converges to a matrix that gives the probability of being in state j after k iterations if the system has been started in state i. After convergence the stationary distribution can be computed as $P^{(0)}P^{(k)}$ where $P^{(0)}$ is a one dimensional matrix containing the values of $p(r_0)$.

However in some cases there is no convergence of the matrix. A typical example is if the transitions are deterministic and generate a cycle. For example if the matrix is $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ then multiplying it by itself will lead to the matrix $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Each time k is increased one moves from one matrix to the other in an endless cycle, making it impossible to compute the stationary distribution. However it is obvious that the stationary distribution for this system is $\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \end{pmatrix}$ whatever the initial distribution is.

In order to solve this problem we use the following algorithm:

- $P^{(1)} \leftarrow P_{ij}$.
- $k \leftarrow 1$.

- $l \leftarrow$ empty list of distributions $p(r)$.
- While true
  - Compute $p(r_k) = P^{(0)}P^{(k)}$.
  - If l contains a distribution $p(r_i)$ for which $D_{KL}\big(p(r_k)||p(r_i)\big) \leq \epsilon$ then exit while.
  - Else $p(r_k)$ is added to l, $k \leftarrow k+1$ and $P^{(k)} \leftarrow P^{(k-1)}P_{ij}$.
- Identify the latest $p(r_i)$ that matches $p(r_k)$ (in terms of KL distance).
- Compute $p(r_\infty) = \frac{\sum_{j \geq i} p(r_j)}{N}$ where N is the size of the list l minus i.

Appendix D

Adaptation of the Perception-Action Loop Using Active Channel Sampling

Adaptation of the Perception-Action Loop Using Active Channel Sampling

Philippe Capdepuy[1], Daniel Polani[1,2], Chrystopher L. Nehaniv[1,2]
[1]Adaptive Systems Research Group
[2]Algorithms Research Group
School of Computer Science, University of Hertfordshire
College Lane, Hatfield, Herts, AL10 9AB, UK
{P.Capdepuy,D.Polani,C.L.Nehaniv}@herts.ac.uk

Abstract

During the lifetime of a real world agent or robot, many changes unforeseen at design time can occur. Whether these are due to a change in environmental conditions or to alterations of the embodiment of the robot, flexibility and adaptation are essential qualities that can help it to keep operating in this new situation. This work is based on an information-theoretic approach and introduces an exploration strategy that allows an agent to detect and adapt to changes in its perception-action loop by actively sampling areas of interest. We define the problem of exploring the sensorimotor channel and establish a measure of the distance between the observed and the real model of the channel. An optimal Oracle-based strategy is used to compare performances of the adaptive sampling strategy and a random strategy. Results for different scenarios of change in a binary channel show that the proposed strategy is highly effective in many cases. We also outline principles to adapt this mechanism to the exploration of multiple channels and we give preliminary results for such a scenario.

1 Introduction

The development of adaptive sensorics (and actuatorics) is a topic of high current interest and relevance. The advent of increasing powerful and ubiquitous computational resources has brought about the ability to construct hardware of many different sizes for a variety of use niches. This makes it increasingly important to provide this growing number of individual (and interconnected) devices with the ability to interact flexibly and adaptively. At this point, most of the activities in this direction have to be explicitly engineered: any adaptivity of a device has been planted into it by the manufacturers, any flexibility of reaction requires a protocol that specifies how a device is to handle novel stimuli and unforeseen situations. “True” adaptivity, in the sense of a device “learning on its own” is still very much elusive; existing device adaptivity relies on engineered failure/success models of devices.

In this dilemma, inspiration from biology is sought: biology has a seemingly unmatched reservoir of successful adaptation strategies. Evolution is probably the most celebrated of these, but there are many more: whether Neural Networks, Ant Colony Optimization, Artificial Immune Systems, or other paradigms, there is a rich variety of methodologies that have originally been motivated from the biological example.

While these paradigms share the general biological motivation, they have, structurally, little in common and it seems difficult to formulate a common principle which gives rise to them. This implies that any even bio-inspired adaptive algorithm used in an engineering problem needs to be hand-fitted to the problem at hand.

However, in the last years evidence has been mounting that even the convoluted dynamics of biological adaptation may be governed by simple fundamental principles; even more interestingly, some of these principles are well established in engineering, namely as principles of (Shannon) information optimization. For instance information maximization principles (infomax) give rise to biologically plausible neural receptive fields [16], or neural codes [18, 4, 7, 3, 21]. The latter seem to operate at the trade-off curve between information transmission and metabolic cost [15] and, more than that, organisms are ready to trade off a very significant amount of information (in typical cases of the order of magnitude of 10%-20% of the organism's total metabolic energy) to acquire sensoric and process it [14]. This indicates that (Shannon) information is a vital resource for organisms, almost on par with its metabolic energy. Why should that be the case? The main hypothesis is that of a principle of *parsimony*: of two organisms which e.g. utilize the same amount of metabolic energy it is likely that the organism which makes better use of the available information will have an evolutionary advantage. In absence of any evolutionary advantage of that information, the metabolic cost of processing the given information can be deevolved by degenerating the associated neural and sensoric apparatus (as happens with cave fish).

Such a parsimony principle provides a way of understanding what needs to happen in an adaptive system that mimics biological operation. However, there is another interesting factor involved: the influence of the environment on the organism does not reflect the standard view of a sender and receiver communicating with each other using a common code [8]. Rather environment and organism/agent interact in a quite intricate manner which nevertheless can be captured by novel mathematical formalisms: the treatment of information processing in the perception-action loop of agents can be modeled transparently by the use of causal Bayesian Networks [11, 10] which extend Ashby's Law of Requisite Variety [1, 22, 23] to general sensorimotor loops. This provides a handle for a quantitative treatment of general infomax scenarios of an agent and thus an approach towards a systematical, but yet biologically relevant methodology for constructing adaptive devices.

As a particularly promising path, the use of *empowerment* has been suggested [12, 13, 5], a concept similar to the channel capacity of the external part of the perception-action loop of an agent (we discuss this formally and in detail in Sec. 2). Empowerment measures by how much (in terms of information) an agent can *potentially* modify its environment so that it is able to register this modification. Essentially, empowerment quantifies a combined controllability/observability [17, 19] in information-theoretical terms.

Empowerment has been shown in a range of scenarios to constitute a universal utility that, if maximized locally, provides behaviours consistent with the natural choice of humans in a "self-motivated" way (not unlike the homeokinetic principle [6], autotelic principle [20] or the learning progress maximization [9] or the predictive information [2]). The reason for the success of empowerment is not fully understood at this time, although first hypotheses are emerging.

The current paper, however, will not preoccupy itself with this question — it will assume that, as evidence seems to indicate, the central hypothesis is valid that empowerment is indeed a quantity of interest to induce adaptive behaviour in an agent embedded in an environment via its sensorimotor loop. Up to now, all earlier scenarios studied calculated empowerment separately or externally. Once done, they assumed that, for the duration of a particular behaviour strategy, the empowerment profile of the system would stay unchanged. Real systems will be different — the reaction of the environment to the actions of an agent (even if in the same states) may change with time. In such cases, all the relevant quantities of the perception-action loop need to be reestimated for empowerment to be up-to-date. The current paper will discuss how to adaptively and efficiently estimate the relevant signatures of a perception-action loop. Section 2 introduces the information-theoretic perspective of the perception-action loop. In Sec. 3 we define the exploration problem and introduce a measure of the performance of exploration. The optimal Oracle-based policy and the adaptive exploration strategy are then introduced. The performance of the latter is then evaluated in different scenarios against the optimal strategy and a random one. Section 4 describes an adaptation of the exploration problem to multiple channels related through a topology of contexts and shows some preliminary results for a simple grid world.

2 The Information-Theoretic Picture of the Perception-Action Loop

We will refer to the perception-action loop of the agent as a causal Bayesian network which describes the relationships between the environment, the sensors and the actuators of the agent. The perception-action loop can then be unrolled in time (see Fig. 1) and some of its properties can be assessed using information-theoretic tools. One central aspect of our work is to investigate the sensorimotor channel, i.e. the channel that goes from actions to future perceptions through the environment. An important characterization of this channel is provided by the concept of empowerment [12, 13]. The idea is to measure how much information can be injected by an agent into its environment and then perceived back through its sensors. More precisely it is defined as the channel capacity from the sequence of actions $A_t, A_{t+1}, \ldots, A_{t+n-1}$ to the perceptions S_{t+n} after a fixed number of time steps. The channel capacity is defined as the maximum mutual information between the sent message and the received message, where the maximization is made with respect to the probabilities

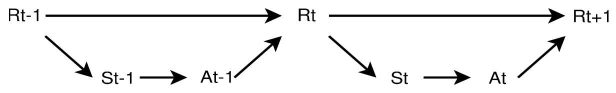

Figure 1. Representation of the perception-action loop as a causal Bayesian network unrolled in time. R_t stands for the environment of the system, S_t is the sensor of the agent and A_t its actuator.

for the sent message. In the context of this work, we will restrict ourselves to the simplest case where only the current action and the next sensoric state matter. Empowerment can then be written as

$$\mathfrak{E}(A_t \to S_{t+1}) = \sup_{p(a_t)} I(A_t; S_{t+1}) \qquad (1)$$

with $p(a_t)$ the probability distribution function of the action. Empowerment can be described as the maximum potential information an agent can transfer into its own sensors through the environment.

In the perception-action loop, the properties of the channel that goes from actions to future perceptions depend on both the embodiment of the agent and its coupling with the environment. In the case of a real agent these properties, described as the conditional probability distribution $p(s_{t+1}|a_t)$, are subject to changes due to alterations of the embodiment or changes in the environment. If only observational data are available and if the channel is unstable, estimating empowerment becomes a difficult task. To get good estimates of empowerment, an accurate model of the environment is necessary. The purpose of this work is to provide an active exploration strategy that maximizes the accuracy of the constructed model.

3 Exploration as Sampling of the Perception-Action Loop

In all this section we will use a conceptually simple case, the single channel case, to define the basic principles of our exploration strategy. The perspective taken in this work is to consider an agent that constructs a statistical model of its perception-action loop by collecting samples. This model is represented by a probability distribution $p(s|a)$ (precisely it is $p(s_{t+1}|a_t)$ but for the sake of clarity we will use the short version) with $s \in \mathcal{S}$ being the perceptual space, and $a \in \mathcal{A}$ the set of possible actions. To construct this model, the agent has to explore the channel by acting on it. At each time-step it picks an action and sends it into the channel, through the environment, and then perceives back a particular sensor value.

