Thermodynamics and the Structure of Living Systems

Nathaniel Virgo

Submitted for the degree of D.Phil.

University of Sussex

March, 2011

Declaration

I hereby declare that this thesis has not been submitted, either in the same or different form, to this or any other university for a degree.

Signature:

Acknowledgements

This thesis could never have existed without the support and guidance of my supervisor Inman Harvey. Many of the ideas in this thesis arose in conversations with colleagues, and I am particularly indebted to Simon McGregor, Mike Beaton, Matthew Egbert, Tom Froese, Alex Penn, Chrisantha Fernando, Xabier Barandiaran, Ezequiel Di Paolo, James Dyke, Lucas Wilkins, Nick Tomko, Manuela Jungmann and my colleagues at the Centre for Computational Neuroscience and Robotics for discussions that have formed the inspirational basis of this work.

Thermodynamics and the Structure of Living Systems

Nathaniel Virgo

Summary

Non-equilibrium physical systems, be they biological or otherwise, are powered by differences in intensive thermodynamic variables, which result in flows of matter and energy through the system. This thesis is concerned with the response of physical systems and ecosystems to complex types of boundary conditions, where the flows and intensive variables are constrained to be functions of one another. I concentrate on what I call negative feedback boundary conditions, where the potential difference is a decreasing function of the flow.

Evidence from climate science suggests that, in at least some cases, systems under these conditions obey a principle of maximum entropy production. Similar extremum principles have been suggested for ecosystems. Building on recent work in theoretical physics, I present a statistical-mechanical argument in favour of this principle, which makes its range of application clearer.

Negative feedback boundary conditions can arise naturally in ecological scenarios, where the difference in potential is the free-energy density of the environment and the negative feedback applies to the ecosystem as a whole. I present examples of this, and develop a simple but general model of a biological population evolving under such conditions. The evolution of faster and more efficient metabolisms results in a lower environmental energy density, supporting an argument that simpler metabolisms could have persisted more easily in early environments.

Negative feedback conditions may also have played a role in the origins of life, and specifically in the origins of individuation, the splitting up of living matter into distinct organisms, a notion related to the theory of autopoiesis. I present simulation models to clarify the concept of individuation and to back up this hypothesis.

Finally I propose and model a mechanism whereby systems can grow adaptively under positive reinforcement boundary conditions by the canalisation of fluctuations in their structure.

Submitted for the degree of D.Phil.

University of Sussex

March, 2011

Contents

List of Figures

Chapter 1

Introduction

Living systems — organisms and ecosystems — are physical systems far from thermal equilibrium and share many of their properties with this broader class of systems. This thesis concerns the study of living systems from a physical point of view, and in particular the attributes they have in common with other types of physical system. In this respect this research is very much in the spirit of Schrödinger (1944), Morowitz (1968, 1978), Schneider and Kay (1994) and other authors who have explored the continuity between life and physics.

The results in this thesis largely concern the response of complex physical and biological systems to a particular form of negative feedback, where the energy gradient that drives the system is a decreasing function of the system's overall activity. I argue that such negative feedback might have played vital roles in the origins and evolution of early life.

Because of its nature this thesis contains results relevant to non-equilibrium thermodynamics as well as to the study of living systems. In particular, I advance the theory behind a conjectured principle of maximum entropy production, an hypothesis which has arisen based on evidence from climate science and which also applies to systems under this type of negative feedback.

Additionally I argue that, under the right circumstances, some physical systems can respond to what might be called positive reinforcement feedback — an increase in the driving force when the system behaves in a particular way — by growing structures that cause the system to behave in the specified way. This has implications for biological development.

1.1 Negative Feedback in Ecological and Pre-Biotic Scenarios

A central theme of this thesis is the effect of what I call *negative feedback boundary conditions* upon physical and ecological systems. I will define this term more formally below, but in the context of an ecosystem it means that the amount of energy that can be extracted from the environment is a decreasing function of the overall rate at which it is used up. I will argue that this situation often arises naturally in ecological situations. As an example, consider a flow reactor, a container into which some concentrated food chemical is added at a constant rate. The contents of

the reactor are kept at a constant volume by allowing the container to overflow, removing whatever mixture of food and other molecules is present.

If there are no processes using up the food within the reactor then its concentration will build up to a high level. However, if a population of organisms is present that feeds on the food, converting it into some other molecules, then the concentration of food in the container will become lower. The steady-state concentration of food decreases with the overall rate at which it is used up, which is given by the number of organisms present, multiplied by the rate at which each individual is able to consume food. I will present a simple model of this situation, in which the organisms are modelled as thermodynamic machines whose energetic output is used to maintain their own structures, as well as to produce new individuals. In this model the overall rate at which the population consumes food increases over evolutionary time, with a concomitant drop in the free-energy density of the surrounding chemical mixture.

These ideas have an interesting set of implications for the origins of life. The first living organisms' metabolisms are likely to have been powered by extracting free energy from their chemical environment. There are many possible sources of this chemical energy, including photochemistry in the atmosphere of the early Earth, or geothermal sources. However, whatever the source, it seems reasonable to think that some of the molecules created would have been stable enough to build up to very high concentrations in the absence of any process to consume them, either globally or in localised areas such as lakes. Thus the availability of energy in the chemical environments of the early Earth could have been higher, perhaps much higher, than it is in any environment we observe today.

The simple model described above shows that a higher free-energy density in the early environment means that organisms (or proto-organisms) with slower, less efficient metabolisms are able to persist. This is because less material has to be processed in order to extract enough energy to maintain and re-produce the organism's structure. This suggests that early organisms could have had much simpler structures and metabolisms than today's. As the speed and efficiency of metabolisms increased over evolutionary time, the energy density of the environment would have dropped, making it impossible for all but the fastest metabolising and most efficient organisms to persist, leading to the very complex and specific molecular mechanisms found in modern cells.

The presence of negative feedback in the chemical environments in which life arose has another important implication, to do with what I call *individuation*, the splitting up of living matter into individual organisms. I argue that living organisms can be characterised as *individuated, precarious dissipative structures*. This notion is related to Maturana and Varela's (1980) theory of autopoiesis, as well as to the philosopher Simondon's (1964/1992) concept of biological individuation and Jonas' (1968) concept of biological individuality.

However, the concept of an individuated, precarious dissipative structure as I construe it applies not only to living organisms but to a wider class of phenomena. One example of a non-living structure that fits this description is a hurricane; another can be found in reaction-diffusion systems, a simple type of physical/chemical system that is easily simulated. With appropriate parameter settings, such models exhibit a "spot pattern" consisting of spatially concentrated regions

of autocatalytic chemical activity, separated by regions in which no, or very little, autocatalyst is present. Following the approach of Beer (2004), who analyses gliders in Conway's game of life in terms of autopoiesis, I argue that studying individuation in simple physical systems can help us to understand the concept in a more complete way than would be possible if living organisms were the only example available to study. This can be seen as an extension of the Artificial Life methodology from "life as it could be" (Langton, 1989) to "near-life as it is" (McGregor & Virgo, 2009).

Like the chemically driven ecosystems discussed above, these reaction-diffusion systems are powered by a continual input of a "food" chemical. A central result is that, when negative feedback of the form described above is applied to such systems, individuation can occur as a result. In effect, the negative feedback tunes the system's parameters to the region where individuated spots of autocatalyst can persist. This happens because, in the absence of any autocatalyst in the system, the availability of energy becomes very high. When an amount of autocatalyst is added it grows, increasing the rate at which food is used up, and thus decreasing the availability of energy at each point in the system. Eventually the energy availability drops to the point where a homogeneous concentration of autocatalyst cannot persist in the system, but persistence in a heterogeneous, patterned form is possible. This process continues until the pattern breaks up into spatially separate individuals, with each spot drawing in food from its surrounding area by diffusion.

This phenomenon of individuation under negative feedback can occur in reaction-diffusion systems whose chemical components form complex autocatalytic sets, and can result in the formation of individuals with a more complex structure than a single spot. I argue that it should occur more generally, in physical systems in which processes other than reaction and diffusion can occur.

If this is the case it suggests a novel hypothesis about the origin of life. The work of authors such as Kauffman (2000) has shown that large, complex autocatalytic sets are a natural consequence of the chemistry of non-equilibrium systems. If the free-energy density of an early environment was high enough, such an autocatalytic set would have been able to grow in concentration, spreading homogeneously throughout the environment. The energy availability would then have dropped due to the negative feedback process, until the autocatalytic set was no longer able to persist in a homogeneous configuration. Individuation is a possible result of such a situation, and I hypothesise that the first ancestors of modern organisms arose in this way. As an hypothesis about the origin of life this has the advantage of being a general principle, independent of more specific hypotheses about what type of molecules the original autocatalytic set and its energy source might have been made of. This generality makes it amenable to further experimental testing.

This hypothesis concerns the origins of metabolism, but says nothing about the origins of genetic material. One possibility is that information-carrying macromolecules were already part of the homogeneous autocatalytic set, but another possibility is that they arose after individuation as a result of evolution. In order for this to be the case the earliest proto-organisms must have had some form of heredity despite the lack of genetic material. I demonstrate that, even in reaction-diffusion systems with simple sets of possible reactions, a very limited form of heredity is possible.

In the example I give there are two types of individuated pattern, one consisting of spots with tails, and one consisting of tail-less spots. Each type is adapted to a different environment, and mutation from one form to the other can happen via the loss of a tail to produce a tail-less spot.

1.2 Non-Equilibrium Systems and Maximum Entropy Production

In this section I will describe the approach taken to non-equilibrium thermodynamics in this thesis and introduce the results pertaining to the Maximum Entropy Production Principle (MEPP). The latter is a hypothesis which arose from an empirical discovery by Paltridge (1979), who was able to reproduce features of the Earth's atmosphere to an impressive degree of accuracy using a simple model, by adjusting its parameters so that the entropy production was maximised. A similar technique was applied by R. D. Lorenz, Lunine, and Withers (2001) to the atmospheres of Titan and Mars. Although the principle currently lacks a widely accepted theoretical justification, some progress has been made towards a statistical-mechanical explanation by Dewar (2003, 2005a, 2005b). The work of Dewar and others suggests that maximisation of entropy production may be a physical principle akin to the second law of thermodynamics, but applicable to non-equilibrium systems in a steady state rather than to isolated systems. This opens up the exciting possibility that it could be used to make predictions about ecosystems in a manner similar to that suggested by H. T. Odum and Pinkerton (1955). The maximum entropy production principle should not be confused with Prigogine's (1955) *minimum* entropy production principle, which is a different type of principle with a different domain of applicability.

The MEPP is an exciting hypothesis because it suggests a powerful and general methodology by which complex, far-from-equilibrium systems can be studied and understood. However, there is a conceptual problem with its application, which is that it makes different predictions depending on which parts of a system it is applied to. When modelling atmospheres one must apply the principle to a set of processes that includes the transport of heat from the planet's tropical to its polar regions, but which excludes the conversion of high-energy photons from the sun into heat in the atmosphere (Essex, 1984). If one instead includes both processes, the MEPP predicts implausibly fast rates of atmospheric heat transport which do not match the observed values. Dewar's (2003) statistical-mechanical argument claims to justify this choice of system boundary, but I will argue that it does not.

I will present an argument in favour of the MEPP which, like Dewar's, builds on the work of Jaynes (e.g. 1957a, 1965, 1980, 1985, 2003) and suggests a principled way to resolve this system boundary problem. This represents a step towards applying the principle to ecosystems, since the question of where to draw the boundary must be answered before the principle can be applied to any system.

The maximum entropy production principle, if it turns out to be a genuine physical principle, can be used to make specific numerical predictions about systems subjected to negative feedback boundary conditions. I discussed some of the consequences of such conditions for ecosystems above, and will now define the concept more formally.

In order for a physical system to be sustained out of thermal equilibrium indefinitely it must

continually exchange matter and/or energy with another system. Such an exchange is characterised by two types of quantity: the rate of flow of matter or energy across the boundary, and what I will refer to as the "temper" of the flow, for want of an established term. This is the per-unit entropy of the flowing substance. For a flow of energy the temper is $1/T$, the reciprocal of temperature, whereas for an exchange of some chemical species it is given by $-\mu/T$ (the negative chemical potential divided by temperature)[1]. The concept can be generalised to flows of other types of quantity, such as volume, momentum or charge.

In order for a flow thorough the system to be sustained, there must be a difference in tempers between different parts of the system. A difference in $1/T$ between parts of a system may cause a flow of heat, whereas chemical reactions and flows of particles are enabled by differences in chemical temper $-\mu/T$. As matter or energy flows from lower to higher temper its entropy increases (by definition), so the statement of the second law applied to non-equilibrium systems in the steady state is that the sum of the products of the flows with their associated tempers must be positive.

In near-equilibrium thermodynamics a difference in tempers is called a "thermodynamic force", but I consider this a misnomer. For a simple system such as a electrical resistor it might make sense to say that the voltage difference "forces" a flow of current, although the units are not the same as for a mechanical force. However, for a terrestrial ecosystem the temper difference is between the radiation field of the sun (at approximately 6000 K, or a temper of 1.6×10^{-4} K^{-1}) and the outgoing radiation field of the ecosystem, at around 300 K. It makes more sense to say that the solar radiation enables the growth of plants than that it "forces" the flow of energy through the system, because in the absence of any plants this flow will not take place.

Ordinarily in the study of non-equilibrium thermodynamics one considers some or all of the tempers across the system's boundary to be fixed and asks what will be the resulting rates of flow, or else one holds the flows fixed and asks the value of the temper difference. This can be generalised by instead allowing both the flows and the tempers to vary, but constraining them to have some particular functional relationship to one another.

Negative feedback boundary conditions are a special case of this, where a difference in tempers is constrained to be a decreasing function of the associated flow, dropping eventually to zero for a fast enough flow. In the ecosystem example discussed above, the difference in chemical temper between the food chemical and the products to which it is converted is a decreasing function of the overall rate at which the conversion takes place; in the Earth's atmosphere the Stefan-Boltzmann law constrains the inverse temperature difference to be a decreasing function of the overall rate of heat transport from the tropics to the polar regions. In both cases the relationship between temper and flow is a property of the boundary conditions and not of the processes that cause the flow (a population of organisms, and a network of convective heat transport processes, respectively).

Boundary conditions of this functional form permit a variety of possible steady-state behaviours: the temper difference can be small and the flow large, or vice versa, with a continuous range of values permitted in between. Applying the maximum entropy production principle to

[1] I take the term "temper" from Tribus (1961), who uses it to refer to $1/T$ only.

such a situation means predicting that the observed steady-state rate of flow will be close to the one which maximises the entropy production (flow times temper difference). The negative feedback boundary conditions ensure that the entropy production has a definite peak for some particular value of the flow.

The advantage of this approach is that it effectively ignores all the complex details of the processes that occur inside the system. Within a non-equilibrium system there are, in general, many circulating flows that do not cross the boundary, but these do not need to be specifically modelled in order for the MEPP to be applied. Quantities from equilibrium thermodynamics such as temperature may not be well defined in the interior of complex far-from-equilibrium structures such as living organisms, but the MEPP can still be applied as long as these quantities are well-defined on the boundaries of the system under consideration.

According to the theories of Dewar (2003, 2005a) and of this thesis, this maximisation of entropy production is somewhat analogous to the maximisation of entropy in isolated physical systems. As in equilibrium thermodynamics, the maximisation is performed subject to whatever constraints are acting on the system. In equilibrium thermodynamics these constraints usually take the form of conservation laws. In non-equilibrium systems the boundary conditions are one kind of constraint but there may be others. However, unlike the entropy in equilibrium systems, the entropy production is not expected to increase monotonically over time (in particular, if the system is not in a steady state then its entropy production can temporarily be higher than the maximum possible steady-state value).

This means that, if a system is observed in a steady state other than that predicted by the MEPP, it does not invalidate the maximum entropy production principle but instead indicates the presence of additional constraints that were not taken into account when making the prediction. The evolutionary model presented in this thesis does not exhibit maximisation of entropy production, and so the question of whether the constraints that act upon ecosystems are such that the MEPP can be practically applied to make predictions about their behaviour is left open.

1.3 Positive Reinforcement Feedback in Physical Systems

The concept of functional boundary conditions can be extended beyond negative feedback. One interesting possibility is that a system could be "rewarded" by increasing the gradient (or the flow) when the system performs some externally specified task. I argue that some types of purely physical system will increase their ability to perform the task under these conditions, via a mechanism whereby (dissipative) structures within the system fluctuate over time. In some systems, fluctuations that increase the gradient will tend to become canalised into more permanent structures. In particular I propose this as an explanation for the results of Pask (1958), who was able to grow an electrochemical "ear" capable of detecting sound using this type of technique.

The idea also has implications for biological growth and development, because it provides a way in which structures can adapt to particular circumstances even if these circumstances have not been encountered during the organism's evolutionary history. In this respect it can be seen as a form of learning, but in the domain of form rather than behaviour.

1.4 Major Contributions

The major contributions of this thesis are as follows:

1. The statistical argument in favour of the maximum entropy production principle and the resulting clarification of its domain of application, presented in Chapter 3.
2. A technique for explicitly modelling the interaction of evolutionary and population dynamics with thermodynamic constraints on organisms' metabolism and the system's boundary conditions, developed in Chapter 4.
3. The hypothesis, developed in Chapters 4 and 5, that the chemical energy density of prebiotic environments was much higher than in any modern environment, allowing proto-organisms to persist with much simpler structures and metabolisms than those of modern organisms.
4. The methodological proposal in Chapter 5 of using simple dissipative structures such as spot patterns in reaction-diffusion systems to illustrate and investigate certain properties of living systems, particularly the maintenance of spatially distinct individuals.
5. The result in Chapter 5 that, in such a reaction-diffusion spot model, the range of parameter space in which individuation occurs is much greater when a negative feedback is applied between the system's overall activity and the thermodynamic gradient that drives it.
6. The demonstration of dissipative structures that exhibit the simplest form of limited heredity with variation in Chapter 5

1.5 Structure of the Thesis

The thesis proceeds roughly from the most general to the most specific results, progressing from pure physics, through the physical behaviour of ecosystems and the notion of biological individuality, and ending with the notion of adaptive growth. Chapter 2 gives an overview of the background to this thesis from various disciplines. Additional, more detailed treatments of some aspects of the background material are given in each chapter.

Chapter 3 concerns the response of complex physical systems to non-equilibrium boundary conditions, regardless of whether life plays any role. The results concerning the maximum entropy production principle are presented in this chapter. This chapter also serves to make formal the concept of negative feedback boundary conditions. The results of the following chapters are independent of the arguments given in this chapter.

In Chapter 4 I focus specifically on the thermodynamics of ecosystems as physical systems, developing a simple model of population and evolutionary dynamics in which thermodynamic constraints on individuals' metabolisms are explicitly modelled. One of the primary goals of Chapters 3 and 4 is to investigate whether the Maximum Entropy Production Principle can be applied to ecosystems. I cannot give a definitive answer to this question but I nevertheless believe the models and arguments I have developed represent significant progress towards one.

Chapter 5 is concerned with the phenomenon of individuation in biological and physical systems. I elucidate this concept and discuss its relationship to autopoiesis, using spot patterns in reaction-diffusion systems as an example. I demonstrate the phenomenon of individuation under

system-wide negative feedback in simple and more complicated reaction-diffusion systems, and discuss the relevance of these results for the origin of life. I also use reaction-diffusion systems to demonstrate that a very limited form of heredity is possible in dissipative structures that contain only a few types of molecule.

Chapter 6 concerns a proposed mechanism by which some physical systems under some circumstances may be able to respond to energetic "rewards" for behaving in a particular way, through a process of adaptive growth whereby fluctuations in structure that increase the energy flow are more likely to become fixed. The conclusions of the thesis are summarised and discussed in Chapter 7.

1.6 Relationship to my Previous Publications

Parts of this thesis are based on previously published work. In particular, Chapter 3 is based on (Virgo, 2010), with few changes since the previously published version; Chapter 4 is based in part on (Virgo & Harvey, 2007) but has been substantially expanded since its prior publication; and Chapter 6 is based on (Virgo & Harvey, 2008a), with a few significant modifications to the text. The research upon which Chapter 5 is based has been published previously in abstract form (Virgo & Harvey, 2008b). A full manuscript based on the text of Chapter 5 is in preparation, to be submitted for journal publication.

In addition to these there are a number of my previous publications that are relevant to this work. The methodology of looking for life-like properties such as individuation in the physical world is introduced and argued for in (McGregor & Virgo, 2009), and much of this thesis can be seen as a continuation of that work.

(Virgo, Egbert, & Froese, 2009) discusses the interpretation of autopoiesis, and in particular points out the need to distinguish between the physical membrane of a cell and the more conceptual boundary that determines which processes are to be considered part of the operationally closed network that constitutes the system. This paper argues that virtually all organisms cause processes to occur outside their physical boundary that they also rely on, and hence the operationally closed network should be seen as extending far beyond the physical boundary. This is important background for the discussion in Chapter 5, where I examine membrane-less (but nevertheless spatially individuated) reaction-diffusion structures from an autopoiesis-like point of view.

Finally, (Virgo, Law, & Emmerson, 2006) is an investigation of various goal functions in ecology, using a population-dynamic model of ecological succession. The concepts in this paper are relevant to the discussion of the maximum entropy production principle in an ecological context in Chapter 4.

Chapter 2

Background: Thermodynamics and Its Application to Living Systems

2.1 Introduction

This chapter collects background material on thermodynamics and its application to living systems, and on the development of the hypothesised Maximum Entropy Production Principle (MEPP) in climate science.

I have departed from tradition slightly in making this chapter a high-level overview of the background topics, which are then covered in more detail in the individual chapters. This is because the material in this thesis is not based on one single body of work. It arose from a synthesis of several trains of thought from several disciplines, and this background chapter must therefore survey a fairly diverse collection of material. Most of the chapters rely on some background material that is not as relevant to the other chapters. In those cases I have chosen to present that material in the chapter concerned, rather than collecting it all here. In particular, the planetary atmosphere model of R. D. Lorenz et al. (2001) is given a detailed treatment in Chapter 3, along with Essex's (1984) objection to the maximum entropy production principle. The theory of autopoiesis (Maturana & Varela, 1980) is discussed in more detail in Chapter 5, as is the background on reaction-diffusion systems, which are used in the model presented in that chapter. Chapter 6 concerns the cybernetic experiments of Gordon Pask (1958, 1960, 1961), which are described therein.

This chapter introduces the thermodynamic concept of entropy in Section 2.2, and discusses its application to organisms (Section 2.3) and to ecosystems (Section 2.4). In particular, the history of "extremum function" or "goal function" hypotheses in ecology is discussed in Section 2.4.2. The hypothesised maximum entropy production principle is discussed in Section 2.5.

In Section 2.6 I present some notes on the structure of thermodynamics. In addition to presenting background material, the goal of this section is to explain how, with some small changes in notation, thermodynamics can be expressed with entropy rather than energy as the central concept. This section also introduces some terminology that will be used in later chapters. In Section 2.7 I

summarise Jaynes' approach to statistical mechanics, which is important to the results of Chapter 3 and their discussion in later chapters.

2.2 Entropy and the Microscopic/Macroscopic Distinction

Before we discuss the application of thermodynamics to life we must first discuss a central concept in thermodynamics, namely entropy. This chapter cannot serve as a complete introduction to thermodynamics for the uninitiated reader (an excellent brief introduction can be found in (Jaynes, 1988)), but the concept of entropy is so fundamental that it is worthwhile to cover it here. Additionally, I wish to clarify exactly what I mean by the distinction between "microscopic" and "macroscopic" systems, since this distinction can be approached in more than one way. I will make further comments on the structure of thermodynamics later in this chapter.

To understand entropy one has to understand the distinction between microscopic states and macroscopic states of a system. The distinction between a system's microscopic and macroscopic states is due primarily to the work of Boltzmann and Gibbs. Some significant clarifications of the concept were provided by Jaynes (e.g. 1957a, 1992), whose work has been a major influence on the approach to thermodynamics taken in this thesis. In particular, the definition of a macrostate given below is Jaynes' version of the concept, and differs somewhat from more traditional versions.

"Macroscopic" literally means "large," in contrast to "microscopic," meaning very small. However, I will use the terms in a slightly different way. The essential difference between a microscopic state (microstate) and a macroscopic state (macrostate) of a system is that the former is a property of the system alone - the microstate is the underlying state which determines the system's behaviour - whereas a macrostate is a property of what we, as observers, are able to measure and predict about the system. For example, in an electrical system the microstate would include the precise position of every electron. Quantities such as voltage and current, on the other hand, are properties of the macrostate. Thus the word "microstate" usually does refer to a description of a system on a much smaller scale than the macrostate.

In classical physics the microstate of a physical system can be thought of as a vector composed of the precise positions and velocities of every elementary particle in the system, along with the precise value of any relevant field at every point in space. In quantum mechanics the microstate is a vector of complex numbers whose correspondence to familiar physical quantities is a little more abstract, but in either case the assumption is that if the system is completely isolated from the outside world then its dynamics are completely determined by its microstate. In other words, if one had complete knowledge of an isolated system's microstate, along with knowledge of the laws of physics, then one could predict the system's future behaviour with complete success. The system behaves as a deterministic and reversible dynamical system, with the microstate playing the role of the state variable.

The concept of a system's macroscopic state arises because we do not usually have such complete knowledge of a system's microscopic state. The example usually given is that of a gas in a sealed chamber. There are many properties of the gas that can be known or measured with standard physical instruments: its mass, its volume, which types of molecule it is made of, how much force

it exerts on its container, etc. However, if we have a complete knowledge of all these parameters we are still very far away from knowing the system's microstate — that is, the precise position, configuration and velocity of every molecule of gas. This is because there are many possible microstates which, when measured, would yield the same set of values for the volume, composition, temperature, etc. Knowledge of a macrostate of a system therefore represents incomplete knowledge of its microstate.

An important thing to note about macrostates in this sense is that they depend not only on the system itself but also (unlike the microstate) on which instruments are available to measure it. See Jaynes (1992) for an exploration of this subtlety.

There are many microstates compatible with a given macrostate. Suppose we have measured the macrostate of a particular system (using some particular set of measuring instruments). Let the number of microstates[1] compatible with this microstate be W. In other words, we know, based on our macroscopic measurements, that the system is in one of W possible microstates. Thus W measures an experimenter's uncertainty about the microstate, given the macroscopic measurements available.

The most important quantity in thermodynamics, the entropy, is defined in terms of the logarithm of W:

$$S = k_{\mathrm{B}} \log W, \tag{2.1}$$

where $k_{\mathrm{B}} \approx 1.38 \times 10^{-23}$ JK^{-1} is Boltzmann's constant, which is present in the definition of entropy for purely historical reasons (as will be seen below, it would make more sense to define the units of temperature in terms of the units of entropy than the other way around). S represents a measure of the uncertainty about the microstate with the property that the entropy of two independent systems is the sum of the entropies of the individual systems[2].

An alternative unit of measurement for entropy, which I will often use in this thesis, is the *bit*. Entropy in bits is given by $S = \log_2 W$, so that a system with two equally possible microstates has an entropy of one bit. The advantage of using bits as the unit of measurement for entropy is that it makes clear its relationship to information: the entropy is to be conceived as the number of bits of information one would have to learn about a system in order to know its precise microstate.

An extremely important property of the entropy is that, as a system's macrostate evolves over time, its entropy never decreases. The reasons for this are rather subtle and still debated; one explanation is discussed in Section 2.7.5. This rule breaks down for small systems and over short time scales, but the probability of such fluctuations can be calculated and the rule always applies on average. This is the so-called second law of thermodynamics.

The rate at which entropy increases can be vastly different for different systems. For example, a solid iron object may take centuries or more to rust, but if the same mass of metal is finely

[1] In classical mechanics, this "number of microstates" should be replaced by a continuous volume of phase space. In quantum mechanics, finite systems tend to have a discrete number of possible microstates.

[2] Equation 2.1 is due to Boltzmann. The concept of a macrostate was made yet more precise by Gibbs (whose arguments were later made much more clear by Jaynes), who considered macrostates as defining probability distributions over the set of possible microstates. In this case the entropy formula is extended to $S = -k_{\mathrm{B}} \sum_i p_i \log p_i$, where p_i is the probability of the i^{th} microstate, conditional on the current macrostate, and $0 \log 0$ is understood to be 0. This formula reduces to Boltzmann's in the case where all the probabilities are equal to either 0 or $1/W$.

powdered and mixed with air it can explosively combust. The change in entropy in each case is similar (because the products and reactants and the total heat given off are essentially the same) but the time scale on which the process occurs is dramatically different.

Another important property of entropy is that under many circumstances it can be seen, very roughly, as a measure of how disorganised, or unstructured, a system is. Consider an experimental set-up in which two different gases occupy two sides of a chamber, separated by a thin membrane. If the membrane is removed, the two gases will immediately start to mix due to diffusion, resulting eventually in a homogeneous mixture of the two gases, occupying the whole chamber. The entropy of the system increases continually as this process occurs; the macrostate in which the two gases are completely mixed is the one with the highest entropy. One can also say that, out of the macrostates observed in such an experiment, the one in which the two gases are separate is the one with the greatest degree of structure.

This rule of entropy-as-disorder does not always hold (or at least, it depends on one's definition of disorder). In particular, a thoroughly shaken mixture of oil and water might intuitively seem less ordered than a macrostate in which the oil is separated into a distinct layer on top of the water, yet the latter is the higher entropy macrostate, and such a mixture will eventually separate out in this way. Nevertheless, it is a general rule that complex structures represent low-entropy configurations of matter. The tendency of entropy to increase means that such structures have a tendency to decay.

The rule that entropy must always increase applies only to isolated systems — those which cannot exchange heat, matter or any other quantity with the outside world. Living organisms, of course, are not isolated, since they continually exchange matter and energy with their environment, by eating, breathing and photosynthesising. On a larger scale, the biosphere as a whole is not isolated, as it continually takes in solar radiation while radiating its own heat into space. The next two sections are concerned with the application of thermodynamics to organisms and to ecosystems, respectively.

2.3 Entropy and Organisms

Living organisms are complex structures and therefore have a low entropy. We can get a quantitative feel for this from Morowitz' (1968) estimate of the probability of a living *Escherichia coli* cell forming spontaneously due to a thermal fluctuation in an equilibrium system (that is, a chemical system consisting of the chemical elements that make up an *E. coli* cell in the correct quantities, but with no source of energy flow). Morowitz calculates this value to be approximately $10^{-10^{11}}$. An equivalent way to state this is that, for every arrangement of the atoms that makes up an *E. coli* cell there are $10^{10^{11}}$ possible arrangements of the same atoms that do not make up such a cell. Applying Boltzmann's formula (Equation 2.1 above), this corresponds to an entropy difference of about 3.2×10^{-12} JK^{-1}, or about 3.3×10^{11} bits, between an *E. coli* cell and its constituent matter in the equilibrium state.

For comparison, a similar difference in entropy is involved in heating 2.3×10^{-13} kg of water from 300 K to 301 K. (The per-kilogram entropy change is given by $c_p \ln(T_{\text{final}}/T_{\text{initial}})$, where c_p

is the specific heat capacity, about 4181 $JK^{-1}kg^{-1}$ for water.) The mass of an *E.coli* cell is of the order 10^{-15} kg, so the entropy change in changing matter from an inanimate to a living state is roughly 200 times greater than the entropy change involved in heating an equivalent weight of water by one degree.