By collecting such data it is possible to approximate the real probability distribution of the channel (if it is stationary). However, if one supposes that the channel can sometimes be changed (e.g. external damage, change in the environment) then the agent has to reevaluate its statistical model to reflect the changes and match the new real model. We make the assumption that the channel is changed to another almost stationary channel.

In the following subsections, we formalize what are the real and the observed model and define a measure of their distance. Using this measure we can establish an Oracle-based optimal strategy for exploration. Subsequently we propose a simple heuristic that allows to approximate this strategy. Efficiency of this heuristic is then evaluated against the optimal strategy and a purely random one.

3.1 Real and Observed World

The whole point of an exploration strategy when used on its own is to provide the explorer with an accurate model of its environment. Basically the world can be described as a model, and the subjective vision of the explorer is another model. The purpose of exploration is to minimize the distance between the real and the observed model. In the single channel case, the real world model is represented by a probability distribution $p_r(s|a)$ and the agent model is constructed by sampling the channel, leading to another probability distribution $p_o(s|a)$.

As our goal is to maximize the accuracy of the observed channel, we need a way of measuring how much the two models match. For this purpose we use the Jensen-Shannon distance between the two distributions, averaged over all actions (which we will consider equiprobable). The Jensen-Shannon distance is based on the Kullback-Leibler divergence between two distributions p and q, defined by

$$D_{KL}(P||Q) = \sum_i p(i) \frac{\log p(i)}{\log q(i)}, \qquad (2)$$

but where the distance is the average of the divergence between each distribution and their average M:

$$D_{JS}(P||Q) = \frac{1}{2}\Big(D_{KL}(P||M) + D_{KL}(Q||M)\Big). \qquad (3)$$

where $M = \frac{1}{2}(P+Q)$. We can therefore measure the distance ϵ between the real and the observed model using

$$\epsilon(P_o||P_r) = \frac{1}{|\mathcal{A}|}\sum_a D_{JS}\big(P_o(S|a)\big|\big|P_r(S|a)\big). \qquad (4)$$

3.2 Defining an Optimal Sampling Strategy

Now that the problem has been stated and that we have a measure of the distance between the observed and the real model, we can define what we will consider as an optimal strategy. The goal of the exploration strategy is to match as quickly as possible the real world model by sampling it with actions. An optimal strategy is one that would maximally reduce this distance at each sampling.

If one considers that there exists an Oracle who knows the real model of the environment, one can define a strategy that will use this Oracle to pick the actions which are more likely to have an informative outcome (in the sense that it will change our current knowledge). Formally we define the change in accuracy δ_ϵ when performing action a and observing outcome s (i.e. by adding a new sample at time t) by

$$\delta_\epsilon(a,s) = \epsilon\Bigg(P_o|S_{t+1}=s, A_t=a \quad \Bigg|\Bigg| \quad P_r\Bigg) - \epsilon(P_o||P_r) \qquad (5)$$

where $P_o|S_{t+1}=s, A_t=a$ is the observed model after being updated with the new sample. According to the real model of the environment we can define for each action a the expectation of change in accuracy by

$$E\big[\delta_\epsilon|A_t=a\big] = \sum_s p_r(s|a)\delta_\epsilon(a,s). \qquad (6)$$

As our goal is to minimize the distance between the observed and the real model, the optimal Oracle-based strategy is to pick the action a which has minimum $E[\delta_\epsilon|A_t=a]$. Of course a real agent does not have access to the Oracle, however this strategy will be useful in our case to evaluate the performance of other strategies.

3.3 Approximating the Optimal Sampling Strategy

Now comes the central question. How can an agent that has no access to an Oracle discover an efficient sampling strategy. The goal of the agent is also to minimize the distance between his observed model and the real one, but it has no access to this distance measure. One way to obtain information that is relevant to this problem is to consider not only the current observed model, but also how it evolves in time.

In the case of an agent that has a model that perfectly matches the environment, the sampling process will not bring anything new, i.e. it will not change the model (apart from small fluctuations, but this problem is addressed at the end of the paper). However if the model of the agent is not accurate for a particular action, sampling this action will provoke strong changes in the distribution of sensoric outcomes. By taking into account this time evolution, the agent can estimate how accurate the different parts of its model are, and therefore have an idea about the ϵ function that only the Oracle detains. From the agent perspective we can make the following assumption: if a part of our model changed due to recent sampling, then our model was (and probably still is) not accurate. Therefore if we want to maximize the accuracy of our model, this part needs more sampling in order to converge to the real distribution.

The key idea of our approach is to quantify these changes in the distribution and then to use this quantity as a guide to pick the action that is most likely to get us to the real model. To measure the change in the probability distribution, we use the variation of the entropy of the distribution. Formally, for a given action a, and an observed outcome s, the entropy variation of the corresponding distribution is

$$\delta_H(a,s) = H(P_o|S_{t+1}=s, A_t=a) - H(P_o|a) \qquad (7)$$

where H stands for the Shannon entropy

$$H(X) = -\sum_{x\in\mathcal{X}} p(x)\log p(x). \qquad (8)$$

To use this heuristic, the agent simply has to favor actions which are changing the model, i.e. actions for which $|\delta_H|$ is maximum. The absolute value is taken because we do not care if the entropy is increasing or decreasing, what we care about is if it changes at all.

3.4 Results for the Single Channel Case

To evaluate our heuristic (δ_H), we compare it with a random strategy (which always converges after a sufficiently long time) and the Oracle-based strategy. The experiment consists in providing initial data from a particular channel, assuming that it is known perfectly by the agent, and then changing the channel and letting the agent explore it. We measure how much time each strategy takes to converge to 1% of the initial error ϵ.

We use a collection of different binary channels described in table 1 that have different properties in term of randomness. For each pair of different channels (one used as initial channel and the other used as the

Table 1. Binary channels used for evaluation, separated in deterministic channels, half deterministic, and completely random.

Name	$p(S=1\|A=0)$	$p(S=1\|A=1)$
ID	0	1
NOT	1	0
ZERO	0	0
ONE	1	1
HID0	0	$\frac{1}{2}$
HID1	$\frac{1}{2}$	1
HNOT0	1	$\frac{1}{2}$
HNOT1	$\frac{1}{2}$	0
RAND	$\frac{1}{2}$	$\frac{1}{2}$

changed channel) we perform 100 experiments and average the measures. We use the Oracle-based strategy as a baseline for the speed of convergence, and we express the result for the random strategy and the δ_H as the ratio between their convergence time and the baseline. The δ_H strategy is in fact an ϵ-greedy strategy with $\epsilon = 0.1$, meaning that 90% of the time the agent picks the action that has maximum $|\delta_H|$ and a random action the rest of the time. Results are described in tables 2 and 3.

For every combination of channels studied, the δ_H strategy clearly outperforms the random strategy. On average the δ_H strategy takes 9% more time than the baseline Oracle-based strategy, whereas the random strategy takes on average 62% more time. Qualitatively it is possible to classify the different scenarios into two main groups. The first group includes all the channel changes that involve a modification of the outcomes for both actions. In this group the random and the δ_H strategy have close results, but the δ_H strategy still outperforms the random one, having an average ratio of 1.05 against 1.23. But the real effectiveness of the δ_H strategy appears when changes are only partial (in this case when only one of the actions has a different outcome after the channel change). In this case it has an average ratio of 1.13 against 2.02 for the random strategy.

If one action has been changed but the other stayed the same, then only for the first one will the entropy change and therefore it will be sampled more often. In the case where both actions are changed we obtain a slightly more complex behaviour. This is the case for the scenario ID to HNOT0 (see Fig. 2). In this scenario the outcome of both actions are changed. For action 0 the outcome changes from a deterministic (only 0) to the opposite deterministic distribution (only 1). On the other hand, action 1 changes from a deterministic outcome (1) to a random one. We can observe on the graph that the behaviour of the δ_H strategy differs quite a lot from the Oracle-based and the random ones. The two latter strategies sample both actions at very similar frequencies whereas the δ_H strategy strongly changes over time. To understand it better we now describe in detail what is happening (we suggest the reader to first have a look at Fig. 3 to have a graphical representation of the problem).

Table 2. Ratio between the convergence time of ($Random$;δ_h) and the baseline time provided by the Oracle-based strategy for each scenario. Rows represent the initial channel, columns correspond to the channel after the change. The second part of the results is shown in table 3.

	ID	NOT	ZERO	ONE	HID0
ID	—	1.0;1.0	2.0;1.1	2.0;1.1	2.0;1.1
NOT	1.0;1.0	—	2.0;1.1	2.0;1.1	1.5;1.1
ZERO	2.0;1.1	2.0;1.1	—	1.0;1.0	2.0;1.0
ONE	2.0;1.1	2.0;1.1	1.0;1.0	—	1.5;1.1
HID0	2.0;1.1	1.0;1.0	2.0;1.1	1.0;1.0	—
HID1	2.0;1.1	1.0;1.0	1.0;1.0	2.0;1.1	1.5;1.1
HNOT0	1.0;1.0	2.0;1.1	1.0;1.0	2.0;1.1	2.0;1.2
HNOT1	1.0;1.0	2.0;1.1	2.0;1.1	1.0;1.0	1.5;1.1
RAND	1.0;1.0	1.0;1.0	1.0;1.0	1.0;1.0	2.0;1.4

Table 3. Continuation of table 2.