Some of this entropy difference arises from the physical structure of the cell. Structure is a macroscopic property and can be thought of as a set of constraints on the system's microscopic state. On a microscopic level one can only talk about state: there is an atom here, another there and so on, and the concept of structure is not relevant. But moving to a macroscopic level and throwing away some information we are able to make structural statements such as "there is a membrane here, beyond which the concentration of this compound is higher." This statement is macroscopic: there are many possible ways in which the atoms could be arranged which are compatible with the concentration being higher behind the membrane. We can think of these as a vast number of hypotheses about the system's (unobservable) microscopic state. These hypotheses can in principle be enumerated, and the logarithm of their number is the entropy associated with the structure.

However, the vast majority of the entropy difference between a cell and its constituent matter in equilibrium is due to the formation of chemical bonds required to produce the amino acids of which it is composed. Values of a similar order of magnitude to Morowitz' estimate should therefore be expected for all types of living cell. (For cells of a different size from *E. coli* the figure scales linearly with size.)

The low value of Morowitz' estimate for the probability of a cell forming due to fluctuations in an equilibrium indicates that the first living cells did not arise from such fluctuations. The Earth is not an isolated system — it is held out of equilibrium by the flow of incident sunlight — and the formation of complex structures in such non-equilibrium systems is part of the subject of this thesis. The origins of living organisms in particular will be discussed in Chapter 5.

The entropy of an isolated system increases until it reaches a maximum, at which point we say that the system has reached a state of *thermodynamic equilibrium*. The low entropy of living systems means that they are very far from this equilibrium. When organisms grow or reproduce, the amount of this low-entropy matter increases, and so it might appear at first sight that the growth of biological populations runs counter to the second law of thermodynamics. The resolution to this was apparent to Boltzmann (1886), but was clarified and popularised by Schrödinger (1944).

Schrödinger noted that living organisms are not isolated systems. Instead they rely on an environment that is far from thermodynamic equilibrium. If the entropy of one system decreases while that of another increases by a greater amount, the entropy of the combined system increases overall. An organism can therefore keep its own entropy at a low value by increasing the entropy of its environment at a greater rate. Plants achieve this by absorbing high-frequency (and thus low entropy) photons from sunlight, converting them ultimately to high-entropy heat. Animals take in oxygen and food from their environment, converting them into a higher-entropy combination of carbon dioxide and other waste products. Chemotrophic organisms similarly increase the entropy of their environment by converting chemical substances into a higher-entropy form.

Schrödinger described this overall process as organisms "feeding on negative entropy." The intuitive picture is that in order to counter the continual production of entropy in their own tissues and keep their entropy at a low level, organisms must take in "negative entropy" from their environment[3]. The idea of a low-entropy structure being maintained by continually increasing the entropy of the environment has been central to all subsequent applications of thermodynamics to biology.

This idea was made clearer by Prigogine, whose work on what he called "dissipative structures" (Prigogine, 1978) had an important impact on the development of non-equilibrium thermodynamics. Loosely speaking, a dissipative structure is one that works in the way Schrödinger described, by taking in low-entropy exporting high-entropy matter or energy. For Prigogine, this included not only living systems but also oscillating chemical reactions, weather systems and a wide variety of other phenomena. The thermodynamic similarity between living organisms and other types of dissipative structure will play an important role in this thesis.

Prigogine's formalism of non-equilibrium thermodynamics was characterised by the equation

$$\mathrm{d}S = \mathrm{d}_i S + \mathrm{d}_e S, \tag{2.2}$$

which is supposed to apply to any system, whether it be an organism, an ecosystem, or any non-biological physical system with a well-defined boundary. The equation indicates that the rate of change $\mathrm{d}S/\mathrm{d}t$ of a system's entropy (which may be positive or negative) is equal to the rate at which the system produces entropy $\mathrm{d}_i S/\mathrm{d}t$ (which must be positive), plus the rate at which it exchanges entropy with its environment, which can give rise to a positive or negative term $\mathrm{d}_e S/\mathrm{d}t$.

Entropy can thus be created, and once created it can be moved between systems, becoming more or less concentrated. In this way it can be thought of as a compressible fluid which can be created but not destroyed. However, Jaynes (1980) gives a convincing argument that this picture is only an approximation, and in systems sufficiently far from thermodynamic equilibrium, entropy cannot correctly be conceived of in this way. Nevertheless it is often useful to keep the compressible fluid picture of entropy in mind.

In addition to Morowitz' estimate of the entropy of a living cell it is also useful to have an idea of the rate at which living systems produce entropy, i.e. the value of $\mathrm{d}_i S/\mathrm{d}t$ for a living organism. Aoki (1989) calculates an entropy production rate of 0.259 $\mathrm{JK}^{-1}\mathrm{s}^{-1}$ (or 2.7×10^{22} $\mathrm{bits} \cdot \mathrm{s}^{-1}$) for a particular naked human subject under experimental conditions. This is calculated by measuring the rate at which heat and chemical substances (whose entropy can be calculated) enter and leave

[3] In a footnote in a later edition of his book, Schrödinger comments that his argument about negative entropy was met with "doubt and opposition from physicist colleagues," and says that it would have been better to let the discussion "turn on free energy instead." However, free energy (as will be seen in Section 2.6.1) is just a formal way to reason about entropy when a system is connected to a "heat bath," holding its temperature constant. In Schrödinger's footnote it is evident that he is considering this heat bath to be separate from the system and its environment, so that the total system is divided into three entities: the system of interest, its environment, and the heat bath. Schrödinger then conceded that his argument should be seen in terms of the decreasing free energy of the system and its environment, but not of the heat bath. However, if one counts the heat bath as part of the system's environment then it is correct to speak of the total entropy of the system and its environment increasing rather than the total free energy decreasing. Seen in these terms, Schrödinger's argument was correct in its original form.

the subject (the experiment measured the exchange of O_2, CO_2 and H_2O but not food and faeces, since the subject did not eat during the experiment). The figure thus produced is comparable to the rate at which entropy is produced by a 75 W light bulb operating at room temperature.

However, unlike Morowitz' figure, this figure cannot be scaled to other organisms. Organisms' metabolisms differ widely in the rate at which they operate, and the rate of entropy production is heavily dependent on the organism's metabolic rate, as well as upon how it operates and what it uses for food. Even in humans the rate of entropy production varies substantially depending on factors such as whether exercise has been performed recently (Aoki, 1990) and the subject's age (Aoki, 1991).

It is worth mentioning Maturana and Varela's (1980) theory of autopoiesis in this context. I will discuss this only briefly here because it is covered in some depth in Chapter 5. The theory of autopoiesis has multiple interpretations, but the basic idea is that living organisms can be thought of as a network of processes, the result of whose action is to continually re-generate itself, both as a network of processes and as a concrete "unity", or individual, in space. Maturana and Varela's work included very little about physical and thermodynamic constraints on such networks of processes. However, later authors such as Moreno and Ruiz-Mirazo (1999) and Ruiz-Mirazo and Moreno (2000) have developed the theory in a direction that does include such constraints.

A closely related idea is Kauffman's (2000) concept of the "work-constraint cycle". Kauffman conceives of living organisms as thermodynamic machines consisting of physical constraints that cause the extraction of thermodynamic work from their environment. This work is in turn used to re-generate the constraints that make up the machine. I will give a very similar characterisation of organisms in Chapter 4.

2.4 Thermodynamics and Ecosystems

Like organisms, ecosystems are physical systems and must obey the same thermodynamic restrictions as any other physical system. However, throughout the history of ecology there have been suggestions that, in addition to this, ecosystems act so as to maximise some thermodynamic "goal function" over time. If such a general law exists it will allow predictions to be made about ecosystems without the need to know the precise details of every species and process within them, which would clearly have a high degree of practical usefulness. However, such principles have tended to lack both a solid theoretical justification and convincing empirical evidence. This thesis will attempt to address this by giving a theoretical justification of the Maximum Entropy Production Principle in Chapter 3.

2.4.1 Material cycling in Physical Systems

A great inspiration for this work is that of Morowitz (1968, 1978). These works focus largely on the concepts of the flow of energy and the cycling of matter in ecosystems. Importantly, Morowitz showed that material cycling is a natural consequence of an imposed energy flow on any physical system. This leads to the view that the cycling processes that are so conspicuous in ecosystems lie on a continuum with the cycling processes that occur spontaneously in driven abiotic systems.

Morowitz (1968) gave a very simple example of this abiotic cycling that is worth repeating. Let us imagine a chemical reaction $A \leftrightarrow B$ that takes place in the gas phase. At chemical equilibrium the two species will both be present, each in a certain proportion. We may imagine that this gas is held in two chambers, call them 1 and 2, separated by a permeable membrane through which both A and B can pass. If the two chambers are held at the same temperature then A and B will be present in the same proportion in each chamber.

However, we now imagine that chamber 1 is held at a higher temperature than chamber 2. This involves continually heating chamber 1 and cooling chamber 2, because heat flows from the hotter chamber to the cooler one. The proportion of each species in chemical equilibrium is in general dependent on the temperature. Thus, if we imagine for a moment that the membrane is not porous, the ratio of the two species in chamber 1 will be different from from that in chamber 2. Let us say that species A happens to be present in a higher proportion in chamber 1 than in chamber 2. Now, if we again make the membrane porous, this will result in a flow of A from chamber 1 to 2, and a corresponding flow of B in the opposite direction.

Because of the flows of gas through the membrane, neither chamber quite reaches chemical equilibrium, but because of the on-going chemical reactions the concentrations on each side of the membrane never reach equilibrium either. The flows and the reactions therefore take place continually, driven by the flow of energy through the system. The result is a cycle: a molecule of B, starting out in chamber 1, will tend to undergo a reaction to form a molecule of A, which will tend to flow through the membrane into chamber 2 where the concentration of A is lower. But now it will tend to react to form a molecule of B again, and then to flow back to chamber 1. All of these transitions can happen in reverse, but they have a definite tendency to occur in the directions indicated. The presence of this macroscopic cycling behaviour increases the rate at which heat is transported across the system, because heat is absorbed and released in the reactions, and thus also increases the production of entropy.

Another example of material cycling in a system driven by energy flow is convection. Schneider and Kay (1994) used a particular type of convective flow — so-called Bénard cells — as a metaphor for ecosystems. Bénard cells are convection cells that form in regular (usually hexagonal) patterns when a thin layer of fluid is heated from below. Schneider and Kay compare the difference between convection cells and diffusive flow to the difference between an abiotic system and an ecosystem: in both cases the more structured state has the greater flow of energy and produces the most dissipation (entropy production). The authors hypothesise a general tendency of physical systems to "utilize all avenues available to counter the applied gradients", with the implication that that life exists because it provides an efficient way to dissipate the gradient applied to the Earth by the temperature difference between incoming solar radiation and the outgoing thermal radiation.

2.4.2 Extremum Functions in Ecology

Part of this thesis (Chapter 3 and part of Chapter 4) concerns the possibility of applying a principle used in climate science known as the Maximum Entropy Production Principle (MEPP) to ecosys-

tems. The history and formulation of the MEPP will be described below, but first I will describe some precursors to the idea that have arisen within the field of ecology.

These ideas have a lot in common with the development of the Maximum Entropy Production Principle in climate science. Indeed, Dewar (2003, 2005a) explicitly suggests that the MEPP can be applied to ecosystems; and Schneider and Kay (1994) cite Paltridge (1979) in justifying their claims about ecosystems. However, in discussing the subject informally with ecologists one finds that they tend to be much more sceptical of such ideas than climate scientists.

One suspects that there are two main reasons for this cultural difference. The first is that, despite their long history, extremum principles have provided few if any experimentally verifiable numerical predictions within the field of ecology. This is in stark contrast to the convincing predictions made in climate science using such methods by Paltridge and by R. D. Lorenz et al. (2001). The second is that the idea of an ecosystem as a whole acting so as to maximise a single variable is hard to reconcile with most ecologists' intuitions about the behaviour of evolution by natural selection.

The notion that ecosystems act so as to maximise dissipation (or some other related quantity) is an old one in ecology, though it has remained controversial throughout its history. Perhaps its earliest form is Lotka's (1922) Maximum Power Principle. Lotka claimed that there would always be evolutionary pressure towards greater energy use, since organisms that were more efficient at gathering and using energy would have an advantage over others. I will work through a similar argument in Chapter 4. Lotka himself later drew back from the idea, warning that "a prematurely enunciated maximum principle is liable to share the fate of Thomsen and Berthelot's [incorrect] chemical 'principle of maximum work.' " (Lotka, 1925).

Lotka's maximum power idea was built upon by H. T. Odum and Pinkerton (1955), who demonstrated that in many situations, both in ecosystems and in engineering, there is an optimal rate at which some process should proceed, in order to extract the greatest amount of work. As an example, consider extracting work using a heat engine placed between a cold reservoir of constant temperature (e.g. the outside world) and an imperfectly insulated hot reservoir to which heat is supplied at a constant rate. If no heat is extracted from the hot reservoir its temperature will increase until the rate of heat flow through the insulation into the cold reservoir balances the rate at which it is added to the hot reservoir. No work is being performed at this point. Let us now imagine that we start to extract heat from the hot reservoir at some constant rate, in order to perform work. As the rate of heat extraction increases, so does the rate at which work can be performed. However, the increased rate of heat extraction causes a drop in the steady-state temperature of the hot reservoir, which decreases the efficiency of the heat engine. Therefore there exists a rate of heat extraction beyond which the rate at which work can be extracted will start to decrease. In the extreme, if heat is extracted at the same rate it is added to the reservoir then the reservoir's temperature will equilibrate with that of the outside world and the rate at which work can be extracted will again drop to zero. We will see an ecological example of such a trade-off in Chapter 4.

It may be noted that if once the work has been performed it eventually becomes heat (via pro-

cesses such as friction) and is returned to the outside world then this maximum in work extraction is also a maximum in entropy production, since the entropy production due to the work-extracting proces and the downstream processes that use the work it produces is equal to the rate of work extraction divided by the (constant) outside temperature. The entropy produced by the diffusion of heat through the insulation must be excluded in the example above in order for this to be the case.

Odum and Pinkerton's theory suffers from a problem which haunts many extremum function approaches to ecology: it demonstrates that a maximum exists but offers no argument for why the ecosystem should reach a state for which the maximum is attained. This gives the hypothesis a teleological sort of character: it is easy to read Odum and Pinkerton as saying that ecosystems behave in particular ways *because* this leads to a maximisation of the work extraction rate. The maximum rate of work extraction is perhaps, in some sense, optimal for the ecosystem as a whole but, as we will see in Chapter 4, this does not mean that it is optimal for individual organisms, and selfish evolution can drive the system toward states where energy is extracted faster than the optimum. In Chapter 3 of this thesis I will propose a statistical-mechanical explanation for Odum and Pinkerton's principle (or rather, the maximum entropy production principle, to which it is roughly equivalent), and in Chapter 4 I will discuss possible ecological mechanisms by which the maximum might be attained.

There have been various other proposed "goal functions" in the ecological literature (see Fath, Jœgensen, Patten, and Straškraba (2004) for a review). Examples are the increase over time of biomass and primary production over time as ecosystems develop (E. P. Odum, 1969); a hypothesised increase over time of a quantity called "ascendency" (Ulanowicz, 1980); and Schneider and Kay's (1994) hypothesis that ecosystems should become more complex over time, increasing their rate of dissipation. Virgo et al. (2006) investigate a number of these hypotheses using a population-based model of ecosystem development.

It should be noted that all these hypotheses involve the increase in some quantity over time. The hypothesis proposed by H. T. Odum and Pinkerton (1955) has a slightly different character (as does the maximum entropy production hypothesis in climate science). Rather than simply proposing that the rate of power production will increase over time, Odum and Pinkerton suggest that the values of other parameters (such as the combined rate of all the organisms' metabolism - see Chapter 4 of this thesis) will eventually take on values such that this maximum is attained. In many cases the extremum function has a maximum for some particular value of a parameter. This means that Odum and Pinkerton's hypothesis can, in principle, be used to make specific numerical predictions about the value of these parameters, allowing it to be tested, either experimentally or through field observations, in a rigorous manner. However, to the best of my knowledge, such numerical testing of Odum and Pinkerton's hypothesis has not been performed.

Another hypothesis worth mentioning in this context is Prigogine's principle of minimum entropy production (1955). Although Prigogine knew when he formulated it that this principle could only be applied to a very restrictive range of near-equilibrium systems, many later authors tried to apply it to ecosystems (see, e.g., the review of Fath et al. (2004)), which has caused a certain

amount of confusion in the literature, especially since the idea of a *minimum* in entropy production runs directly counter to almost all the other proposed goal functions (particularly the *maximum* entropy production hypothesis with which this thesis is concerned), and to direct experience. It has since been proven beyond doubt that Prigogine's principle cannot be generalised to complex far-from-equilibrium systems such as ecosystems (Richardson, 1969).

2.5 The Maximum Entropy Production Principle in Climate Science

This section discusses the development of a hypothesised Maximum Entropy Production Principle (MEPP) in the field of climate science. This material is covered in greater depth in Chapter 3. Like Odum and Pinkerton's (1955) maximum power principle in ecology, the MEPP is an extremum principle, suggesting that the rates of processes in the large and complex system of the Earth's atmosphere are such that a particular quantity is sustained at its maximum possible value.

In contrast to the development of Odum and Pinkerton's principle, however, the MEPP arose from an empirical discovery: Paltridge (1979) simply tried inserting a number of arbitrary extremum principles into a simple model of the Earth's climate until he found one that was successful (Paltridge, 2005). Paltridge's impressive results and those of later authors have lent the principle a high degree of credibility within climate science, but the principle still lacks a widely accepted theoretical justification.

The maximum entropy production principle should not be confused with Prigogine's similarly named minimum entropy production principle. Although the principles' names suggest contradicting hypotheses they are actually completely different types of principle, with nonintersecting domains of applicability (to be discussed below). Prigogine's principle applies only to linear, near-equilibrium systems and is not relevant to the complex, far-from-equilibrium systems considered in this thesis.

The MEPP was applied to the atmospheres of Mars and Titan in addition to Earth by R. D. Lorenz et al. (2001). This work uses a simplified version of Paltridge's (1979) model. This simplified model is the basis of much of Chapter 3 and is explained in some detail there, but it is worth examining its basic features here.

In Lorenz et al.'s model a planet's atmosphere is modelled by two "boxes", representing the atmosphere over the planet's tropical and polar regions. (Paltridge's original model had 10 boxes.) Heat from the sun flows into these boxes at different rates, because of the oblique angle the incoming solar radiation makes with the Polar region. Heat flows out of each box at a rate determined by its temperature. (This rate is proportional to the fourth power of temperature according to the Stefan-Boltzmann law but is linearly approximated in Lorenz et al.'s model.) The higher temperature of the equatorial box causes a flow of heat which takes place at a rate Q, assumed to be constant over time. In Earth's atmosphere this heat flow is produced mainly by convection, but the complex details of the transport mechanism are not modelled. The value of Q cannot be determined from these constraints: the model is underdetermined in the sense that it permits a range of possible steady states, each with a different value for Q.

The rate of entropy production σ associated with the heat transport is given by $Q/T_{\text{poles}} -$

Q/T_{tropics}. One extreme case is $Q = 0$, in which case the entropy production is zero. With higher values of Q the temperature of the tropical region becomes lower because of the loss of heat, which also raises the temperature of the poles. Eventually another extreme is reached where the heat transport is so fast that T_{poles} and T_{tropics} become equal, and σ is again zero. Plausible values of Q lie between these extreme values.

Application of the maximum entropy production principle consists of choosing the value of Q that maximises σ. Lorenz et al. show that this gives answers close to the observed heat transport rates of Mars, Earth and Titan.

As Kleidon and Lorenz (2005) point out, the results of Paltridge (1979), E. N. Lorenz (1960) and others correspond to applications of the same principle to more detailed energy balance models, and show that not only can the intensity of the atmospheric heat transport and the temperature difference between the equator and the poles be predicted using the MEPP but so can the meridional distribution of cloud cover. This, combined with R. D. Lorenz et al.'s (2001) successful application of a two-box MEPP model to Mars and Titan as well as to Earth, seems to strongly suggest that MEPP is a genuine physical principle which can be used under some very general conditions to make predictions about various physical parameters of steady-state systems.

However, the specific conditions under which the principle can be applied are currently unknown. In particular, (Essex, 1984) pointed out that σ as defined above excludes the considerable production of entropy due to the absorption of high-frequency photons by the atmosphere. If one tries to apply the MEPP to a system that includes this process as well as the heat transport from tropics to poles, one obtains an answer that does not correspond to the observed value. Every application of the MEPP requires a choice to be made about where to draw the system boundary, and a different choice of boundary will produce different numerical predictions about the behaviour of the system.

It has been clear to most authors that in order to make progress in applying the MEPP to new physical systems, we need a theoretical explanation of the principle which can help us to understand the conditions under which it can be applied. In particular, a successful explanation must tell us how to determine which boundary to use when applying the principle.

Currently, the closest thing we have to a theoretical explanation for the MEPP is the work of Dewar (2003, 2005a, 2005b), which I will examine in detail in Chapter 3. Dewar's theorems provide us with some answers about the conditions under which MEPP can be applied, by making the claim that steady state systems maximise entropy production *subject to constraints*, in an analogous way to the increase of entropy in isolated physical systems. An assumption of maximum entropy production thus corresponds to a "best guess" about a system's behaviour. If this guess is contradicted by empirical measurements of the system's behaviour it indicates that additional constraints are in operation that were not taken into account when calculating the entropy production. (This reasoning may seem somewhat circular at a first glance. The full explanation involves the use of Bayesian probability theory and is analogous to the maximum entropy framework for equilibrium thermodynamics as described in Section 2.7 below.)

However, in my view Dewar's theory does not satisfactorily answer the question of where to

draw the system boundary when applying the MEPP. It appears to be applicable to any physical system, but this means that contradictory results can be obtained from it depending on where the boundary is drawn.

There is therefore still a need for a theoretical explanation of the MEPP. Such a theory must not only explain the effectiveness of MEPP in climate science but must also give us a prescriptive procedure for deciding where to draw the boundary when applying MEPP in new domains. Only when these issues are fully understood will it be possible to say whether and in what way MEPP can be applied in the biological sciences.

I will make some progress towards such a development in Chapter 3, in which I will give an explanation for MEPP that is different from Dewar's, although based on many of the same principles.

2.5.1 Maximum Entropy Production versus Minimum Entropy Production

It is important to be clear about the difference between the MEPP and Prigogine's (Prigogine, 1955) *minimum* entropy production principle. Prigogine's principle applies only to systems close to equilibrium that admit only a single possible steady state, and for which the flows across them can be approximated by a (known) linear function of the gradients that drive them. Prigogine's theorem states that, if one assumes some of the gradients to be held constant, the steady state can be found by minimising the entropy production with respect to the flows and the other gradients.

Thus the minimum entropy production principle is simply equivalent to the assumption that the system is in a steady state and has no physical content beyond this assumption (Jaynes, 1980). It was originally hoped that the principle would admit some generalisation to systems further from thermal equilibrium but it has since been proven that no principle of this form can apply to such systems (Richardson, 1969).

The MEPP, however, is not affected by this non-existence theorem because it is an entirely different kind of principle. Whereas the minimum entropy production principle applies to systems that only admit one steady state and asserts that all neighbouring, non-steady, states must have a greater entropy production, the MEPP applies to systems far from equilibrium that admit many possible steady states and asserts that out of these possible steady states the observed one is likely to be the one with the greatest entropy production. Unlike the minimum entropy production principle, the MEPP embodies a hypothesis that cannot be derived from the steady state assumption and the conservation laws; see (Kleidon & Lorenz, 2005) for a further comparison.

2.6 Some Notes on the Structure of Thermodynamics

Thermodynamics is usually expressed in a way that makes energy, rather than entropy, appear to be the central quantity. The goal of this section is to show that the formalism of thermodynamics can be expressed more symmetrically with entropy as the central quantity. This will become important later, with the discussion of non-equilibrium thermodynamics and statistical mechanics below and in Chapter 3.

The relationships between the central quantities of thermodynamics can be summed up by the so-called *fundamental equation of thermodynamics*,

$$\mathrm{d}U = T\mathrm{d}S - p\mathrm{d}V + \sum_{i=1}^{p} \mu_i \mathrm{d}N_i, \tag{2.3}$$

where U represents the internal energy of a system, V its volume, S its entropy and N_i the molar quantity of particles of the i^{th} type. Each of these quantities is "extensive", meaning that their values scale linearly with the size of the system. The "intensive" quantities (whose value does not change with the size of the system) temperature T, pressure p and the chemical potentials $\{\mu_i\}$ are defined as the change in energy with respect to an extensive quantity, while holding all the other extensive quantities constant. For example,

$$T = \frac{\partial U(S,V,N_1,\ldots,N_p)}{\partial S}, \tag{2.4}$$

where the notation makes explicit that U is to be considered a function of the extensive variables, all of which except for S are to be held constant when performing the differentiation. If necessary, the fundamental equation can be extended to include terms for additional extensive quantities (such as electric charge or angular momentum, for example), which are then given their own conjugate intensive quantities.

There is a slightly odd asymmetry to the fundamental equation as written, however, in that the entropy S has a different character to the other extensive variables (and, as a matter of fact, is not not necessarily extensive in the case of very small systems). The quantities U, V and $\{N_i\}$ are all conserved for an isolated system; they can neither increase nor decrease in a system except by being exchanged with another system. In this respect there is nothing particularly special about energy, and its place on the left-hand side of Equation 2.3 is, arguably, an historical accident. Entropy, on the other hand, always increases up to a maximum. The symmetries inherent in the formalism of thermodynamics can be made clearer by moving S to the left-hand side:

$$\begin{aligned} \mathrm{d}S &= \frac{1}{T}\mathrm{d}U + \frac{p}{T}\mathrm{d}V - \sum_i \frac{\mu_i}{T}\mathrm{d}N_i, \\ &= \lambda_U \mathrm{d}U + \lambda_V \mathrm{d}V + \textstyle\sum_i \lambda_{N_i} \mathrm{d}N_i, \end{aligned} \tag{2.5}$$

where the entropy is now considered a function of all the other extensive variables and the quantities $\lambda_U = 1/T$, $\lambda_V = p/T$ and $\{\lambda_{N_i} = -\mu_i/T\}$ are again defined as partial derivatives, e.g.

$$\lambda_U = \frac{\partial S(U,V,N_1,\ldots,N_p)}{\partial U}. \tag{2.6}$$

Note that, if entropy is measured in bits, the units of λ_U are $\text{bits}\cdot\text{J}^{-1}$, those of λ_V are $\text{bits}\cdot\text{m}^{-3}$, etc. Extending the terminology of Tribus (1961), I will refer to these quantities as the *tempers* of their respective extensive quantities.

We will see in Section 2.7.4 that the fundamental equation in this form arises naturally from statistical mechanics. It makes clear that the role of energy in thermodynamics is similar to the roles of the other conserved quantities and that the role of entropy is different. Another advantage of working with tempers rather than the traditional intensive quantities is that there may be some situations where the temperature cannot be defined but the tempers of quantities other than energy can.

The quantities $1/T$, p/T and $\{-\mu_i/T\}$ also arise in the linear non-equilibrium thermodynamics of Prigogine (e.g. 1955), in which the central quantities are *fluxes* J_i of extensive quantities across a system, and *thermodynamic forces* F_i, which are essentially small differences in temper across a system. The fluxes and forces are assumed to be linear functions of one another, which is a good approximation if the system is close to thermodynamic equilibrium.

I will not use the "thermodynamic force" terminology, for two reasons. Firstly, the approach that I will use to non-equilibrium thermodynamics can be applied outside the linear regime, and in this case the absolute values of the tempers, rather than just the differences between them, are relevant. Secondly, the word "force" implies that the direction of causation is from the force to the flux. To my mind this is not the correct way to picture the situation. Systems are not "driven" by a need to increase entropy, but rather the entropy (and the tempers, which are derived from it) is a tool that we use to make predictions about the system's macroscopic state.

2.6.1 Entropy and the Free Energies

In traditional thermodynamics there are a family of concepts called "free energies," the most commonly used being the Gibbs and Helmholtz free energies. The idea is that when a system is not isolated but instead is held at a constant temperature, or constant temperature and pressure, then its entropy can decrease as well as increase and will not in general tend towards a maximum. This is because in order for its temperature to remain constant the system must exchange energy with another system (usually referred to as its "surroundings"), and it is the combined entropy of the system and its surroundings which cannot decrease. From this one can derive that an appropriately defined free energy of the system cannot increase but must always decrease; most results in chemical thermodynamics are found by minimising a free energy function rather than directly maximising an entropy.

In this section I will briefly derive the free energy quantities, and argue that, like the intensive quantities of thermodynamics, they are better expressed in terms of entropy rather than energy.

Let us suppose we have a system with extensive variables U_1, V, $\{N_i\}$, which is connected to another system, which we will call the heat bath, whose internal energy is U_2. The fundamental equation for the combined system is

$$\mathrm{d}S_{\text{total}} = \lambda_{U_1}\mathrm{d}U_1 + \lambda_{U_2}\mathrm{d}U_2 + \lambda_V \mathrm{d}V + \sum_i \lambda_{N_i}\mathrm{d}N_i. \tag{2.7}$$

If the systems are large then $\lambda_{U_1} = 1/T_1$ behaves as a property of system 1 only, and $1/T_2$ a property of the heat bath only. (This can be derived from statistical mechanics, and arises from

the fact that the motions of the particles in one system are, to a good approximation, uncorrelated with the particles in the other.) We can thus write $S_{\text{total}} = S_1(U_1, V, N_1, \ldots, N_p) + S_2(U_2)$.

We further assume that the heat bath is so large that λ_{U_2} can be considered a constant. To find the equilibrium state of the combined system we must maximise S_{total}, subject to the constraint that the total energy, $U_1 + U_2$ is a constant, or $\mathrm{d}U_2 = -\mathrm{d}U_1$. This gives rise to

$$\mathrm{d}S_{\text{total}} = \mathrm{d}S_1 + \mathrm{d}S_2 = \mathrm{d}S_1 - \lambda_{U_2}\mathrm{d}U_1, \tag{2.8}$$

or $S_{\text{total}} = S_1 - \lambda_{U_2}U_1$, plus a constant of integration. So one can calculate the equilibrium state of the system by maximising the quantity $S^* = S_1 - \lambda_{U_2}U_1$.

In equilibrium $\lambda_{U_1} = \lambda_{U_2}$, i.e. the temperature of the system must equal that of the heat bath. If we assume that this equalisation of temperature occurs much more rapidly than the changes in V and $\{N_i\}$, so that the temperature of system 1 is effectively held constant as its other variables change, then we can write $S^* = S_1 - \lambda_{U_1}U_1$. But now we have eliminated all properties of the heat bath from the equation, and we can conclude that the quantity $S - \lambda_U U$ increases toward a maximum for any system whose temperature is held constant.

In the traditional language of thermodynamics this would be written $S^* = S - U/T$. However, the standard practice in physics is to multiply this quantity through by $-T$ to produce the *Helmholtz free energy* $A = U - TS$, which is to be minimised rather than maximised. Multiplying through by $-T$ is a somewhat arbitrary operation. It is permissible because under the assumptions used, T is a constant, and so maximising S^* is the same as minimising $-TS^*$. It changes the quantity into the units of energy, but this is not necessarily desirable since it obscures the purely entropic nature of the quantity. One could just as easily multiply through by T/p to produce a "free volume". Moreover, maximising S_{total} is valid even when the temperature is not constant.

The above reasoning easily generalises. A very common situation is a chemical system for which both temperature and pressure are held constant, which corresponds to constant λ_U and λ_V. In this case one can find the equilibrium by maximising $S - \lambda_U U - \lambda_V V$, which again is usually multiplied through by $-T$ to obtain the *Gibbs free energy* $G = U + pV - TS$.