	HID1	HNOT0	HNOT1	RAND
ID	2.0;1.0	1.5;1.1	1.5;1.1	1.1;1.0
NOT	1.5;1.1	1.8;1.0	2.2;1.0	1.1;1.0
ZERO	1.5;1.1	1.4;1.1	1.8;1.0	1.0;1.0
ONE	2.0;1.0	2.1;1.1	1.5;1.1	1.1;1.0
HID0	1.6;1.1	2.0;1.6	1.5;1.1	1.9;1.3
HID1	—	1.5;1.2	2.0;1.2	2.1;1.2
HNOT0	1.5;1.1	—	1.5;1.1	2.2;1.3
HNOT1	2.0;1.2	1.6;1.1	—	2.3;1.4
RAND	2.0;1.4	2.0;1.3	2.0;1.4	—

At the beginning, both distributions have first to move from a deterministic low-entropy distribution to a high-entropy random one. However as the outcome of action 0 is always 1, it moves faster toward the maximum entropy state than action 1 does, leading to higher δ_H. Therefore during the first 40 time-steps of simulation sampling is dominated by action 0. When this distribution gets close to the maximum entropy one, its derivate diminishes, making action 1 the most sampled during the next 300 hundreds time-steps. At

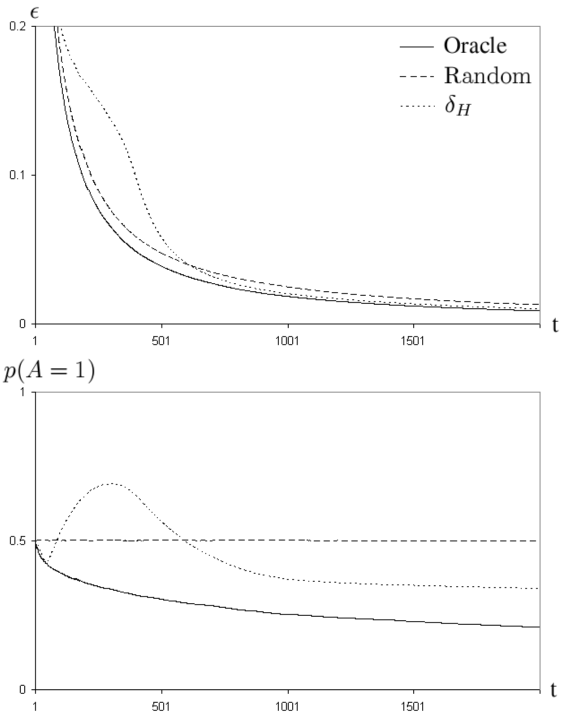

Figure 2. Scenario with a complete change of the channel (ID to HNOT0) averaged over 100 experiments (2000 time-steps). Top: time evolution of distance between observed and real model for the different strategies. During the first 600 time-steps, the δ_H-strategy deviates from the two others. After this time it overtakes the random strategy and gets close to the optimal one. Bottom: proportion of actions sampled (see text and Fig. 3 for details)

.

this point, action 1 is getting very close to the maximum entropy level where it has to converge. However sampling of action 0 starts to move from the maximum entropy (and low derivative) toward the lower entropy value where it converges. By doing so the derivative grows leading to a positive feedback effect that reinforces exploration of action 0. Eventually both distributions converge toward the changed channel.

4 Multiple Channels

Now we consider a more complex case: exploration of multiple channels. By multiple channels we do not mean that the agent has a number of constantly accessible channels for which it has to get a model, in

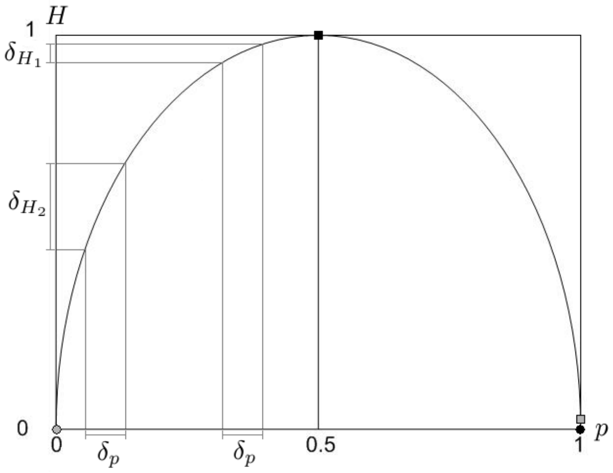

Figure 3. Entropy function of a binary distribution (using base two logarithm). For a given probability change δ_p due to a new sample, the entropy change δ_H depends on the previous state of the observed distribution. When entropy is maximum (i.e. $p = 0.5$), δ_H reaches a minimum. Dots represent the distribution for action 0 (circle) and 1 (square) of the initial (gray) and changed (black) channels in the scenario ID to HNOT0. During the sampling process, the gray dots converge toward their black couterparts.

that case we would simply consider them as one composite channel and use exactly the same strategy as for the single channel case. In this section we are interested in situations where channels are not all directly accessible to the agent but instead it has to move between channels by performing actions (and sampling at the same time). For such a case we will refer to the concept of *contexts*. We assume that the agent is able to distinguish different contexts (for example based on the current sensoric state) and that each context c is associated with a particular channel.

We first define how the contexts are related to each other through a topology and we translate the problem of channel exploration to this topology. Two cases are distinguished, the first one is the general case where the channels and their topology are not related. The second one is a particular case where the channels and their corresponding topology are completely intertwined. This case has very important connections with models of the perception-action loop and empowerment maximization. We then introduce a simple mechanism to use the δ_H-strategy in such topologies. Again, simulation results for simple scenarios are used to compare the different strategies.

4.1 Context Topology

We introduce a principle which we refer to as *context topology*. The idea is the following, for the sampling agent the world is represented as a collection of separate channels $c \in \mathcal{C}$ similar to the ones described in the previous section but uniquely identified by a context. When the agent is in a particular context, it performs an action to sample the corresponding channel. The difference with the single channel case is that the action will not only bring a new perceptive sample but it might also move the agent in a different channel. The context topology is described by the probability distribution $p(c_{t+1}|a_t, c_t)$ and it can also be subject to changes.

4.2 Propagating the Sampling Strategy

The goal of the agent is still to maximize the accuracy of its model $p(s_{t+1}|a_t)$ where a_t is an action and s_{t+1} is the sensor state obtained after performing the action. However now there are multiple channels and for all of them we have to maximize the accuracy. To adapt the δ_H-strategy to this topology of channels, we use a framework similar to that of reinforcement learning. For each channel-action pair we associate a 'reward' value which is simply the last entropy change of the distribution associated with this action in this context $\delta_H(a, c)$. This value is then propagated into the topology by using a value-iteration algorithm:

```
foreach $c \in \mathcal{C}$ do
    $V(c) \leftarrow 0$;
end
repeat
    $\Delta \leftarrow 0$;
    foreach $c \in \mathcal{C}$ do
        $V'(c) \leftarrow max_a \Big( \delta_H(a, c) +$
        $\gamma \sum_{c_{t+1}} p(c_{t+1}|c_t, a_t)V(c) \Big)$;
        $\Delta \leftarrow max(\Delta, |V'(c) - V(c)|)$;
    end
    $V = V'$;
until $\Delta < \theta$ ;
```

Algorithm 1: Value iteration algorithm in the multichannel case.

In this algorithm γ is the discount factor and θ is a small number that stops the algorithm when a sufficient precision has been reached. When the agent is in context c, the action-selection process consists in picking the action a that maximizes the utility quantity

$$U(a, c) = \delta_H(a, c) + \gamma \sum_{c_{t+1}} p(c_{t+1}|c_t, a_t)V(c). \qquad (9)$$

4.3 Preliminary results

We evaluate this model in a simple grid world with a moving agent. The agent senses its absolute position in the world and it can move to any neighboring cell (if not occupied by a block) or stay in the same cell. The current sensor value is used as the context. Initially the grid world is surrounded by blocks, preventing the agent to move out of it, but the inside is empty. We allow the agent to collect statistics about this initial environment. After some time we introduce a block inside the box, changing the channels that are located next to this block.

The experimental setup consists of a 11 by 11 grid world and we performed 100 experiments during which we measured the distance between the observed and the real model during 1000 time-steps. To avoid being stuck sampling areas already very close to the real value, we used a Boltzmann selection instead of the ϵ-greedy strategy. In a given context c the probability of picking action a is defined as $p(a) = \frac{1}{Z} e^{U(a)/T}$ where Z is a normalization factor $Z = \sum_{a'} e^{U(a')/T}$, T is a temperature parameter, and $U(a, c)$ is the utility calculated by the value-iteration algorithm.

Parameter values used for this experiment are $T = 0.01$, $\gamma = 0.8$ and $\theta = 0.001$. We measured the distance between the observed and the real model for the δ_H strategy and the random strategy at the end of the experiment. Values obtained for the δ_H strategy are significantly better, 24% of the initial distance, than the random strategy that reaches on average 58% of the initial distance. These preliminary results are encouraging but a more systematic study is needed to properly assess the effectiveness of the δ_H strategy in such multichannel case.

5 Conclusion

In the context of agents constructing a model of their perception-action loop by collecting statistics, we have proposed an active sampling strategy (δ_H) based on the temporal change of the entropy of the model. This strategy allows an agent to quickly adapt to changes of their perception-action loop. As the perception-action loop of the agent reflects its embodiment and the coupling with the environment, any change in the environment or any damage to the sensoric of actuatoric apparatus of the agent can impact the model of the perception-action loop. Using the proposed adaptive sampling strategy, the agent will reinforce exploration of these changes in order to quickly converge to the new model.

We first performed a set of experiments on different scenarios of change with a single binary channel case and measured the convergence time for the different strategies. The results for the δ_H strategy are very close to the optimal Oracle-based strategy (9% more time); comparatively, the random strategy performed quite poorly (62% more time). The behaviour of this strategy has been detailed in some particular scenarios.

We extended the δ_H strategy to the exploration of multiple channels related to each other by a context topology. Preliminary results on a simple grid world show that the proposed strategy performs significantly better than a random one. However more results are needed to validate its efficiency in different scenarios.

Future investigations will focus on the use of such an exploration strategy for maximization of empowerment. We expect this model to extend results in the area of self-organization in collective systems (as has been investigated in [5]). Useful applications of this model also include sensor evolution scenarios, where different sensorimotor apparatus can be evaluated in a given environment and compared on such criteria as stability of perception-action loop model and potential capacity to inject information in future sensoric states.

References

[1] W. R. Ashby. *An Introduction to Cybernetics.* Chapman & Hall Ltd., 1956.

[2] N. Ay, N. Bertschinger, R. Der, F. Güttler, and E. Olbrich. Predictive information and explorative behavior of autonomous robots. *European Journal of Physics B*, 2008. Submitted.

[3] W. Bialek, R. R. de Ruyter van Steveninck, and N. Tishby. Efficient representation as a design principle for neural coding and computation. arXiv.org:0712.4381 [q-bio.NC], December 2007.

[4] N. Brenner, W. Bialek, and R. de Ruyter van Steveninck. Adaptive rescaling optimizes information transmission. *Neuron*, 26:695–702, 2000.

[5] P. Capdepuy, D. Polani, and C. Nehaniv. Maximization of potential information flow as a universal utility for collective behaviour. In *2007 IEEE Symposium on Artificial Life.* IEEE, 2007.

[6] R. Der. Self-organized acquisition of situated behavior. *Theory Biosci.*, 120:1–9, 2001.

[7] A. L. Fairhall, G. D. Lewen, W. Bialek, and R. de Ruyter van Steveninck. Efficiency and ambiguity in an adaptive neural code. *Nature*, 412:787–792, 2001.

[8] J. J. Gibson. *The Ecological Approach to Visual Perception.* Houghton Mifflin Company, Boston, 1979.

[9] F. Kaplan and P.-Y. Oudeyer. Maximizing learning progress: an internal reward system for development. In F. Iida, R. Pfeifer, L. Steels, and Y. Kuniyoshi, editors, *Embodied Artificial Intelligence*, volume 3139 of *LNAI*, pages 259–270. Springer, 2004.

[10] A. Klyubin, D. Polani, and C. Nehaniv. Representations of space and time in the maximization of information flow in the perception-action loop. *Neural Computation*, 19(9):2387–2432, 2007.

[11] A. S. Klyubin, D. Polani, and C. L. Nehaniv. Organization of the information flow in the perception-action loop of evolved agents. In *Proceedings of 2004 NASA/DoD Conference on Evolvable Hardware*, pages 177–180. IEEE Computer Society, 2004.

[12] A. S. Klyubin, D. Polani, and C. L. Nehaniv. All else being equal be empowered. In *Advances in Artificial Life, European Conference on Artificial Life (ECAL 2005)*, volume 3630 of *LNAI*, pages 744–753. Springer, 2005.

[13] A. S. Klyubin, D. Polani, and C. L. Nehaniv. Empowerment: A universal agent-centric measure of control. In *Proc. IEEE Congress on Evolutionary Computation, 2-5 September 2005, Edinburgh, Scotland (CEC 2005)*, pages 128–135. IEEE, 2005.

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

[15] S. B. Laughlin, R. R. de Ruyter van Steveninck, and J. C. Anderson. The metabolic cost of neural information. *Nature Neuroscience*, 1(1):36–41, 1998.

[16] R. Linsker. Self-organization in a perceptual network. *Computer*, 21(3):105–117, March 1988.

[17] A. I. Mees. *Dynamics of feedback systems.* John Wiley & sons, Ltd., New York, 1981.

[18] F. Rieke, D. Warland, R. de Ruyter van Steveninck, and W. Bialek. *Spikes.* A Bradford Book. MIT Press, 1999.

[19] E. D. Sontag. *Mathematical control theory; Determistic finite dimensional systems (Texts in Applied Mathematics)*, volume 6. Springer-Verlag, New York, 1990.

[20] L. Steels. The autotelic principle. In F. Iida, R. Pfeifer, L. Steels, and Y. Kuniyoshi, editors, *Embodied Artificial Intelligence: Dagstuhl Castle, Germany, July 7-11, 2003*, volume 3139 of *Lecture Notes in AI*, pages 231–242. Springer Verlag, Berlin, 2004.

[21] S. F. Taylor, N. Tishby, and W. Bialek. Information and fitness. arXiv.org:0712.4382 [q-bio.PE], December 2007.

[22] H. Touchette and S. Lloyd. Information-theoretic limits of control. *Phys. Rev. Lett.*, 84:1156, 2000.

[23] H. Touchette and S. Lloyd. Information-theoretic approach to the study of control systems. *Physica A*, 331:140–172, 2004.

Appendix E

Construction of an Internal Predictive Model by Event Anticipation

Construction of an Internal Predictive Model by Event Anticipation

Philippe Capdepuy*, Daniel Polani*†, and Chrystopher L. Nehaniv*†

Adaptive Systems* and Algorithms† Research Groups
School of Computer Science, University of Hertfordshire
College Lane, Hatfield, Herts, AL10 9AB, UK
{P.Capdepuy,D.Polani,C.L.Nehaniv}@herts.ac.uk

Abstract. We introduce information-theoretic tools that can be used in an autonomous agent for constructing an internal predictive model based on event anticipation. This model relies on two different kinds of predictive relationships: time-delay relationships, where two events are related by a nearly constant time-delay between their occurrences; and contingency relationships, where proximity in time is the main property. We propose an anticipation architecture based on these tools that allows the construction of a relevant internal model of the environment through experience. Its design takes into account the problem of handling different time scales. We illustrate the effectiveness of the tools proposed with preliminary results about their ability to identify relevant relationships in different conditions. We describe how these principles can be embedded in a more complex architecture that allows action-decision according to reward expectation, and handling of more complex relationships. We conclude by discussing issues that were not addressed yet and some axis for future investigations.

1 Introduction

Designing agents that can act intelligently in a previously unknown environment is one of the most challenging issues in behavioral robotics. Such an agent must have the ability to construct an internal model describing the dynamics of the environment and the effect of its own actions on this environment. This can be mainly understood as extracting predictive relationships between events occurring in the perceptive field of the agent, whether these events are under its control (its actions) or if they are externally generated. This internal model allows the agent to predict forthcoming events, as well as the effect of its own actions on the environment. Such a predictive ability paves the way to anticipation and smart decision making by allowing the agent to decide which action to perform to obtain or avoid a given outcome. According to the classification of [1], these agents are said to perform *state anticipation.*

Our main focus in this paper is to define and evaluate tools that allow the construction of such an internal model regardless of any reinforcement. In this sense we are very close to latent learning and the concept of *expectancies* proposed by Tolman [7]. We describe an architecture that uses these tools to effectively construct the internal model and we explain how this model can be used to anticipate events. The robustness of this architecture to length and variability of time-delays between relevant associations is evaluated in two experiments. A last experiment shows the temporal dynamic of the model and more especially the forgetting mechanism.

The paper is structured as follows: in Sec. 2 we formulate the problem of event anticipation along with some examples and what we would expect from our anticipation system. Section 3 introduces the main information-theoretic concepts used in our model and the two kinds of relationships they allow us to identify. Section 4 describes the anticipation architecture embedding these concepts. In Sec. 5 we describe preliminary results concerning the predictive efficiency of the proposed tools in a simple simulation experiment. Section 6 describe some possible extensions to the actual model, mainly considering the problem of action-selection and how our predictive model can be used in a reward-based behavior. We also introduce a possible mechanism for handling more complex situations involving sequences of events and non-occurrence. Section 7 summarizes the issues our model addresses and we discuss some of those that will be investigated in future work.

2 Event-Based Anticipation

2.1 Stating the Problem

Here we refer to anticipation in a very general way as the ability to predict, more or less accurately, the future occurrence of *perceptive events*. These events can be seen as different stimuli that the agent can encounter in its environment. We consider as a preliminary simplification that the agent is not allowed to act onto its environment, he can only observe it (handling of actions will be described in Sec. 6). Different from other approaches, the agent is not provided here with a continuous flow of sensoric values for different modalities. Instead we consider that the agent perceives discrete events in discrete time (0 to n different events can be observed at a given time-step). We will denote the set of possible events by $\mathcal{E}$. The agent is then observing a stream of events such as the one represented in Fig. 1. The only relevant information that can be extracted from this stream are the relationships in time between similar or different events. The purpose of our work is to find an efficient way to identify these relationships in an anticipatory perspective.

2.2 Expected Properties

We want to infer a predictive model from observing the stream of events. According to a given recent past, the predictive model could then be used to anticipate

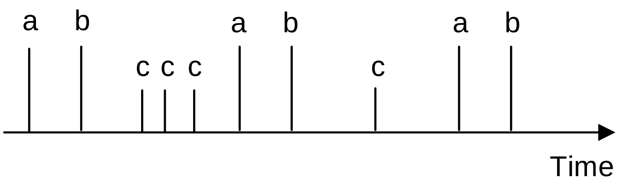

Fig. 1. Example of a stream of events $\mathcal{E} = \{a, b, c\}$ over time. The height is not relevant but just for clarity. In this particular example, we can observe that a is always directly followed by b with a fixed time delay. The event c seems to have a more complex pattern.

what the next events should be, and when they will occur. One of the constraints we put on our model is that it should be robust to noise and variations in the relationships. Also time-scale variations should have no effect on the efficiency of the predictive model construction (if for example all events have their delay multiplied by 2). Figure 2 shows three different cases where there exists a predictive relationship (a predicts b). One is rather obvious but different configurations of the time-delay between a and b and other noise events can lead to more difficult situations. To allow the extraction of these relationships, we will split our analysis in two different components. The first one is the relation from one event to all the others, the idea is to identify the most probable event than will occur shortly after another one (or shortly before if we look toward the past). The second component considers only pairs of events and its role is to measure the precision of the time-delay between these events.

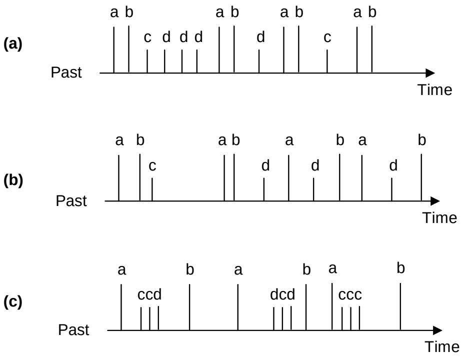

Fig. 2. Three different streams of events with $\mathcal{E} = \{a, b, c, d\}$. For each of them our aim is to identify the predictive relationship from a to b. Example **(a)** is quite obvious, events a and b follow each other very closely in time, and with a constant delay. Example **(b)** is more tricky as the delay between a and b varies, anyway it seems that b always follows a. In example **(c)** the delay between a and b is very large, providing room for many events to occur in between, nevertheless as this delay is constant we would like to identify such a relationship.

3 Information Theory and Anticipation

The goal of constructing an internal predictive model is to minimize the uncertainty of the predictions that the model will make. This construction can only be based on information acquired through experience, and therefore on a partial view of the environment, leading to probabilistic representations. Tools for dealing with such representations have been increasingly used in the context of sensorimotor coordination (for example Bayesian modelling in [6]), to analyze properties of the coupling between an agent and its environment (information-theoretic approach in [5]) and also to describe conditioning processes with information theory (see [3] and [4]). In our particular context, information theory is a very valuable tool because it is a natural framework to deal quantitatively with uncertainty.