2.6.2 Entropy and Work

One of the most central concepts in thermodynamics is that of *work*, or usable energy. Work comes in two distinct varieties: *mechanical work* and *chemical work*. In this section I will point out the differences between these two concepts and argue that the latter is best thought of in terms of entropy rather than energy.

One way to think of mechanical work is as energy whose associated temper is zero. For example, if a heat engine extracts an amount Q_1 of heat from a reservoir of temperature T_1 and deposits an amount Q_2 of heat into a colder reservoir of temperature $T_2 < T_1$ in order to perform an amount W of work, then we know from the second law that $Q_2/T_2 - Q_1/T_1 \geq 0$. (Combined with the first law, $W = Q_1 - Q_2$, this tells us that the efficiency, W/Q_1, of the engine cannot be

more than $1 - Q_2/Q_1$.) This second law inequality is equivalent to

$$\Delta S = \sum_{i=1}^{3} X_i \lambda_i, \tag{2.9}$$

where $X_1 = Q_1$, $\lambda_1 = 1/T_1$, $X_2 = Q_2$, $\lambda_2 = 1/T_2$, $X_3 = W$ and $\lambda_3 = 0$.

Formally, work behaves like heat whose temperature is infinite (zero temper is equivalent to infinite temperature). One intuitive way to see this is to use a result from statistical mechanics that temperature is (roughly) proportional to the average amount of energy in each mechanical degree of freedom of a system. Thus heat can be thought of as energy spread throughout a huge number of degrees of freedom — the velocities and rotations of all the atoms in a substance — whereas work is energy concentrated into a single degree of freedom. Thus the energy per degree of freedom is not infinite, but it is so large compared to the corresponding figure for heat that we might as well treat it as infinite.

It should be noted that it is not possible to reduce the entropy of a system by performing mechanical work on it alone. For instance, if the heat engine in the previous example is run in reverse, so that W, Q_1 and Q_2 become negative then the second-law condition is still $Q_2/T_2 - Q_1/T_1 \geq 0$. This means that, although the entropy of the cold reservoir decreases as a result of work being performed, the entropy of the combined system of both reservoirs must increase overall (or, in the limit of perfect efficiency, remain the same).

There is nothing in the formalism of thermodynamics to prevent amounts of quantities other than energy from occurring with zero (or approximately zero) temper. However, energy is somewhat unique in being able to be concentrated into a single degree of freedom in this way, which is why one does not usually come across work-equivalents with units of volume or particle number rather than energy.

Chemical work is a distinct concept from mechanical work. Like the free energies, but unlike mechanical work, chemical work can be seen as a change in entropy rather than energy. As an example of chemical work, suppose we have a solution of two chemical species, A and B, such that $\mu_A > \mu_B$, or $\lambda_B > \lambda_A$, assumed for the moment to be in an adiabatic (heat-proof) container at constant volume. (The chemical potential μ is a function of concentration; see Section 4.2.) In this case a reaction $A \leftrightarrow B$ will proceed in the forward direction, converting A into B, since the change in entropy per mole of A converted into B is given by $\lambda_B - \lambda_A > 0$.

But now let us suppose that the solution contains another two species, C and D, such that $\lambda_C < \lambda_B$, and let us replace the reaction $A \leftrightarrow B$ with $A + C \leftrightarrow B + D$. If $\lambda_B + \lambda_D > \lambda_A + \lambda_C$ then the reaction will still proceed in the forward direction, with a concomitant increase in entropy. However, in some situations it makes sense to consider the concentrations of A and B as being a separate system from the concentrations of C and D. In this case we say that the reaction $A \leftrightarrow B$ is coupled to the reaction $C \leftrightarrow D$ and that the entropy of the C-D system is being reduced, with this entropy reduction being offset by a greater entropy increase in the A-B system.

If such a system takes place in conditions of constant temperature and/or volume then it would usually be discussed in terms of free energy rather than entropy: one would say that the free

energy of the *C*-*D* system has been increased at the cost of reducing the free energy of the *A*-*B* system at a greater rate. This makes the process appear to involve a change in energy rather than entropy, which is why it is often referred to as "work". However, as we saw above, changes in free energy are really just changes in entropy that have been divided by temperature to give a quantity in energy units, so chemical work can always be thought of as a reduction in entropy of part of a system.

We can now see an important difference between mechanical and chemical work. It is impossible to reduce the entropy of a system by performing mechanical work upon it, although it is possible to use mechanical work to transfer entropy from one part of a system to another. Performing chemical work, on the other hand, always reduces the entropy of a system, by definition. It is important to be aware of the difference in meaning of these two terms.

2.6.3 Non-equilibrium thermodynamics

The approach taken to non-equilibrium thermodynamics in this thesis was described in Chapter 1 and will be developed more thoroughly in Chapter 3, but it is worth making a few important points here.

The concept of entropy is well-defined in equilibrium thermodynamics but does not generalise uniquely to far-from-equilibrium situations. The early approaches of authors such as Onsager and Prigogine (e.g. 1955) involved a near-equilibrium assumption, such that the system could be considered as having a classically defined entropy at each point in space. Such an assumption is almost certainly inapplicable to living systems, and in any case would be little help for modelling ecosystems, since the near-equilibrium methodology requires explicitly modelling the system's behaviour at every point in space. The approach used in this thesis does not require this assumption.

An excellent summary of the type of approach taken here can be found in a review by Jaynes (1980), who, describing the work of Tykodi (1967), asks us to consider an experiment performed by Joule and Thomson in 1842. In this experiment, gas is forced through a porous plug in such a way that it has no contact with the outside world except through the gas in the entry and exit pipes. The gas heats up as it is forced through the plug. The temperature and pressure of the gas are measured in the incoming pipe, and far enough downstream of the plug in the outgoing pipe that the gas has locally come to equilibrium.

Such a setup allows the use of equilibrium thermodynamics to reason about the potentially very complex and far-from-equilibrium processes taking place in the plug. In particular, if the system is in a steady state, the total amount of energy going in must equal the total amount coming out, and the entropy of the exiting gas must be equal or greater than that of the incoming gas. As Jaynes points out, the system need not be a simple plug and could contain any number of processes. Its internal state also need not be in a steady state, as long as the properties of the flows coming in and going out do not change over time.

This setup can be generalised to multiple flows of multiple types of substance, which can be flows of heat or chemical compounds. This is the essence of the approach taken to non-equilibrium

thermodynamics in this thesis. We consider systems where the flows in and out are close to thermodynamic equilibrium and changing little over time, allowing the steady-state entropy production to be calculated, even though the internal state of the system itself may be very far from equilibrium, with an entropy that is impossible to calculate in practice.

2.7 Jaynes' approach to Statistical Mechanics

The argument in Chapter 3 is based upon an approach to statistical mechanics pioneered by Jaynes (1957a, 1957b, 1965, 1979, 1980, 1985, 2003), known as the maximum entropy, or MaxEnt, approach (not to be confused with the maximum entropy *production* principle). The results in later chapters do not specifically involve statistical mechanics, but Jaynes' work was nevertheless an important influence upon them. I will briefly summarise the MaxEnt approach here.

The difference between Jaynes' approach and more traditional approaches to statistical mechanics is subtle but important, and regards the interpretation of the meaning of probability. This may initially seem a trifling philosophical point, but in fact has direct practical implications, as we will see below and in Chapter 3. In particular, it frees us from so-called "ergodic" assumptions which were long thought necessary to derive many important formulae in statistical mechanics, and this in turn opens up the possibility of applying these formulae to non-equilibrium systems where the ergodic assumptions do not apply.

Jaynes saw statistical mechanics as a specialised application of a very general form of Bayesian inference. Accordingly, his procedure was to develop the mathematical part of the theory in general terms, without reference to the physics, and then later show how its application to physical microstates resulted in the standard formulae of statistical mechanics and thermodynamics.

An interesting historical account of the development of this method, as well as some impressive examples of its power, can be found in (Jaynes, 1979). I will loosely follow the more pedagogically oriented approach found in (Jaynes, 2003).

2.7.1 Probability Theory as Logic

Probabilities are often thought of as representing the frequency with which particular events occur in repeated experiments. On this view, the statement that the probability of rolling a 3 on a six-sided die is $1/6$ means that, if the die is rolled a great many times, it will come up with a 3 in one sixth of these trials. This view, which is now known as "frequentism", could be seen as obsoleted by Bayesian approaches, one form of which is described below. Historically, what are now called Bayesian approaches to probability pre-date the development of frequentism as a formal approach to probability theory, much of which took place during the 19^{th} and 20^{th} centuries (Fienberg, 2006; Jaynes, 1979), although one suspects the intuitions underlying it are older. The Bayesian approach began with Bayes' (1763) paper and was developed in more detail by Laplace, among other authors (see Fienberg, 2006). Although frequentism is now declining in popularity it is worth mentioning here because it was the orthodoxy during the development of much of statistical mechanics (see Jaynes, 1979), as well as classical statistical inference and Kolmogorov's

axiomatisation of probability theory. As such, it has left its mark on the terminology used in these subjects, and the way in which they are traditionally taught.

Bayesian views of probability see probability statements as expressing something much more general than the frequency of results in repeated experiments. Under a Bayesian interpretation, probabilities are used to represent "degrees of belief" about statements or hypotheses. On this view the statement that the probability of rolling a 3 is $1/6$ is a statement about the knowledge of an observer: it represents our degree of certainty about the outcome of a single roll of the die[4]. Bayesian probability can thus be applied in any situation in which one has a lack of knowledge. This includes the outcomes of random experiments but also includes many other situations.

Jaynes (2003) begins his exposition of the MaxEnt approach with a discussion of the theorems of Cox (1946, 1961), who showed that, under certain mild assumptions, probability theory arises as the only consistent extension of propositional logic to deal with uncertainty. Cox' theorems represent an alternative to the standard Kolmogorov axioms of probability theory. In Cox' framework, probabilities apply not to "events" as in Kolmogorov's formalisation, but to *statements of propositional logic*. The statement $p(A) = 1$ is to be thought of as meaning the same as A, whereas $p(A) = 0$ means "not A", which can also be written $p(\neg A) = 1$.

If we restrict ourselves temporarily to statements whose probability is either 0 or 1 (i.e. those about whose truth or falsehood we are certain), we can confirm that probability theory is indeed equivalent to propositional logic. For example, $p(AB)$, to be read "the probability of A and B", is 1 if and only if both $p(A) = 1$ and $p(B) = 1$.

The power of Bayesian inference begins to arise when we allow probabilities to take on values between 0 and 1, to represent a degree of uncertainty about whether the statement is true or false. For instance, $p(A) = 0.5$ indicates complete uncertainty about the truth of A, with neither A nor $\neg A$ more likely than the other. The other important component of Bayesian inference is the conditional probability. $p(A|B)$ represents the probability that A is true, assuming that B is true. It can be thought of as the extent to which B implies A.

In Cox's framework all probabilities are conditional upon some "state of knowledge", which can be thought of as a statement that asserts everything a particular observer knows or believes to be true. Thus, where we wrote $p(A) = 1$ above we should strictly have written $p(A|U) = 1$, where U is the state of knowledge of some observer who believes that A is true.

Cox proved that, under very mild assumptions, the unique extension of logic to deal with uncertainty is given by the following two rules, or a system equivalent to them. For any statements A, B and C,

$$\begin{aligned} p(AB|C) &= p(A|C)p(B|AC) \\ &= p(B|C)p(A|BC), \end{aligned} \tag{2.10}$$

[4]By "observer" in this section I mean an idealised reasoner with some particular state of knowledge that can be expressed formally. This discussion should not be taken to imply that real human observers represent incomplete knowledge as probability distributions or formally process information.

which is referred to as the product rule, and

$$p(A|C) = 1 - p(\neg A|C), \tag{2.11}$$

which is called the sum rule. From these two rules all of probability theory can be deduced.

The resulting formalism behaves almost identically to the more widely cited Kolmogorov axiomatisation, but has an entirely different interpretation and therefore a much wider range of application. (The only formal difference is that, under Kolmogorov's axioms, the conditional probability $p(A|B)$ is defined as $p(AB)/p(B)$, and is therefore undefined if $p(B) = 0$, whereas under Cox's axioms conditional probabilities are the primary quantities, and hence $p(A|B)$ can have a definite value even in cases where B is known to be false.)

Equation 2.10 can be rearranged to give Bayes' theorem, which tells us how to update probabilities upon receiving new information. Let U represent the state of knowledge of some observer. This observer is concerned with the probability of some hypothesis X. Initially the observer considers this to be some number $p(X|U)$. But now let us suppose that the observer learns a new fact D, referred to as the data. Now the observer's state of knowledge has changed from U to DU (that is, D and U). Bayes' theorem states that the observer should now consider the probability of X to have a new value,

$$p(X|DU) = p(X|U)\frac{p(D|XU)}{p(D|U)}. \tag{2.12}$$

The usefulness of this is that the quantities on the right-hand side are often easy to calculate.

The quantity $p(X|U)$ is called the observer's *prior* for X, since it represents the observer's degree of belief in X prior to learning the data. $p(X|DU)$ is called the *posterior*. If another piece of data comes to light, the posterior will become the prior for a further application of Bayes' theorem.

Bayes' theorem provides a powerful method for updating probabilities from priors to posteriors, but it does not in itself give a method for determining what to use as the prior in a given situation. However, there are techniques for determining priors in some situations. One of these is the maximum entropy (MaxEnt) principle.

2.7.2 Maximum Entropy Inference

Bayesians can be broadly divided into two schools: subjective Bayesians, who believe that there are no situations in which there is an objectively correct way to determine priors, and objective Bayesians, who believe that such situations do exist. Jaynes was a key member of the objective school. "Objective" is something of an unfortunate term, however, since it could be taken to mean that probabilities have an objective meaning independent of any observer, which is not what is meant by it. (In fact, Jaynes considered the idea that probabilities have objective, observer-independent values to be a symptom of what he called the "mind projection fallacy", a logical fallacy that involves assuming one's knowledge of the world is a property of the world alone, rather than of one's relationship to it.)

One example of a situation where there seems to be one particular correct prior is as fol-

lows: we are given the information that a ball has been placed under one of three cups, and no other information. In this case it seems intuitively clear that we should assign the probabilities $p(\text{ball under cup 1}) = p(\text{ball under cup 2}) = p(\text{ball under cup 3}) = 1/3$. This can be seen more formally by noting that only with this assignment of probabilities can we swap around the labels "cup 1", "cup 2" and "cup 3" while keeping our state of knowledge the same. Since the information we were given is symmetric with respect to the three cups — it gives us no reason to think that any of them are in any way different from the other two — our probability distribution should have the same property. This argument can be found in (Jaynes, 2003).

This prior is objective in the sense that it uniquely expresses a particular state of knowledge about the ball's location. A different state of knowledge would result in a different probability distribution, and so in this sense it is still a subjective property of the observer. The ball itself is in fact under a particular cup, and so has no intrinsic probability distribution.

The maximum entropy principle generalises this example in a powerful way. It can be seen as a method for converting specific information into a probability distribution. It is used when the information we have takes the form of *constraints* on a probability distribution. An important type of constraint takes the form of a known value for an expectation. For example, suppose our knowledge consists of the following two statements: x is an integer between 1 and 4 inclusive, and the expectation of x is 3. We must therefore assign probabilities p_1, p_2, p_3, p_4 such that $\sum_{i=1}^{4} i p_i = 3$. There are many possible probability distributions with this property, and MaxEnt is a principle for choosing between them.

The basic idea behind the maximum entropy principle is that we do not want the probability distribution to encode any additional assumptions other than the ones we specifically include. This is done by choosing the probability distribution with the greatest value for the *information entropy*

$$H = -\sum_i p_i \log_2 p_i, \tag{2.13}$$

subject to the constraints we wish to include. The justification for this comes from Shannon's (1948) proof that the entropy is the only measure of the amount of uncertainty (or spread-outness) of probability distributions that satisfies certain desirable properties such as consistency. The entropy can be thought of as an amount of information, measured in bits (changing the base of the logarithm corresponds to choosing a different "unit of measurement" for the entropy). The greater the entropy, the less information can be said to be encoded in the probability distribution. We want to encode only the information that corresponds to the specific knowledge we have, and this is achieved by maximising the entropy. To choose a distribution with n bits less entropy than the maximum corresponds to making n bits' worth of additional unnecessary assumptions.

The information entropy should not be confused with the physical, thermodynamic entropy. The two concepts are intimately related (in fact, as Jaynes (e.g. 1979) points out, Equation 2.13 was used by Gibbs in the context of physics long before its generality was demonstrated by Shannon's proof), but they are not always numerically equal. The difference between the two concepts will be explained below. For now the information entropy should be seen as a conceptual tool for

use in inference with no particular physical interpretation.

In general, let us suppose we wish to determine a probability distribution for some quantity x, which can take on integer values from 1 to n (which might be infinite). Let us further assume that our knowledge consists of the expectation values of m functions of x, labelled $f_1(x), \ldots, f_m(x)$. The MaxEnt procedure consists of maximising H subject to the constraints that

$$\sum_{i=1}^{n} p_i f_j(i) = F_j, \qquad (\text{for } j = 1 \ldots m) \tag{2.14}$$

where $F_1 \ldots F_m$ are the specified expectation values. We must also take account of the additional constraint that the probability distribution be normalised, i.e.

$$\sum_{i=1}^{n} p_i = 1. \tag{2.15}$$

This can be solved using the method of Lagrange multipliers (Jaynes, 1957a, 2003), and gives probability distributions of the form

$$p_i = \frac{1}{Z} \exp_2 \Big(- \sum_{j=1}^{m} \lambda_j f_j(i) \Big), \tag{2.16}$$

where

$$Z = \sum_{i=1}^{n} \exp_2 \Big(- \sum_{j=1}^{m} \lambda_j f_j(i) \Big) \tag{2.17}$$

is a normalising factor (called the "partition function") and $\lambda_1, \ldots, \lambda_m$ are parameters whose values can in principle be calculated. (I have used base-2 exponential functions, in keeping with specifying entropy in bits. These expressions more commonly use base-e exponentials.)

For instance, in the example above this tells us that $p_i = \frac{1}{Z} 2^{\lambda i}$. The value of λ that fits the constraint that $\sum_{i=1}^{4} i p_i = 3$ is approximately 0.605, with a corresponding $Z = 12.7$, leading to probabilities $p_1 = 0.12$, $p_2 = 0.18$, $p_3 = 0.28$ and $p_4 = 0.42$.

This is the only probability distribution that encodes only the mean-value constraint(s) and no other information. We can now use this probability distribution to make predictions about other quantities (usually other functions of x). These predictions are, in a formal sense, our best guesses given the information we have.

2.7.3 The Connection Between MaxEnt Distributions and the Frequency of Random Events

The maximum entropy method was presented above as a technique for assigning a probability distribution for the value of an unknown quantity. In such a case the quantity has no intrinsic probability distribution, since it has only one actual value. There is no objective "correct" distribution that is approached or approximated by the maximum entropy method. The distributions produced by maximising the entropy are correct only in that they correspond to the limited knowledge we have as observers about the true value.

However, a common situation in science is that x represents an outcome of a repeatable experiment. This need not be an experiment in physics — it could just as easily be an experiment in biology or psychology (the specific connection to physics will be given below). In such repeatable experiments the value of x might not be the same on each iteration, due issues of measurement, or to an inability to precisely control all the details of the experimental conditions. In such a case the distribution produced by the entropy maximisation principle can be compared to the observed frequencies with which the possible values of x occur.

In general there is no reason to suppose that the two would match[5]. However, if they do then it has an important implication: it means that the information constraints taken into account were sufficient to predict the experiment's outcome. No extra information is necessary, since the given information could already be used to predict the results up to the limit of experimental accuracy. Conversely, any additional non-redundant constraints would change the MaxEnt distribution so that it would no longer match the observed frequencies of experimental results.

Thus a match between a maximum entropy distribution and an observed distribution of experimental results tells us that we have identified all the factors that need to be taken into account in order to predict the experiment's results. Any additional details we learn or measure must be redundant as far as predicting the distribution of x is concerned.

This suggests an important methodological point: when predicting the results of an experiment, one should always try the probability distribution with the maximum entropy subject to constraints formed by one's knowledge, but this is not guaranteed to reproduce the results correctly. If it does not, this indicates that the given information is not sufficient to predict the result, suggesting that further effects or measurements need to be taken into account. Jaynes' book (Jaynes, 2003) argues for a general scientific methodology based upon these principles.

Because of this relation between maximum entropy distributions and experimental results, MaxEnt type distributions often arise in experimental results. For example, the normal distribution is the continuous distribution with the greatest entropy subject to the constraints of a given mean and variance.

2.7.4 Thermodynamics as Inference

Now that the framework of maximum entropy inference has been expressed it can finally be applied to physical microstates, from which the equations of thermodynamics will emerge as a result.

Let us suppose that the possible microstates of a physical system can be enumerated[6], and let x stand for the microstate that a system is in at some particular time. As with the ball-and-cup example above, x has only one correct value at a given time, but our knowledge of it is usually incomplete, being based only on macroscopic measurements.

The state x is thought of as having exact values for a number of quantities. These are usually

[5] Sampling theory must be invoked in order to determine the extent to which the two distributions match; the details of this are beyond the scope of this introductory explanation. Here we will assume that the experiment has been repeated enough times that any mismatch would be obvious.

[6] There are additional complexities involved in considering systems with a continuum of possible microstates, but these do not substantially change the results.

conserved quantities, although the formalism permits the use of quantities that are not conserved. The standard quantities to be considered include energy u, volume v, and molar quantities of particles n_A, n_B, etc. (I will use the convention of lower case letters to denote the precise values of these quantities for each microstate, and capital letters to represent their macroscopic values, to be introduced below). However, our measuring apparatus does not allow us to measure them precisely. We assume for the moment that we can calculate these vales from the value of x (in practice this is almost never the case, and this assumption will be relaxed later). We thus wish to assign a probability distribution over the possible microstates of the system, given the limited knowledge obtained from these imprecise measurements.

Jaynes argued that this can be done by maximising the information entropy, subject to constraints that the expected values of the conserved quantities should be equal to the measured values. That is,

$$\begin{aligned} \sum_i p_i u(i) &= U, \\ \sum_i p_i v(i) &= V, \\ \sum_i p_i n_A(i) &= N_A, \quad \text{etc.} \end{aligned} \tag{2.18}$$

The justification for using expected values in this way is somewhat subtle and will not be discussed in full here, but it hinges on the fact that the resulting probability distributions for physical systems tend to be so highly peaked that the variances of the conserved quantities' values are much smaller than the measurement error.

This results in a probability distribution of the form

$$p(x=i) = \frac{1}{Z} \exp_2\left(-\lambda_U u(i) - \lambda_V v(i) - \lambda_{N_A} n_A(i) - \cdots\right). \tag{2.19}$$

This is simply Distribution 2.16 with $f_1(x) = u(x)$, $f_2(x) = v(x)$, etc. As before, the value of Z is chosen such that the distribution is normalised. Distributions of this form arise via a different train of reasoning in the traditional approach to statistical mechanics, and are known as Boltzmann distributions.

As in Equation 2.16, the quantities λ_U, λ_V, etc. arise as Lagrange multipliers and their values can in principle be calculated if the values of u, v, etc. for each microstate are known. However, they are also equal to the tempers as defined in Equation 2.5, as we will see below.

The thermodynamic entropy S can be defined as the information entropy of Distribution 2.19. We can now see the way in which the thermodynamic entropy and the information entropy differ. The thermodynamic entropy is equal to the information entropy of a probability distribution over a system's microstates (up to a conversion factor of $k_\mathrm{B} \ln 2$ if the thermodynamic entropy is specified in JK^{-1} rather than bits), but *only* if the information entropy of the distribution has been maximised subject to constraints formed by macroscopic measurements. The information entropy is a general concept that applies to any probability distribution, whereas the thermodynamic entropy

is the information entropy of a very specific probability distribution.

For a particular physical system (i.e. a particular set of functions $u(x)$, $v(x)$, etc.), we can maximise the entropy subject to a range of different expectation values for the conserved quantities. Each set of values will give a different value for the maximised entropy. Thus S can be considered a function of the expectation values of the constraints, i.e. $S = S(U,V,N_A,\dots)$. Substituting the distribution of Equation 2.19 into Equation 2.13, we obtain

$$S(U,V,N_A,\dots) = \log_2 Z + \lambda_U U + \lambda_V V + \lambda_{N_A} N_A + \cdots . \tag{2.20}$$

With a little calculation it can be shown from this that

$$\frac{\partial S(U,V,N_A,\dots)}{\partial U} = \lambda_U, \quad \frac{\partial S(U,V,N_A,\dots)}{\partial V} = \lambda_V, \quad \text{etc.}, \tag{2.21}$$

from which we can recover Equation 2.5, the fundamental equation of thermodynamics in entropy-centric form.

A number of other important properties of the function $S(U,V,N_A,\dots)$ can be derived in this manner, but for the most part they will not be needed in this thesis, so I will not repeat their derivation here. Perhaps the most important part is that S is a convex function of its arguments, from which can be derived, among other things, the uniqueness of the thermodynamic equilibrium.

Above I derived the function $S(U,V,N_A,\dots)$ as if the functions $u(x)$, $v(x)$, etc. were known and the tempers and Z could be derived from them. However, we can now see that the function S must have certain properties regardless of what form the functions u, v etc. take. In practice, rather than deriving the values of the tempers λ_U, λ_V, etc. from theoretical principles we usually measure their values using instruments such as thermometers. We can then calculate the function $S(U,V,N_A,\dots)$ from these observations, and work backwards to make inferences about the properties of the individual microstates. The practical use of thermodynamic reasoning usually corresponds to this procedure.

The advantages of Jaynes' approach is that it makes clear the subjective nature of the entropy, and shows that thermodynamics is a special case of the general methodology summarised in Section 2.7.3. This frees us from having to prove that the distribution in Equation 2.19 will actually be observed when a particular system is sampled over time. Such "ergodic" proofs tend to be long and complicated. With Jaynes' procedure we can simply assume that a distribution of this form does apply, because our maximum entropy distribution does not have to match any actual, objective distribution. If we discover that this does not match a repeatable experimental result this simply implies that the constraints encoded in the distribution were not sufficient to predict the result, and new information needs to be taken into account. Jaynes (1992) made a strong argument that this is how thermodynamics has always been applied in practice, and his framework simply makes this explicit.

Jaynes (1980, 1985) argued that this freedom from ergodic assumptions also frees us from the near-equilibrium assumptions that were previously required for statistical mechanics, since our probability distribution no longer has to correspond to a sampling of states over time. This idea

will be used in Chapter 3 of this thesis.

2.7.5 Why Does Thermodynamic Entropy Increase?

One important topic of thermodynamics remains to be discussed in the MaxEnt context: why is it that the thermodynamic entropy S must always increase in large, macroscopic systems? The answer to this question is an important step towards understanding the role of the entropy production in non-equilibrium thermodynamics. In this section I will briefly discuss Jaynes' answer to this question, first given in (Jaynes, 1965), which differs from the more widely invoked "coarse graining" explanation.

So far we have not discussed the change of systems over time. In order to derive the second law of thermodynamics we need one important empirical fact about the microscopic dynamics of physical systems, known as Liouville's theorem. This is a theorem of Hamiltonian mechanics, and is thought to apply to all physical systems, as universally as the conservation of energy. It states that, for an isolated physical system, phase space volume is conserved over time.

The term "phase space" refers to an abstract space whose dimensions are the position and momenta of each particle in the system, or generalisations thereof. A point in phase space corresponds to a complete specification of a system's microstate. Hamiltonian systems' microscopic dynamics consist of following deterministic paths in phase space. Liouville's theorem states that, if we start with a continuous n-volume of points in the n-dimensional phase space of a particular system (corresponding to a continuous range of different initial conditions for the same system) and follow each of them forward for the same amount of time, the resulting set of points in phase space will have the same volume as the original.

More generally, if we start with a probability distribution of points in phase space, the information entropy of this distribution will remain constant over time. Perhaps somewhat counter-intuitively, it is this conservation of information entropy in Hamiltonian dynamics from which the law of increasing thermodynamic entropy can be derived.

To see this, imagine we are performing a repeatable experiment on some isolated physical system. The first step in this experiment is to set up the initial conditions. As an example, we might prepare a hot and a cold body and place them next to each other, but we could equally do something much more complicated. We then make appropriate macroscopic measurements to determine the thermodynamic state (in this example we would measure the two bodies' temperatures) and allow the system to evolve for a period of time in isolation from the rest of the world. (The isolation is required because otherwise we would have to consider the phase space a larger system that includes anything the system of interest interacts with.) Finally, we make the same set of measurements again, to see how the macroscopic state has changed.

Applying Jaynes' procedure for both the initial and final states, we obtain two different probability distributions over the system's microstates, which in general will have different entropies. As the system evolves over time its macroscopic parameters change, but they cannot repeatably change in such a way that the resulting final entropy is smaller than the initial entropy, because of the conservation of information entropy over time. Effectively, the image of the initial distribution

of states cannot "fit into" a distribution with a smaller entropy.

The final entropy can be greater than the initial entropy, however. This corresponds to a loss of information, which is caused by the re-maximisation of the information entropy subject to the final values of the macroscopic variables. This entropy re-maximisation corresponds to the throwing away of information that is no longer relevant to predicting the system's future behaviour. In the example of two bodies exchanging heat, the system's final state is such that, if one were able to precisely measure it and compute the trajectory of the microscopic dynamics backwards in time, one could in theory determine that the thermal energy was initially more concentrated in the hot body. However, such measurements are generally not possible. In practice the only parameters relevant for predicting such a system's future behaviour are the present temperatures of the two bodies. This justifies re-maximising the entropy subject to constraints formed by the new measurements, effectively throwing away the no-longer-relevant information about the system's previous state.

This argument applies for experiments on large systems, which are always repeatable to within the bounds of experimental error. For small systems it is possible for the entropy to decrease sometimes. In these systems it is the expectation of the final entropy that must be greater than the initial entropy; it cannot decrease on average in the long run.

Jaynes' argument is interesting in relation to philosophical questions about the arrow of time (the increase of entropy in one time direction but not the other). Rather than trying to derive the time-asymmetric second law from the time-symmetric underlying laws of microscopic physics (as has historically been attempted more than once; see Price, 1996), it effectively locates the cause of the arrow of time in our ability as experimenters to directly affect the initial but not the final conditions of an experiment. From a philosophical point of view this in itself is still in need of explanation, but from a practical point of view Jaynes' argument makes the reason for the experimentally observed non-decreasing nature of thermodynamic entropy clear.

2.8 Conclusion

This background chapter has covered the application of thermodynamics to living organisms, ecosystems, and general non-equilibrium systems such as the Earth's atmosphere. I have briefly surveyed the use of extremum functions in both ecology and climate science.

I have also commented on the structure of thermodynamics, showing that it can be expressed with entropy rather than energy as the central concept. I have also given a relatively detailed overview of Jaynes' approach to statistical mechanics. This is partly because it is important background material for the following chapter, but also because it represents an especially powerful way to reason about thermodynamics in general.

Chapter 3

The Maximum Entropy Production Principle: Statistical Considerations

---

3.1 Introduction

Evidence from climate science suggests that, in at least some cases, the dynamics of planetary atmospheres obey a principle of maximum entropy production (Kleidon & Lorenz, 2005; Paltridge, 1979; R. D. Lorenz et al., 2001). The rates of heat flow in these systems are such that either increasing or decreasing them by a significant amount would reduce the rate at which entropy is produced within the system. If this principle could be extended to some more general class of physical systems then it would have a wide range of important implications.