3.1 Basis of Information Theory

Shannon's information theory is a mathematical framework that provides quantitative characterizations of probability distributions of events. We refer the reader to [2] for a complete introduction to the field. One of the main quantities we will be using is the entropy of a probability distribution. Consider a random variable X for which each event x can take a value in the set $\mathcal{X}$. The *entropy* of this random variable is defined as

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x), \tag{1}$$

where $p(x)$ is the probability that event x occurs ($\sum_{x \in \mathcal{X}} p(x) = 1$ and $0 \leq p(x) \leq 1, \forall x \in \mathcal{X}$). This value reflects the uncertainty about the outcome of this random variable. The minimum is 0 for an absolutely predictable outcome (for example one outcome has a probability of 1) and the maximum is $\log_2(|\mathcal{X}|)$ if all outcomes are equiprobable.

The *information content* or self-information of one particular event x according to the given probability distribution is defined as

$$I(x) = -\log_2 p(x). \tag{2}$$

The minimum information content is 0 if this outcome has a probability of 1 and goes toward infinity as the probability approaches 0.

Our use of information theory in this model concerns the extraction of relationships between time-located events such as perceptions or actions. For understanding the tools described below, it is only necessary to keep in mind that high entropy H means high uncertainty, and high information content I means a low probability event (or surprising event).

3.2 Time-delay relationships

We will first focus on time-delay relationships between two events. For example, if an event b always occurs 50 timesteps after another event a, then we would

like to identify this relationship. Also we would like the method to have some tolerance for variability, i.e. if b sometimes occurs 49 or 51 timesteps after a, we still consider that there exists a time-delay relationship between them.
For identifying these relationships, we will use information quantities. The principle used is based on the concept of *causal entropy* (see [8]) which in our case should be referred as *predictive entropy*. The idea of predictive entropy is the following: let us consider that we want to identify a time-delay relationship between an event a and an event b always occurring after a. We will then use a random variable $D_{a,b}$ that represents the probability distribution of the observed time-delay for the next occurrence of b after a (i.e. the observed delay between an observation of a and the next subsequent observation of b). The entropy of this random variable $H(D_{a,b})$ reflects the strength of the relationship. The lower the entropy, the stronger the relationship. For example if b always occurs 50 timesteps after a, the entropy of $D_{a,b}$ will be 0 (only one event with a probability of 1, see Fig. 3).

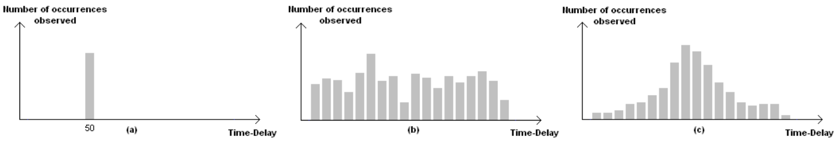

Fig. 3. Histograms of time-delay probability distribution, number of occurrences observed (vertical axis) for each possible time-delay (horizontal axis). **(a)** Histogram of an event b always occurring 50 timesteps after a, $H(D_{a,b}) = -\log_2(1) = 0$. **(b)** Example of a high entropy histogram. **(c)** Example of a low entropy histogram.

The original purpose of causal entropy is to determine whether there may be a relationship between two events from a to b or from b to a. This can be determined by comparing the entropies of $D_{a,b}$ and $D_{b,a}$. In our context, the goal is to identify relationships between many events. Therefore, we need a criterion for saying that there exists a time-delay relationship. In [3], the author states that the baseline from which the information provided by a conditional stimulus can be estimated is the prior estimate of the unconditional stimulus frequency. In our framework this can be translated as saying that the criterion for identifying a relationship from a to b is based on the self-relationship $D_{b,b}$, i.e. the distribution of observed time delays between two successive b events. We will therefore consider that there exists a relationship from a to b if a is a less uncertain predictor for b than b itself, i.e. if

$$H(D_{a,b}) < H(D_{b,b}). \tag{3}$$

Using causal entropy in our context leads to some problems that we need to solve. The first problem is that it is not robust at all to variability in time. If we consider for example two different conditions, in the first one, b occurred

2,10,50 and 100 timesteps after a. In the second case, b occurred 48, 49 ,50 and 51 timesteps after a. For both conditions, $H(D_{a,b}) = 2$ (4 equiprobable outcomes, so $H(D_{a,b}) = \log_2(4) = 2$), therefore, we cannot identify which condition reflects a relationship. Obviously the second one seems to be a relationship where b occurs approximatively 50 timesteps after a, whereas the first condition doesn't seem to be a time-delay relationship.

To solve this problem, the idea is to introduce some variability in the probability distribution. Therefore rather than updating the statistics of $D_{a,b}$ by adding one realization of a given time delay t, we add a gaussian distribution of time-delays centered around t, i.e. we add many realizations of t, then a bit less realizations of $t-1$ and $t+1$, even less for $t-2$ and $t+2$, and so on... Now if we get back to our example, adding gaussian noise around the actual observed values of 48, 49, 50 and 51 will lead to overlapping gaussians, and therefore to less variability than in the first condition, and consequently to a lower entropy (see Fig. 4). For a given time-delay t, the number of realizations to add is computed for growing distances $\varDelta_t$ as

$$\lfloor\left(\frac{\beta}{\sigma\sqrt{2\pi}} exp\left(-\frac{(\varDelta_t - t)^2}{2\sigma^2}\right)\right) \tag{4}$$

until this number reaches 0. The parameters β and σ of this function will be detailed in the Architecture section.

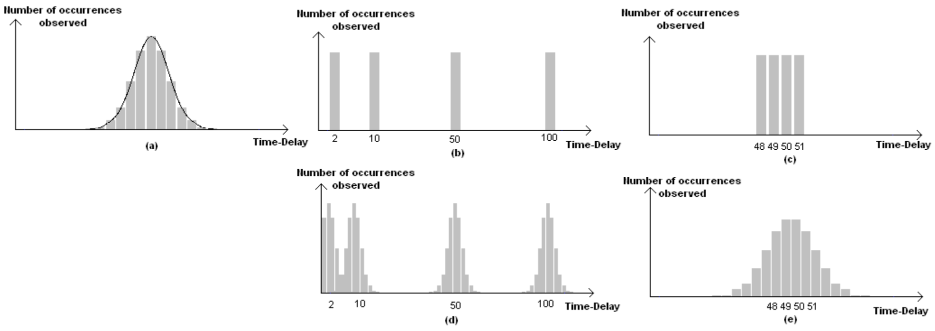

Fig. 4. Usefulness of adding gaussian noise to time-delay events. **(a)** For one occurrence of a time-delay event, we add a discretized gaussian distribution of realizations centered around the occurring event. **(b)** Example: histogram of the first condition without gaussian noise, $H(D_{a,b}) = 2$. **(c)** Example: histogram of the second condition without gaussian noise, $H(D_{a,b}) = 2$. **(d)** Example: histogram of the first condition with gaussian noise, $H(D_{a,b})$ is high. **(e)** Example: histogram of the second condition with gaussian noise, $H(D_{a,b})$ is low.

According to the quantity of information gained from using $D_{a,b}$ rather than $D_{b,b}$, we can compute a confidence value of the time-delay expectation as

$$\tau_{a,b} = \frac{H(D_{b,b}) - H(D_{a,b})}{H(D_{b,b})} \tag{5}$$

We can also compute the average expected time-delay between a and b as

$$\delta_{a,b} = \sum_{t \in D_{a,b}} p(t)t, \tag{6}$$

where $p(t)$ is the observed probability of the time-delay t.

Another problem that has to be solved is the following. Let us suppose that after some time we have identified the time-delay relationship between a and b that has been described in the example above (Fig. 4.e). Now if we consider that a new event c happened 10 timesteps before b, then the histogram of the random variable $D_{c,b}$ would be a perfect gaussian centered on 10. The entropy of this random variable will be lower than the entropy of $D_{a,b}$ because of the small time variation between a and b. But obviously, if we had 4 realizations of b after a (48, 49, 50 and 51 timesteps), then we should be more confident into this relationship than for b after c which had only 1 realization. Put another way, we should be more confident in a relationship that has occurred several times, even with some variation, than into a relationship that occurred only a few times, even with a perfectly constant time-delay. A way to solve this problem is to initialize any random variable $D_{a,b}$ with a uniform probability distribution of time-delays, e.g. an initial white noise. Then multiple realizations of a time-delay, even with some variability, will increase the probability of this time-delay and its neighbourhood, and decrease the probability of the noise values, therefore the entropy of such a random variable will be lower than the entropy of a noisy random variable with only one realization of a time-delay.

3.3 Predictive relationships extracted from contingency

Now we will focus on another type of relationship for which there is no precise delay between events a and b. We consider here relationships of the type "when a occurs, b is likely to occur soon". These relationships can be extracted from the contingency of events in the stream of perceptions. We will speak about them as *contingency relationships*, and we will consider that the closer b occurs after a, the stronger the relationship. Also we will consider that a predicts b if a *mainly* predicts b (relatively to predicting other events) and if b is *mainly* predicted by a (relatively to other events it is predicted by). The purpose of this criterion is the following: let consider an event a that happens all the time, and sometimes an event b, c or d happens. On one hand we can say that b, c and d are well predicted by a, because among all the possible predicting events, a is the most frequent. But on the other hand we cannot say that a usefully predicts b, c or d,

because it predicts nearly everything (even itself), and therefore it is a useless predictor. That is why for establishing a predictive relationship from a to b, our criterion takes into account the future of a *and* the past of b.
We can translate these by the following principle: for each event e, we have two random variables, one is related to its past, i.e. it reflects the probability distribution of events that happened before e, we will refer to it as CP_e; and one is related to its future, i.e. the probability distribution of events that happened after e, we will refer to it as CF_e. In this context we will say that there is a relationship between a and b, such that b is a consequence of a if

$$I_{CF_a}(b) < H(CF_a) \tag{7}$$

and

$$I_{CP_b}(a) < H(CP_b). \tag{8}$$

This means that the information carried by b when occurring after a is less than the average information carried by an event that has occurred after a, thus b is more likely to occur after a than other events; and that a when occurring prior to b carries less information than the average information carried by an event in the past of b, i.e. a is more likely to have occurred before b than other events.