However, the principle as currently formulated suffers from an important conceptual problem, which I term the system boundary problem: it makes different predictions depending on where one draws the boundaries of the system. In this chapter I present a new argument based on Edwin Jaynes' approach to statistical mechanics, which I believe can give a constructive answer to this problem.

In the context of this thesis, this chapter serves two purposes, in addition to presenting a possible proof of the maximum entropy production hypothesis. Firstly, it introduces Jaynes' approach to statistical mechanics, which can be used arbitrarily far from equilibrium and is, in my opinion, the correct framework for applying thermodynamics to systems of the level of complexity of a living ecosystem. Secondly, it introduces the notion of negative feedback boundary conditions, whose implications for the dynamics of ecosystems and pre-biotic physico-chemical systems will be examined in the following two chapters.

This chapter is based on (Virgo, 2010), and has been changed very little since the previously published version.

If the maximum entropy production principle (MEPP) turns out to be a general property of physical systems under some yet-to-be-determined set of conditions then it could potentially have a huge impact on the way science can deal with far-from-equilibrium systems. This is because

planetary atmospheres are extremely complex systems, containing a great number of processes that interact in many ways, but the MEPP allows predictions to be made about them using little information other than knowledge of the boundary conditions. If this trick can be applied to other complex far-from-equilibrium physical systems then we would have a general way to make predictions about complex systems' overall behaviour. The possibility of applying this principle to ecosystems is particularly exciting. Currently there is no widely accepted theoretical justification for the MEPP, although some significant work has been carried out by Dewar (Dewar, 2003, 2005a) and other authors (Martyushev & Seleznev, 2006; Attard, 2006; Niven, 2009; Zupanovic, Botric, & Juretic, 2006).

However, maximising a system's entropy production with respect to some parameter will in general give different results depending on which processes are included in the system. In order for the MEPP to make good predictions about the Earth's atmosphere one must include the entropy produced by heat transport in the atmosphere but not the entropy produced by the absorption of incoming solar radiation (Essex, 1984). Any successful theoretical derivation of the MEPP must therefore provide us with a procedure that tells us which system we should apply it to. Without this the principle itself is insufficiently well defined. In Section 3.2 I present an overview of the principle, emphasising the general nature of this problem.

Like the work of Dewar (Dewar, 2003, 2005a), the approach I wish to present is based on Jaynes' (1957a, 1957b, 1965, 1979, 1980, 1985, 2003) information-theory based approach to statistical mechanics, which is summarised in Section 3.3. Briefly, Jaynes' argument is that if we want to make our best predictions about a system we should maximise the entropy of a probability distribution over its possible states subject to knowledge constraints, because that gives a probability distribution that contains no "information" (i.e. assumptions) other than what we actually know. Jaynes was able to re-derive a great number of known results from equilibrium and non-equilibrium thermodynamics from this single procedure.

In section 3.3.1 I argue that if we want to make predictions about a system at time t_1 (in the future), but we only have knowledge of the system at time t_0, plus some incomplete knowledge about its kinetics, Jaynes' procedure says we should maximise the entropy of the system's state at time t_1. This corresponds to maximising the entropy produced between t_0 and t_1. No steady state assumption is needed in order to make this argument.

This simple argument does not in itself solve the system boundary problem, because it only applies to isolated systems. The Solar system as a whole can be considered an isolated system, but a planetary atmosphere cannot. However, in Section 3.4 I present an argument that there is an additional constraint that must be taken account of when dealing with systems like planetary atmospheres, and that when this is done we obtain the MEPP as used in climate science. If this argument is correct then the reason we must exclude the absorption of radiation when using MEPP to make predictions about planetary atmospheres is not because there is anything special about the absorption of radiation, but because the dynamics of the atmosphere would be the same if the heat were supplied by some other mechanism.

The maximum entropy production principle, if valid, will greatly affect the way we concep-

tualise the relationship between thermodynamics and kinetics. There is also a pressing need for experimental studies of the MEPP phenomenon. I will comment on these issues in Section 3.5

3.2 Background: The Maximum Entropy Production Principle and Some Open Problems

The Maximum Entropy Production Principle (MEPP), at least as used in climate science, was first hypothesised by Paltridge (1979), who was looking for an extremum principle that could be used to make predictions about the rates of heat flow in the Earth's atmosphere in a low resolution (10 box) model. Paltridge tried several different such principles (Paltridge, 2005) before discovering that adjusting the flows so as to maximise the rate of production of entropy within the atmosphere gave impressively accurate reproductions of the actual data. The MEPP was therefore an empirical discovery. It has retained that status ever since, lacking a well-accepted theoretical justification, although some significant progress has been made by Dewar (Dewar, 2003, 2005a), whose work we will discuss shortly.

The lack of a theoretical justification makes it unclear under which circumstances the principle can be applied in atmospheric science, and to what extent it can be generalised to systems other than planetary atmospheres. The aim of this chapter is to provide some steps towards a theoretical basis for the MEPP which may be able to answer these questions.

3.2.1 An Example: Two-Box Atmospheric Models

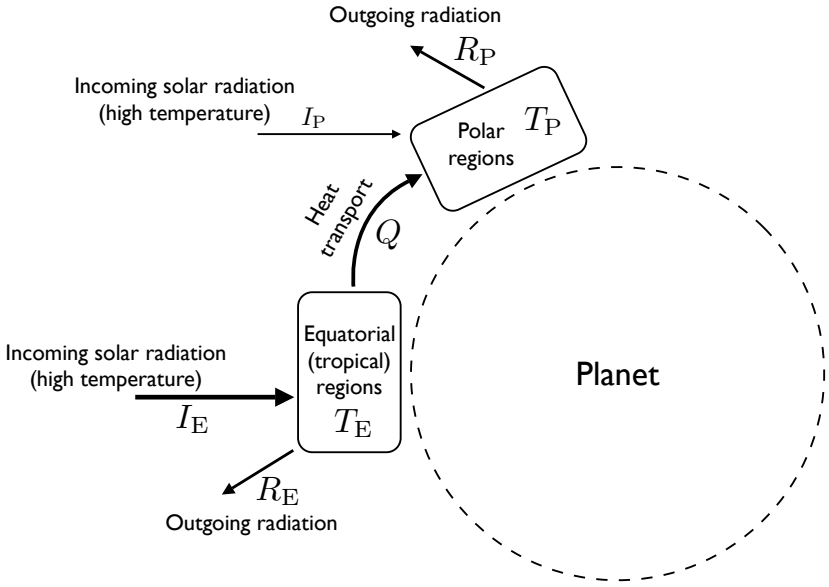
A greatly simplified version of Paltridge's model has been applied to the atmospheres of Mars and Titan in addition to Earth by R. D. Lorenz et al. (2001). The authors state that the method probably also applies to Venus. It is worth explaining a version of this simple two-box version of the MEPP model, because it makes clear the form of the principle and its possible generalisations. Readers already familiar with such models may safely skip to Section 3.2.2

The model I will describe differs from that given in (R. D. Lorenz et al., 2001), but only in relatively minor details. Lorenz et al. use a linear approximation to determine the rate of heat loss due to radiation, whereas I will use the non-linear function given by the Stefan-Boltzmann law. This is also done in Dewar's (2003) version of the model.

In these models, the atmosphere is divided into two boxes of roughly equal surface area, one representing the equatorial and tropical regions, the other representing the temperate and polar regions (see Figure 3.1). Energy can be thought of as flowing between the boxes, mostly by convection, and each box can be very roughly approximated as having a single temperature.

Energy flows into the equatorial region at a rate I_{E}. This represents the flux of incoming solar radiation (insolation). This radiation is effectively at the same temperature as the sun, which is about 5700 K. The atmospheric system is assumed in this model to be in a steady state, so this incoming energy must be balanced by outgoing flows. One outgoing flow is caused by the equatorial region radiating heat to space. The rate at which this occurs depends on the temperature of the equatorial region and is approximately determined by the Stefan-Boltzmann law, which says that the outgoing radiative flux R_{E} is proportional to T_{E}^4. The other outflow from the equatorial

Figure 3.1: A diagram showing the components of the two-box atmospheric heat transport model. Labels written next to arrows indicate rates of flow of energy, whereas labels in boxes represent temperature.



region is the heat flow Q transported by atmospheric processes to the polar region. We will treat Q as a parameter of the model. We will eventually use the MEPP to estimate the value of Q.

Putting this together, we have the energy balance equation for the equatorial region $I_{\mathrm{E}} = Q + kT_{\mathrm{E}}^4$, where k is a parameter whose value depends on the surface area of the equatorial region and on the Stefan-Boltzmann constant. For a given value of Q, this can be solved to find T_{E}. Larger values of Q result in a lower temperature for the equatorial region.

Similarly, for the polar region we have $Q + I_{\mathrm{P}} = kT_{\mathrm{P}}^4$. I_{P} is smaller than I_{E} because the polar regions are at an angle to the incoming radiation, which is why the heat flow Q from the equator to the poles is positive.

We now wish to calculate the rate of entropy production due to the atmospheric heat transport. This is given by the rate at which the heat flow Q increases the entropy of the polar region, minus the rate at which Q decreases the entropy of the tropical region. Entropy is heat divided by temperature, so the rate of entropy production due to Q is given by

$$\sigma_Q = Q/T_{\mathrm{P}} - Q/T_{\mathrm{E}}, \tag{3.1}$$

or $Q(1/T_{\mathrm{P}} - 1/T_{\mathrm{E}})$. The quantity $1/T_{\mathrm{P}} - 1/T_{\mathrm{E}}$ will be referred to as the *gradient*.

Solving the steady state energy balance equations for T_{P} and T_{E} we obtain the entropy production as a function of Q:

$$\sigma_Q = Q\left(\left(\frac{I_{\mathrm{P}} + Q}{k}\right)^{-\frac{1}{4}} - \left(\frac{I_{\mathrm{E}} - Q}{k}\right)^{-\frac{1}{4}}\right). \tag{3.2}$$

The exact form of this equation is not important, but it has some important properties. $\sigma_Q = 0$ when $Q = 0$. At this point the gradient is large, but the heat flow is zero. The inverse temperature difference decreases with increasing heat flow, so that σ_Q is also zero when $Q = \frac{1}{2}(I_E - I_P)$. At this value of Q the temperatures of the equator and the poles are equal, and so there is no entropy production because there is no change in temperature of the energy being transferred. A value of Q greater than this would imply that the temperature of the poles must be greater than that of the equator, and the entropy production would become negative. An external source of energy would thus be required to maintain such a flow, and so $\frac{1}{2}(I_E - I_P)$ can be thought of as the maximum value for Q. The observed value for Q in real atmospheres is somewhere between these two extremes.

The value of σ_Q reaches a peak at some value of Q between 0 and $\frac{1}{2}(I_E - I_P)$. (Numerically, it looks to be about $\frac{1}{4}(I_E - I_P)$.) The work of Lorenz et al. essentially consists finding the value of Q which produces this maximum for models parameterised to match the physical situations of the atmospheres of Earth, Mars and Titan. This produces good predictions of the rate of atmospheric heat transport for these three bodies, which provides good empirical support for the MEPP.

3.2.2 Generalisation: Negative Feedback Boundary Conditions

The two-box atmospheric model described above has an interesting and important feature: the presence of what I will call *negative feedback boundary conditions*. By this I mean that if one thinks of the system as including the convective transport processes in the atmosphere, but not the emission and absorption of radiation, then the dependence of the temperature gradient on the flow rate can be thought of as a constraint imposed on the system from outside.

Usually in non-equilibrium thermodynamics we would consider one of two different types of boundary condition: we could impose a constant gradient upon the system (corresponding to holding T_E and T_P constant), and ask what would be the resulting rate of flow; or else we could hold the flow rate constant and ask the value of the gradient. The boundary conditions on the atmospheric heat transport in the two box model are a generalisation of this, whereby neither the temperature gradient nor the heat flow rate are held constant, but instead the gradient is constrained to be a particular function of the flow. The situation in Paltridge's original 10 box model is more complicated, with more variables and more constraints, but in either case the result is an underspecified system in which many different steady state rates of flow are possible, each with a different associated value for the temperature gradient. As emphasised by Dewar (2003, 2005a), the MEPP is properly seen as a principle that we can use to select from among many possible steady states in systems that are underspecified in this way.

I use the term "negative feedback boundary conditions" to emphasise that, in situations analogous to these two-box atmospheric models, the gradient is a decreasing function of the flow, which results in a unique maximum value for the entropy production, which occurs when the flow is somewhere between its minimum and maximum possible values.

The notion of gradient can readily be extended to systems where the flow is of something other than heat. In electronics, the flow is the current I and the gradient is V/T, the potential difference divided by temperature (the need to divide by temperature can be seen intuitively by noting that

IV/T has the units of entropy per unit time). For a chemical process that converts a set of reactants X to products Y, the flow is the reaction rate and the gradient is $(\mu_X - \mu_Y)/T$, where μ_X and μ_Y are the chemical potentials of X and Y.

If the effectiveness of the two-box model is the result of a genuine physical principle, one might hope that it could be extended in general to some large class of systems under negative feedback constraints. This might include the large class of ecological situations noted by H. T. Odum and Pinkerton (1955), which would provide a much-needed tool for making predictions about complex ecological systems. Odum and Pinkerton's hypothesis was that processes in ecosystems occur at a rate such that the rate at which work can be extracted from the system's environment is maximised. If one assumes that all this work eventually degrades into heat, as it must in the steady state, and that the temperature of the environment is independent of the rate in question, then this hypothesis is equivalent to the MEPP.

The MEPP cannot apply universally to all systems under negative feedback boundary conditions. As a counter-example, we can imagine placing a resistor in an electrical circuit with the property that the voltage it applies across the resistor is some specific decreasing function of the current that flows through it. In this case we would not expect to find that the current was such as to maximise the entropy production $\sigma = IV/T$. Instead it would depend on the resistor's resistance R. The constraint that $V = IR$, plus the imposed relationship between V and I, completely determines the result, with no room for the additional MEPP principle.

For this reason, the MEPP is usually thought of as applying only to systems that are in some sense complex, having many "degrees of freedom" (R. D. Lorenz et al., 2001), which allow the system to in some way choose between many possible values for the steady state flow. A successful theoretical derivation of the MEPP must make this idea precise, so that one can determine whether a given system is of the kind that allows the MEPP to be applied.

In Dewar's approach, as well as in the approach presented in the present thesis, this issue is resolved by considering the MEPP to be a predictive principle, akin to the second law in equilibrium thermodynamics: the entropy production is always maximised, but subject to constraints. In equilibrium thermodynamics the entropy is maximised subject to constraints that are usually formed by conservation laws. In Dewar's theory, the entropy *production* is maximised subject to constraints formed instead by the system's kinetics. If we make a prediction using the MEPP that turns out to be wrong, then this simply tells us that the constraints we took into account when doing the calculation were not the only ones acting on the system. Often, in systems like the resistor, the kinetic constraints are so severe that they completely determine the behaviour of the system, so that there is no need to apply the MEPP at all.

An interesting consequence of this point of view is that, in the degenerate case of boundary conditions where the gradient is fixed, the MEPP becomes a maximum flow principle. With these boundary conditions the entropy production is proportional to the rate of flow, and so according to the MEPP, the flow will be maximised, subject to whatever macroscopic constraints act upon it. Likewise, for a constant flow, the MEPP gives a prediction of maximal gradient.

However, there is another serious problem which must be solved before a proper theory of

maximum entropy production can be formulated, which is that the principle makes non-trivially different predictions depending on where the system boundary is drawn. In the next two sections I will explain this problem, and argue that Dewar's approach does not solve it.

3.2.3 The System Boundary Problem

A substantial criticism of MEPP models in climate science is that of Essex (Essex, 1984), who pointed out that the entropy production calculated in such models does not include entropy produced due to the absorption of the incoming solar radiation. This is important because maximising the total entropy production, including that produced by absorption of high-frequency photons, gives a different value for the rate of atmospheric heat transport, one that does not agree with empirical measurements. This will be shown below.