For each of these variables, event realizations are added according to their distance in time, i.e. when close in time, many realizations of the same event are added (for one actual occurrence), the number of realizations added decreasing with the distance. The exact number of realizations follows the same gaussian equation 4, in which we replace t by 0, and Δ_t by the actual distance between the two events (negative values are discarded). Again we can define a confidence value of the contingency expectation, based on the loss of uncertainty, as

$$\kappa_{a,b} = \frac{1}{2}\left(\frac{H(CF_a) - I_{CF_a}(b)}{H(CF_a)} + \frac{H(CP_b) - I_{CP_b}(a)}{H(CP_b)}\right). \tag{9}$$

4 Architecture

The two information-theoretic tools described above are put together in an anticipation architecture. The main components of the architecture are shown in Fig. 5. First saliency evaluation filters perceptive events, forwarding only the unusual events (those that carry most of the information). These perceptive events are used to update the internal model and their last observed occurrence is updated. The internal model and the last event occurrences are then used together to build expectations about forthcoming events.

4.1 Salience Filtering

We introduce a first mechanism that filters out some of the perceptions to avoid overloading the system with useless information. The precise criterion we use is that according to a distribution probability of perceptions E, which is constantly

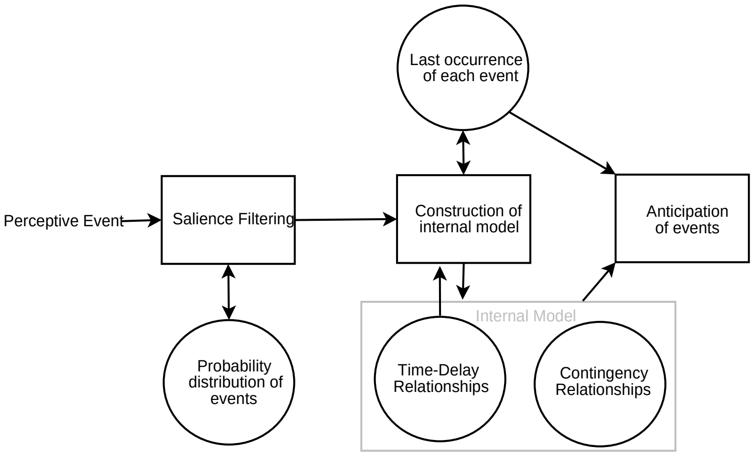
Last occurrence of each event
Perceptive Event
Salience Filtering
Construction of internal model
Anticipation of events
Internal Model
Probability distribution of events
Time-Delay Relationships
Contingency Relationships

Fig. 5. Main architecture. Circles represent stored information, boxes are processes that generate information. See text for details.

updated with new perceptions, we consider salient perceptions those that carry more information than the average information carried. Therefore the saliency criterion can be expressed as

$$I(e) > H(E) \tag{10}$$

where $e \in \mathcal{E}$.

4.2 Construction of the Internal Model

When an event is perceived, it is first stored into memory and replaces any previously stored occurrence of this event. The construction of the internal model is based on the two processes of finding time-delay and contingency relationships. When an event b is processed, for all events a that are in short-term memory, if it is the first occurrence of b since a occurred, we update the statistics of the random variables $D_{a,b}$, CF_a and CP_b. The parameters of the gaussian used for updating the statistics are fixed for $D_{a,b}$ to β_0 and σ_0. For the two other random variables, these are adapted according to the event they concern, i.e. the longer the expected self time-delay between the concerned event, the more the gaussian is flattened (hence the arrow from time-delay relationships to the construction of the internal model). The idea is to adapt to events that occur at very different timescales. Also the β parameter (the height of the gaussian) is adapted according to the frequency of the added event, here the idea is to strengthen the association with rare events and to weaken associations with very common events. Therefore when adding and event b to the statistics of a, the parameters used are

$$\sigma = \sigma_0(1 + \alpha\delta_{a,a}) \tag{11}$$

and

$$\beta = \beta_0(1 + \alpha\delta_{a,a} + \lambda\delta_{b,b}) \tag{12}$$

where α is the range adaptation coefficient and λ is the intensity adaptation coefficient (both low positive values). The higher α, the more the gaussian is flattened for a given self time-delay. The higher λ, the more the added event is important for a given self time-delay.

4.3 Anticipation of Events

The constructed internal model, along with the memory of the last occurrences of events, can easily be used to determine the expected events using the following principles. For each past event a in memory, all the $D_{a,b}$ random variables are evaluated, and for each of them which validate the condition 3, the event b is added into the expectation list, along with its average time-delay $\delta_{a,b}$ and its confidence value $\tau_{a,b}$. Then for each possible event b, if we can find any event a in memory that is valid according to contingency conditions 7 and 8, then b is added to the expectations list, again with its average time-delay $\delta_{a,b}$ and its confidence value $\kappa_{a,b}$.

4.4 Forgetting Mechanism

We introduce a forgetting mechanism to allow for a quick replacement of relationships that are not relevant anymore. The principle of the forgetting mechanism is to define an upper bound to the total number of realizations of the random variables describing the internal model. When a new realization is added and increases the total number above the defined bound, one other realization is removed, by randomly choosing one of the events stored and removing one realization of this event.

5 Experiments

In this section we will evaluate the ability of the architecture described above to extract relevant predictive relationships from the stream of perceptions. The agent is not allowed to act, it can only passively perceive events coming from its environment. We first detail the experimental setup then we analyze the confidence value of relationships of interest.

5.1 Experimental setup

Here we simulate some kind of Skinner box where the agent is situated. The perceptions of the agent are taken from the set $N1, N2, N3, N4, N5, N6, L1, Food$. The events from $N1$ to $N6$ are noise events that have no predictive value, whereas events $L1$ and $Food$ are causally associated, $L1$ predicting the $Food$ event ($L1$

stands for *Light 1*, we consider than when the light is flashed, food will be given to the agent in a given delay). $L1 - Food$ sequence has a probability of 0.02 of being initiated at each timestep. The noise events are generated at each timestep with the respective probabilities $(N1 : 0.2, N2 : 0.1, N3 : 0.05, N4 : 0.025, N5 : 0.0125, N6 : 0.00625)$. Other parameters of the simulation are the following. Gaussian parameters $\sigma_0 = 3$ and $\beta_0 = 100$. Range adaptation coefficient $\alpha = 0.25$. Intensity adaptation coefficient $\lambda = 0.1$. Random variables have an upper bound of 1000 realizations.
The first experiment measures the confidence values of the contingency and time-delay relationships after 10000 steps of simulation for different time-delay of the $L1 - Food$ association. The time-delays evaluated range from 1 to 80 timesteps with a variability of $+/-3$ timesteps.
For the second experiment we use the same procedure but the parameter investigated is the variability of the time-delay of the $L1 - Food$ association. The base time-delay used is 14 timesteps and with a variability ranging from $+/-0$ to 20 timesteps.
The third experiment aims at evaluating the dynamics of the internal predictive model over time. The $L1 - Food$ association has a time-delay of 14 timesteps and a variability of $+/-3$ timesteps. The experiment is running over 100000 timesteps, and during the range 40000 to 60000 $L1$ and $Food$ are not associated anymore, they are both presented at each timestep with the same probability of 0.01.

5.2 Results

Results of the first and second experiment are shown in Fig. 6. We can see from these results that contingency relationships are successfully extracted for short time delays, less efficiently when the time delay increases, but they are robust to variability of this time delay. On the other hand, time-delay relationships have the opposite behavior, i.e. they are robust for long time delays, but they loose efficiency as the variability increases. These results confirm the expected behavior of these two anticipation mechanisms, which used together should allow the extraction of most relevant relationships.

Results of the third experiment are shown in Fig. 7. We can see that both relationships are quickly learned, correctly forgotten when the two stimuli are not associated anymore, and then their confidence value increases as soon as the events are paired again. These results show that the architecture correctly account for forgetting mechanism. We can see that for a long enough time of exposure to the unpaired events during a typical run, the agent can completely forget the contingency relationship. On the other hand, the time-delay relationship is maintained for a longer time and its original confidence value is recovered very quickly when the events are paired again, whereas the contingency relationship shows a slower recovery rate.

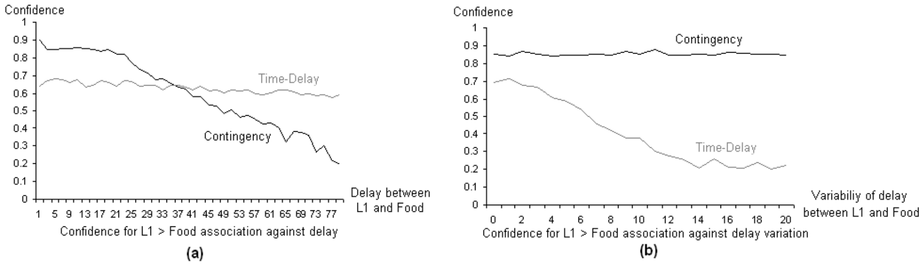

Fig. 6. Plotting of $\kappa_{L1,Food}$ (black) and $\tau_{L1,Food}$ (gray) after 10000 steps simulations. **(a)** Plotting against time delay between $L1$ and $Food$. Time-delay relationship is robust whereas contingency is not. **(b)** Plotting against variability of the time delay between $L1$ and $Food$. Contingency relationship is robust whereas time-delay relationship is not.

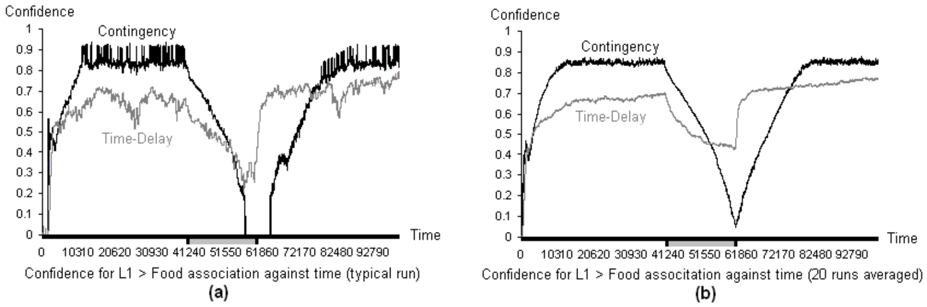

Fig. 7. Plotting of $\kappa_{L1,Food}$ (black) and $\tau_{L1,Food}$ (gray) against time during 100000 steps of simulation. In the range 40000 to 60000 $L1$ and $Food$ are not causally associated (shown in gray on the horizontal axis). **(a)** A typical run. **(b)** Average of 20 experiments.

6 Possible extensions

6.1 Actions and Rewards

Until that point we have only used the anticipation architecture in the context of an agent that can only passively observe the stream of perceptive events. But the point of such an architecture is to be used for action-selection. This aspect can be considered from two different perspectives: goal-oriented behavior and reinforcement learning. For both cases we will consider that actions are special perceptive events (e.g. proprioceptive events) that are generated when the agent performs the action. The particularity of these events is that they cannot be predicted by anything (from the agent's point of view) as they are dependant upon the will of the agent. By propagating these proprioceptive events into the architecture it becomes possible for the agent to extract predictive relationships between his actions and their effect in the environment.

In the case of goal-oriented behavior, we consider that the agent wants to reach a given goal whose definition is outside the scope of the architecture. For simplicity reasons, we can consider that the goal is a particular event. By chaining backward into the predictive model from this event toward possible actions, it is possible to identify which actions should lead to the occurrence of the goal event. It should also be in principle possible to plan more complex sequences of actions to reach intermediate events that will ultimately lead to the goal.

If we now consider the case of action-selection based on reward expectation (as in reinforcement learning), the predictive model can be used in the opposite way. The idea would be to attach reinforcement values to particular events (such as the acquisition of food or an electric shock). When the agent must decide what action to perform, it is possible for him to estimate the effect of each possible action and moreover to compute an expected reinforcement value by chaining forward until reaching events with reinforcement. During the chaining, confidence values of the relationships can be used to estimate the probability of obtaining the reinforcement. The computed values can be used to select the action that will most probably lead to reinforcement. An advantage of this system over classical reinforcement learning is that it possible introduce a complex online modulation of reinforcement values (for example food events are rewarding only if the agent is hungry).

6.2 Handling complex predictive relationships

One of the most difficult issues of anticipatory systems is to be able to identify complex phenomena involving many different events. An example of such a phenomena is that when an event a occurs, doing the action b will result in the event c occurring. One possible way to tackle this issue is to introduce sequence of events. The idea is to construct sequences of events that will be processed as normal events and that can therefore be used as predictors for other events. The problem here is to take care of the combinatorial explosion when grouping events. Therefore we need a criterion for creating new sequences, and also another one for discarding them when they have proved unsuccessful. The idea is to introduce a sequence generation probability p_{sg} that will be used each time an event b is processed to decide if a new sequence has to be created, another event a is then chosen randomly in the recent history and a new sequence a, b is registered. Subsequent occurrences of this sequence would then be recognized and the corresponding event generated and processed by the anticipation system. Using a sequence destruction probability p_{sd} evaluated at each time-step, a randomly chosen sequence may be destroyed if it has no predicting power, with a probability growing with the “age” of this sequence. Forwarding sequence events in the normal events’ pathway allows for the construction of longer sequences by associating already existing sequences with other events.

Another case of complex relationship is when an event c predicted by a can be avoided if the action b is performed before c occurs. In this case we have to take into account the NON-occurrence of an expected event. The idea is that when an expected event did not happen after a sufficiently long time, a special event, opposite of the expected one, is generated and forwarded into the normal pathway. For example if an event a predicts an event c, and if after some time this event c still has not occurred, then we will generate an event $\overline{c}$ and forward it into the event processing pathway. This event can then be associated with another event b that caused this non-occurrence, or to the sequence of events a, b.

7 Conclusion

We have introduced two information theory based tools for extracting time-delay and contingency relationships in the stream of perceptions. These tools have been put together into an architecture that uses them for constructing an internal model of the environment. We have shown two distinct properties of contingency and time-delay relationships, the former is robust to variations of the delay between two stimuli, and the latter keeps its efficiency when the time-delay gets larger. An obvious advantage of these tools is that they allow simultaneous identification of relationships with completely different time scales without suffering from complexity increase. We have also shown the efficiency of this architecture for constructing a relevant internal model that is able to quickly adapt to a changing environment. Nevertheless some more extensive tests have to be carried out to evaluate the architecture in different conditions.

One of the advantages of this internal predictive model is that it can be used in two different ways. On the one hand it can be used to predict which events will occur and then perform appropriate actions to take advantage of this knowledge, such as avoiding a negative reinforcement. On the other hand it can also be used for goal-oriented behaviour. In this case the goal would be a particular event (usually a positive reinforcement) the agent wants to obtain. Using the predictive model it can identify which events predict the goal and then chain back until it can find which actions can initiate the sequence of events leading to the goal. However this last part is a bit more complex as it involves not only predictive relationships between events, but also true causal relationships which are more difficult to identify. For example if we consider that an agent has learned that the sound of a bell predicts food delivery (by the experimenter), ringing the bell will not bring the food because the source of causality is upstream to both events and not from one to the other. Identification of causality requires the agent to actively inject information into the environment by acting upon it. In the example of the bell described above, if the agent can ring the bell by itself, then it would quickly realize that the bell and the food are not causally associated. Such a principle could also be used as a drive toward exploratory behaviour. The idea would be that when a given predictive relationship has been identified, the agent could then try to more precisely evaluate this relationship by provoking the first event and then identify if the relation is causal or not.

One drawback of the architecture is that we use purely symbolic events, so no relation between them can be found apart from the predictive ones; it is impossible to define a notion of similarity between events and therefore impossible to generalize the predictive relationships. For this to be possible, events should not be only symbolic but they should possess a set of properties from which a notion of distance and subsets could be used.

Another problem is that the computational complexity of the model grows quickly with the number of different events that the agent can perceive, that was the reason for us to introduce a saliency filter so as to get rid of irrelevant events. Another possible way to avoid this problem and the previous one would be to map real events defined in a space of properties to a symbolic space by using categorization, i.e. grouping similar perceptions into one symbolic event, hence allowing for some generalization of relationships and also limiting the number of different events.

References

1. Butz, M. V., Sigaud, O., and Gerard,P., (2003). Internal models and anticipations in adaptive learning systems. In Butz, M. V., Sigaud, O., and Gerard,P. (Eds.), Lecture Notes in Computer Science vol. 2684: Anticipatory Behavior in Adaptive Learning Systems. Springer-Verlag, pp. 86-109.
2. Cover, T., and Thomas, J., (1991). Elements of Information Theory. New York: John Wiley and Sons.
3. Gallistel, C. R., (2002). Frequency, Contingency and the Information Processing Theory of Conditioning. In Sedlmeier, P. and Betsch, T. (Eds.), Frequency Processing and Cognition. Oxford, UK: Oxford University Press, pp. 153-171.
4. Gallistel, C. R., (2003). Conditioning from an Information Processing Perspective. Behavioural Processes. 61(3) 1234, pp. 1-13.
5. Klyubin, A. S., Polani, D., and Nehaniv, C. L., (2004). Tracking Information Flow through the Environment: Simple Cases of Stigmergy. In Artificial Life IX: Proceedings of the 9th International Conference on the Simulation and Synthesis of Living Systems, pp. 563-568. The MIT Press.
6. Kording, K.P. and Wolpert D.M., (2004). Bayesian integration in sensorimotor learning. Nature 427, pp. 244-247.
7. Tolman, E. C., (1959). Principles of purposive behavior. In Koch, S. (Ed.), Psychology: A Study of Science pp. 92-157. New York: McGraw-Hill.
8. Waddel, J., Dzakpasu, R., Booth, V., Riley, B. T., Reasor, J. D., Poe, G. R., Zochowski, M., (2007). Causal Entropies - A measure for determining changes in the temporal organization of neural systems. Journal of Neuroscience Methods (In Press).

Bibliography

Almeida e Costa, F. and Rocha, L. M.: 2005, Introduction to the special issue: Embodied and situated cognition, *Artificial Life* **11**(1-2), 13–30.

Anthony, T., Polani, D. and Nehaniv, C. L.: 2008, On preferred states of agents: how global structure is reflected in local structure, *in* S. Bullock, J. Noble, R. Watson and M. A. Bedau (eds), *Artificial Life XI: Proceedings of the Eleventh International Conference on the Simulation and Synthesis of Living Systems*, MIT Press, Cambridge, MA, pp. 25–32.

Arimoto, S.: 1972, An algorithm for computing the capacity of arbitrary discrete memoryless channels, *IEEE Transactions on Information Theory* **18**(1), 14–20.

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

Atick, J. J.: 1992, Could information theory provide an ecological theory of sensory processing, *Network: Computation in Neural Systems* **3**(2), 213–251.

Attneave, F.: 1954, Some informational aspects of visual perception, *Psychological Review* **61**(3), 183–193.

Axelrod, R. and Hamilton, W. D.: 1981, The evolution of cooperation, *Science* **211**, 1390–1396.

Ay, N. and Polani, D.: 2008, Information flows in causal networks, *Advances in Complex Systems* **11**(1), 17–41.

Ay, N., Bertschinger, N., Der, R., Güttler, F. and Olbrich, E.: 2008, Predictive information and explorative behavior of autonomous robots, *European Physical Journal B* **63**(3), 329–339. Submitted.

Barlow, H. B.: 1959, Possible principles underlying the transformations of sensory messages, *in* W. A. Rosenblith (ed.), *Sensory Communication: Contributions to the Symposium on Principles of Sensory Communication*, The M.I.T. Press, pp. 217–234.

Barlow, H. B.: 2001, Redundancy reduction revisited, *Network: Computation in Neural Systems* **12**(3), 241–253.

Bell, A. J.: 2003, The Co-Information Lattice, *ICA 2003*, Nara, Japan.

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

Bonabeau, E., Dorigo, M. and Theraulaz, G.: 1999, *Swarm Intelligence : From Natural to Artificial Systems (Santa Fe Institute Studies on the Sciences of Complexity)*, Oxford University Press, USA.

Boyd, S. and Vandenberghe, L.: 2004, *Convex Optimization*, Cambridge University Press.

Brenner, N., Strong, S. P., Koberle, R. P., Bialek, W. P. and Van Steveninck, R. R. D. R.: 2000, Synergy in a neural code, *Neural Computation* **12**, 1531–1552.

Brooks, R. A.: 1990, Elephants Don't Play Chess, *Robotics and Autonomous Systems* **6**(1&2), 3–15.

Cannon, W. B.: 1939, *The wisdom of the body*, W. W. Norton, New York.

Capdepuy, P., Polani, D. and Nehaniv, C. L.: 2007a, Construction of an internal predictive model by event anticipation, *in* M. V. Butz, O. Sigaud, G. Pezzulo and G. Baldassarre (eds), *Anticipatory Behavior in Adaptive Learning Systems - From Brains to Individual and Social Behavior*, LNCS/LNAI, Springer, pp. 218–232.

Capdepuy, P., Polani, D. and Nehaniv, C. L.: 2007b, Grounding action-selection in event-based anticipation, *in* F. Almeida e Costa, L. M. Rocha, E. Costa, I. Har1vey and A. Coutinho (eds), *Proceedings of the Ninth European Conference on Artificial Life*, Vol. 4648 of *LNCS/LNAI*, Springer, pp. 253–262.

Capdepuy, P., Polani, D. and Nehaniv, C. L.: 2008, Adaptation of the perception-action loop using active channel sampling, *Proceedings of the 2008 NASA/ESA Conference on Adaptive Hardware and Systems.*

Chechik, G.: 2003, Spike-timing-dependent plasticity and relevant mutual information maximization, *Neural Computation* **15**(7), 1481–1510.

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

Deci, E. L. and Ryan, R. M.: 1985, *Intrinsic motivation and self-determination in human behavior*, Plenum Press, New York.

Der, R., Steinmetz, U. and Pasemann, F.: 1999, Homeokinesis — A new principle to back up evolution with learning, *Proceedings of the International Conference on Compu-*

tational Intelligence for Modelling Control and Automation (CIMCA'99), Vienna, 17-19 February 1999.

Gibson, J. J.: 1979, *The Ecological Approach to Visual Perception*, Lawrence Erlbaum Associates, New Jersey, USA.

Heylighen, F.: 1992, Principles of systems and cybernetics: an evolutionary perspective, *Cybernetics and Systems* pp. 3–10.

Jeffery, W. R.: 2009, Evolution of eye degeneration in cavefish, *Developmental Biology* **331**(2), 412 – 412.

Kaplan, F. and Oudeyer, P.-Y.: 2004, Maximizing learning progress: an internal reward system for development, *in* F. Iida, R. Pfeifer, L. Steels and Y. Kuniyoshi (eds), *Embodied Artificial Intelligence*, Vol. 3139 of *LNAI*, Springer, pp. 259–270.

Klyubin, A.: 2007, *Organization of Information Flow Through the Perception-Action Loop*, PhD thesis, School of Computer Science, University of Hertfordshire, UK.

Klyubin, A. S., Polani, D. and Nehaniv, C. L.: 2004, Tracking information flow through the environment: Simple cases of stigmergy, *in* J. Pollack, M. Bedau, P. Husbands, T. Ikegami and R. A. Watson (eds), *Artificial Life IX: Proceedings of the Ninth International Conference on the Simulation and Synthesis of Living Systems*, The MIT Press, pp. 563–568.

Klyubin, A. S., Polani, D. and Nehaniv, C. L.: 2005, Empowerment: A universal agent-centric measure of control, *Proceedings of the 2005 IEEE Congress on Evolutionary Computation*, Vol. 1, IEEE Press, pp. 128–135.

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.

Klyubin, A. S., Polani, D. and Nehaniv, C. L.: 2008, Keep your options open: An information-based driving principle for sensorimotor systems, *PLoS ONE* **3**(12), e4018.

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

Laughlin, S. B., de Ruyter van Steveninck, R. R. and Anderson, J. C.: 1998, The metabolic cost of neural information, *Nature Neuroscience* **1**(1), 36–41.

Linsker, R.: 1988, Self-organization in a perceptual network, *Computer* **21**(3), 105–117.

Lungarella, M. and Sporns, O.: 2006, Mapping information flow in sensorimotor networks, *PLoS Computational Biology* **2**(10), e144+.

Massey, J. L.: 1990, Causality, feedback and directed information, *Proceedings of the International Symposium on Information Theory and its Applications.*

Maturana, H. R., Varela, F. J. and Beer, S.: 1980, *Autopoiesis and cognition: the realization of the living*, Reidel, Dordrecht.

Nehaniv, C. L.: 1999, Meaning for observers and agents, *IEEE International Symposium on Intelligent Control / Intelligent Systems and Semiotics (ISIC/ISAS'99*, IEEE Press, pp. 435–440.

Olsson, L.: 2006, *Information self-structuring for developmental robotics: organization, adaptation and integration*, PhD thesis, School of Computer Science, University of Hertfordshire, UK.

Olsson, L., Nehaniv, C. L. and Polani, D.: 2005, Sensor adaptation and development in robots by entropy maximization of sensory data, *Proceedings of the 6th IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA 2005). IEEE Computer*, Society Press, pp. 587–592.

Oudeyer, P.-Y. and Kaplan, F.: 2004, Intelligent adaptive curiosity: a source of self-development, *in* L. Berthouze, H. Kozima, C. G. Prince, G. Sandini, G. Stojanov, G. Metta and C. Balkenius (eds), *Proceedings of the 4th International Workshop on Epigenetic Robotics*, Vol. 117, Lund University Cognitive Studies, pp. 127–130.

Pearl, J.: 2000, *Causality : Models, Reasoning, and Inference*, Cambridge University Press.

Polani, D., Martinetz, T. and Kim, J. T.: 2001, An information-theoretic approach for the quantification of relevance, *ECAL '01: Proceedings of the 6th European Conference on Advances in Artificial Life*, Springer-Verlag, London, UK, pp. 704–713.

Polani, D., Nehaniv, C. L., Martinetz, T. and Kim, J. T.: 2006, Relevant information in optimized persistence vs. progeny strategies, *Artificial Life X : Proceedings of the Tenth International Conference on the Simulation and Synthesis of Living Systems*, The MIT Press (Bradford Books), pp. 337–343.

Prokopenko, M., Gerasimov, V. and Tanev, I.: 2006, Evolving spatiotemporal coordination in a modular robotic system, *From Animals to Animats 9: 9th International Conference on the Simulation of Adaptive Behavior (SAB 2006)*, Vol. 4095 of *Lecture notes in computer science*, Springer, pp. 558–569.

Rezaeian, M. and Grant, A.: 2004, Computation of total capacity for discrete memoryless multiple-access channels, *IEEE Transactions on Information Theory* **50**(11), 2779–2784.

Schmidhuber, J.: 2006, Developmental robotics, optimal artificial curiosity, creativity, music, and the fine arts, *Connection Science* **18**(2), 173–187.

Shalizi, C. R.: 2001, *Causal Architecture, Complexity, and Self-Organization in Time Series and Cellular Automata*, PhD thesis.

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**, 91–95.

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

Shannon, C. E.: 1958, Channels with side information at the transmitter, *IBM Journal of Research and Development* **2**(4), 289–293.

Shepard, R. N.: 1984, Ecological constraints on internal representation: resonant kinematics of perceiving, imagining, thinking, and dreaming, *Psychological Review* **91**(4), 417–447. PMID: 6505114.

Slonim, N.: 2002, *The Information Bottleneck: Theory and Applications*, PhD thesis, Hebrew University.

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.

Sporns, O. and Lungarella, M.: 2006, Evolving coordinated behavior by maximizing information structure, *in* L. M. Rocha, L. S. Yaeger, M. A. Bedau, D. Floreano, R. L. Goldstone and A. Vespignani (eds), *Artificial Life X : Proceedings of the Tenth International Conference on the Simulation and Synthesis of Living Systems*, International Society for Artificial Life, The MIT Press (Bradford Books), pp. 323–329.

Steels, L.: 2004, The autotelic principle, *in* I. Fumiya, R. Pfeifer, L. Steels and K. Kunyoshi (eds), *Embodied Artificial Intelligence*, Vol. 3139 of *Lecture Notes in AI*, Springer Verlag, Berlin, pp. 231–242.

Still, S.: 2009, Information-theoretic approach to interactive learning, *EPL (Europhysics Letters)* **85**(2), 28005 (6pp).

Sun, X., Janzing, D. and Schölkopf, B.: 2008, Causal reasoning by evaluating the complexity of conditional densities with kernel methods, *Neurocomputation* **71**(7-9), 1248–1256.

Tatikonda, S.: 2000, *Control Under Communication Constraints*, PhD thesis, MIT.

Tatikonda, S. and Mitter, S.: 2009, The capacity of channels with feedback, *IEEE Transactions on Information Theory* **55**(1), 323–349.

Thomas, J.: 1987, Feedback can at most double gaussian multiple access channel capacity, *IEEE Transactions on Information Theory* **33**(5), 711–716.

Tishby, N., Pereira, F. and Bialek, W.: 1999, The information bottleneck method, *Proceedings of the 37-th Annual Allerton Conference on Communication, Control and Computing*, pp. 368–377.

Tononi, G. and Sporns, O.: 2003, Measuring information integration, *BMC neuroscience* **4**(1), 31+.

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

Touchette, H. and Lloyd, S.: 2004, Information-theoretic approach to the study of control systems, *Physica A* **331**, 140.

Uexküll, J. V.: 1934, A stroll through the worlds of animals and men, *in* K. Lashley (ed.), *Instinctive Behavior*, International Universities Press, New York.

Varela, F. J., Thompson, E. and Rosch, E.: 1993, *The Embodied Mind: Cognitive Science and Human Experience*, The MIT Press, Cambridge/MA.

Waddell, J., Dzakpasu, R., Booth, V., Riley, B., Reasor, J., Poe, G. and Zochowski, M.: 2007, Causal entropies–a measure for determining changes in the temporal organization of neural systems, *Journal of Neuroscience Methods* **162**(1-2), 320–332.

Watanabe, Y. and Kamoi, K.: 2002, The total capacity of multiple-access channel, *Proceedings of the 2002 IEEE International Symposium on Information Theory.*, p. 308.

Yang, S., Kavcic, A. and Tatikonda, S.: 2005, Feedback capacity of finite-state machine channels, *IEEE Transactions on Information Theory* **51**(3), 799–810.