This is a serious problem for the MEPP in general. It seems that we must choose the correct system in order to apply the principle, and if we choose wrongly then we will get an incorrect answer. However, there is at present no theory which can tell us which is the correct system.

We can see this problem in action by extending the two-box model to include entropy production due to the absorption of radiation. The incoming solar radiation has an effective temperature of $T_{\text{sun}} \approx 5700$ K. When this energy is absorbed by the equatorial part of the atmosphere its temperature drops to T_{E} (or T_{P} if it is absorbed at the poles), of the order 300 K. The entropy production associated with this drop in temperature is given by $\sigma_{\text{absorption}} = I_{\text{E}}/T_{\text{E}} + I_{\text{P}}/T_{\text{P}} - (I_{\text{E}} + I_{\text{P}})/T_{\text{sun}}$, making a total rate of entropy production $\sigma_{\text{total}} = \sigma_{\text{absorption}} + \sigma_Q = (I_{\text{E}} - Q)/T_{\text{E}} + (I_{\text{P}} + Q)/T_{\text{P}} - (I_{\text{E}} + I_{\text{P}})/T_{\text{sun}}$. No entropy is produced by the emission of radiation, since the effective temperature of thermal radiation is equal to the temperature of the body emitting it. The outgoing radiation remains at this temperature as it propagates into deep space. (Though over very long time scales, if it is not absorbed by another body, its effective temperature will be gradually be reduced by the expansion of space.)

As in Section 3.2.1 we can substitute the expressions for T_{E} and T_{P} to find σ_{total} as a function of Q. A little calculation shows that its maximum occurs at $Q = \frac{1}{2}(I_{\text{E}} - I_{\text{P}})$, which is the maximum possible value for Q, corresponding to a state in which the heat flow is so fast that the temperatures of the tropical and polar regions become equal. This is a different prediction for Q than that made by maximising σ_Q. Indeed, for this value of Q, σ_Q becomes zero.

Thus the results we obtain from applying the MEPP to the Earth system depend on where we draw the boundary. If we draw the system's boundary so as to include the atmosphere's heat transport but not the radiative effects then we obtain a prediction that matches the real data, but if we draw the boundary further out, so that we are considering a system that includes the Earth's surrounding radiation field as well as the Earth itself, then we obtain an incorrect (and somewhat physically implausible) answer.

This problem is not specific to the Earth system. In general, when applying the MEPP we must choose a specific bounded region of space to call the system, and the choice we make will affect the predictions we obtain by using the principle. If the MEPP is to be applied to more general systems we need some general method of determining where the system boundary must be drawn.

3.2.4 Some Comments on Dewar's Approach

Several authors have recently addressed the theoretical justification of the MEPP (Dewar, 2003, 2005a; Attard, 2006; Niven, 2009; Zupanovic et al., 2006); see (Martyushev & Seleznev, 2006) for a review. Each of these authors takes a different approach, and so it is fair to say there is currently no widely accepted theoretical justification for the principle.

The approach presented here most closely resembles that of Dewar (2003, 2005a). Dewar's work builds upon Edwin Jaynes' maximum entropy interpretation of thermodynamics (Jaynes, 1957a, 1957b). Jaynes' approach, discussed below, is to look at thermodynamics as a special case of a general form of Bayesian reasoning, the principle of maximum entropy (*MaxEnt*, not to be confused with the maximum entropy production principle).

Dewar offers two separate derivations of the MEPP from Jaynes' MaxEnt principle, one in (Dewar, 2003) and the other in (Dewar, 2005a). Both derive from Jaynes' formalism, which is probably the fundamentally correct approach to non-equilibrium statistical mechanics.

Dewar's approach consists of applying Jaynes' principle to the microscopic phase space of a system over a finite period of time. This gives rise to a probability distribution whose elements are possible microscopic dynamical trajectories of the system over the given time period, rather than instantaneous microscopic states. This is equivalent to considering instantaneous states subject to constraints formed by the system's history, as discussed in (Dewar, 2005b). The entropy of this probability distribution is maximised subject to constraints formed not of expected amounts of conserved quantities (as in Jaynes' formalism) but of expected rates of flow of conserved quantities across the system's boundary, along with its initial conditions. I will denote the i^{th} flow across the boundary by Q_i, although in (Dewar, 2003) the boundary is considered continuous and has a flow of each quantity at every point.

As shown in Chapter 2 of this thesis, when the information entropy of a probability distribution is maximised subject to mean value constraints, a new quantity arises for each constraint, which I refer to as tempers. I will denote the temper conjugate to the i^{th} flow μ_i.

Dewar's argument consists of two parts. First, he shows that the rate of production of thermodynamic entropy is proportional to $\sum_i \mu_i Q_i$, and then he shows, using approximations, that maximising the information entropy is equivalent to maximising this quantity. The first part of this argument is given in (Dewar, 2003), and the second is given in two different forms, using different approximations, in (Dewar, 2003) and (Dewar, 2005a).

Unfortunately I have never been able to follow the step in Dewar's argument (2003, equations 11–14), which shows that the quantity $\sum_i \mu_i Q_i$ is proportional to the entropy production. It is difficult to criticise an argument from the position of not understanding it, but it seems that both of Dewar's derivations suffer from two important drawbacks. Firstly, they both rely on approximations whose applicability is not clearly spelled out. I am not able to judge the validity of these approximations. However, the second problem is that neither of Dewar's derivations offers a satisfactory answer to the question of how to determine the appropriate system boundary. This is important because, as discussed above, the MEPP makes different predictions depending on where the boundary is drawn.

In (Dewar, 2003), Dewar writes that his derivation tells us which entropy production should be maximised when considering the Earth's atmosphere: the material entropy production, excluding the radiative components, which are to be treated as an imposed constraint. But there seems to be nothing in the maths which implies this. Indeed, in (Dewar, 2005a) the principle is presented as so general as to apply to all systems with reversible microscopic laws — which is to say, all known physical systems — and this condition clearly applies as much to the extended earth system, with the surrounding radiation field included, as it does to the atmosphere alone. Dewar's approach therefore does not solve the system boundary problem and cannot apply universally to all systems, since this would give rise to contradictory results. Dewar's derivations may well be largely correct, but there is still a piece missing from the puzzle.

In the following sections I will outline a new approach to the MEPP which I believe may be able to avoid these problems. Although not yet formal, this new approach has an intuitively reasonable interpretation, and may be able to give a constructive answer to the system boundary problem.

3.3 Thermodynamics as Maximum Entropy Inference

Both Dewar's approach and the approach I wish to present here build upon the work of Edwin Jaynes. One of Jaynes' major achievements was to take the foundations of statistical mechanics, laid down by Boltzmann and Gibbs, and put them on a secure logical footing based on Bayesian probability and information theory (Jaynes, 1957a, 1957b, 1965, 1979, 1980, 1985, 2003).

On one level this is just a piece of conceptual tidying up: it could be seen as simply representing a different justification of the Boltzmann distribution, thereafter changing the mathematics of statistical mechanics very little, except perhaps to express it in a more elegant form. However, on another level it can be seen as unifying a great number of results in equilibrium and non-equilibrium statistical mechanics, putting them on a common theoretical footing. In this respect Jaynes' work should, in the present author's opinion, rank among the great unifying theories of 20^{th} century physics. It also gives us a powerful logical tool — maximum entropy inference — that has many uses outside of physics. Most importantly for our purposes, it frees us from the near-equilibrium assumptions that are usually used to derive Boltzmann-type distributions, allowing us to apply the basic methodology of maximising entropy subject to constraints to systems arbitrarily far from thermal equilibrium.

Jaynes gives full descriptions of his technique in (Jaynes, 1957a, 1957b), with further discussion in (Jaynes, 1965, 1979, 1980, 1985, 2003). The mathematical details of Jaynes' approach were discussed in the Chapter 2. I will briefly summarise its logical basis again here, because of its importance to the arguments that follow.

Like all approaches to statistical mechanics, Jaynes' is concerned with a probability distribution over the possible microscopic states of a physical system. However, in Jaynes' approach we always consider this to be an "epistemic" or Bayesian probability distribution: it does not necessarily represent the amount of time the system spends in each state, but is instead used to represent our state of knowledge about the microstate, given the measurements we are able to take. This

knowledge is usually very limited: in the case of a gas in a chamber, our knowledge consists of the macroscopic quantities that we can measure — volume, pressure, etc. These values put constraints on the possible arrangements of the positions and velocities of the molecules in the box, but the number of possible arrangements compatible with these constraints is still enormous.

The probability distribution over the microstates is constructed by choosing the distribution with the greatest information entropy $-\sum_i p_i \log_2 p_i$, while remaining compatible with these constraints. This is easy to justify: changing the probability distribution in a way that reduces the entropy by n bits corresponds to making n bits of additional assumptions about which microscopic state the system might be in. Maximising the entropy subject to constraints formed by our knowledge creates a distribution that, in a formal sense, encodes this knowledge but no additional unnecessary assumptions.

Having done this, we can then use the probability distribution to make predictions about the system (for example, by finding the expected value of some quantity that we did not measure originally). There is nothing that guarantees that these predictions will be confirmed by further measurement. If they are not then it follows that our original measurements did not give us enough knowledge to predict the result of the additional measurement, and there is therefore a further constraint that must be taken into account when maximising the entropy. Only in the case where sufficient constraints have been taken into account to make good predictions of all the quantities we can measure should we expect the probability distribution to correspond to the fraction of time the system spends in each state.

With this logical basis understood, the formalism of statistical mechanics arises naturally from the application of maximum entropy inference to physical quantities such as energy, volume and particle number. There are some important subtleties regarding the relationship between information entropy and physical entropy, and the status of the second law, which are discussed in Chapter 2.

3.3.1 A New Argument for the MEPP

Once Jaynes' approach to thermodynamics is understood, the argument I wish to present for the MEPP is so simple as to be almost trivial. However, there are some aspects to it which have proven difficult to formalise, and hence it must be seen as a somewhat tentative sketch of what the true proof might look like.

Jaynes' argument tells us that to make the best possible predictions about a physical system, we must use the probability distribution over the system's microstates that has the maximum possible entropy while remaining compatible with any macroscopic knowledge we have of the system. Now let us suppose that we want to make predictions about the properties of a physical system at time t_1, in the future. Our knowledge of the system consists of measurements made at the present time, t_0, and some possibly incomplete knowledge of the system's kinetics.

The application of Jaynes' procedure in this case seems conceptually quite straight-forward: we maximise the entropy of the system we want to make predictions about — the system at time t_1 — subject to constraints formed by the knowledge we have about it. These constraints now

include not only the values of conserved quantities but also restrictions on how the system can change over time.

If the system in question is isolated then it cannot export entropy to its surroundings. In this case, if we consider the entropy at time t_0 to be a fixed function of the constraints then maximising the entropy at t_1 corresponds to maximising the rate at which entropy is produced between t_0 and t_1. Note that no steady state assumption is required to justify this procedure, and so this approach suggests that the MEPP may be applicable to transient behaviour as well as to systems in a steady state.

This appears to show that, for isolated systems at least, maximum entropy production is a simple corollary of the maximum entropy principle. However, it should be noted that the information entropy which we maximise at time t_1 is not necessarily equal to the thermodynamic entropy at t_1 as usually defined. Let us denote by $H(t)$ the entropy of the probability distribution over the system's microstates at time t, when maximised subject to constraints formed by measurements made at time t_0, together with our knowledge about the system's microscopic evolution. Jaynes called this quantity the caliber (Jaynes, 1985). In contrast, let $S(t)$ denote the thermodynamic entropy at time t, which is usually thought of as the information entropy when maximised subject to constraints formed by measurements made at time t only. Then $H(t_0) = S(t_0)$, but $H(t_1)$ can be less than $S(t_1)$. In particular, if we have complete knowledge of the microscopic dynamics then $H(t_1) = H(t_0)$ by Liouville's theorem, which always applies to the microscopic dynamics of an isolated physical system, even if the system is far from thermodynamic equilibrium.

The procedure thus obtained consists of maximising the thermodynamic entropy subject to constraints acting on the system not only at the present time but also in the past. Our knowledge of the system's kinetics also acts as a constraint when performing such a calculation. This is important because in general systems take some time to respond to a change in their surroundings. In this approach this is dealt with by including our knowledge of the relaxation time among the constraints we apply when maximising the entropy.

There is therefore a principle that could be called "maximum entropy production" which can be derived from Jaynes' approach to thermodynamics, but it remains to be shown in which, if any, circumstances it corresponds to maximising the thermodynamic entropy. From this point on I will assume that for large macroscopic systems, maximising $H(t_1)$ and $S(t_1)$ can be considered approximately equivalent, in the sense that maximising the thermodynamic entropy at time t_1 yields the same results as maximising the information entropy. This corresponds to an assumption that all relevant macroscopic properties of the system can be predicted just as well using the distribution formed by maximising $S(t_1)$, which encodes information only about the system's macroscopic state at time t_1, as it can using the distribution attained by maximising $H(t_1)$, which contains additional information about the system's macrostate at previous times. The conditions under which this step is valid remain to be shown.

This suggests that applying Jaynes' procedure to make predictions about the future properties of a system yields a principle of maximum production of thermodynamic entropy: it says that the best predictions about the behaviour of isolated systems can be made by assuming that the

approach to equilibrium takes place as rapidly as possible. However it should be noted that this argument does not solve the system boundary problem in the way that we might like, since it applies only to isolated systems. The heat transport in the Earth's atmosphere is not an isolated system, but one can draw a boundary around the Solar system such that it is isolated, to a reasonable approximation. Thus this argument seems to suggest unambiguously that the MEPP should be applied to the whole Solar system and not to Earth's atmosphere alone — but this gives incorrect predictions, as shown in Section 3.2.3. A possible solution to this problem will be discussed in Section 3.4

3.3.2 Application to the Steady State with a Fixed Gradient

In this section I will demonstrate the application of this technique in the special case of a system with fixed boundary conditions. I will use a technique similar to the one described by Jaynes (1980), whereby a system which could be arbitrarily far from equilibrium is coupled to a number of systems which are in states close to equilibrium, allowing the rate of entropy production to be calculated. In Section 3.4 I will extend this result to the more general case of negative feedback boundary conditions.

Consider a system consisting of two large heat baths **A** and **B**, coupled by another system **C** that can transfer energy between them, about which we have some limited knowledge. Our assumption will be that **A** and **B** are relatively close to thermodynamic equilibrium, so that it makes sense to characterise them as having definite temperatures $T_{\mathbf{A}}, T_{\mathbf{B}}$ and internal energies $U_{\mathbf{A}}, U_{\mathbf{B}}$. This need not be the case for **C**, which could be arbitrarily complex and involve many interacting far-from-equilibrium processes. The only assumption we will make about **C** is that it can be considered to be in a steady state on the time scale under consideration.

We will consider the evolution of the system over a time period from some arbitrary time t_0 to another time t_1. We will assume the duration of this time period to be such that the temperatures of **A** and **B** do not change appreciably.

At a given time t, the (maximised) information entropy of the whole system, denoted $H(t)$, is given by the thermodynamic entropy of the reservoirs, $U_{\mathbf{A}}(t)/T_{\mathbf{A}} + U_{\mathbf{B}}(t)/T_{\mathbf{B}}$, plus the generalised entropy of **C**. By "the generalised entropy of **C**" I mean in this context the entropy of a probability distribution over the possible microstates of **C**, when maximised subject to all knowledge we have about the kinetics of **C**, together with the boundary conditions of constant $T_{\mathbf{A}}$ and $T_{\mathbf{B}}$. This is a generalisation of the usual thermodynamic entropy and does not necessarily have the same numerical value. This generalised entropy is maximised subject to constraints that include information about the system's past as well as its present macroscopic state, so in general it will be lower. Since **C** is potentially a very complex system this generalised entropy may be difficult if not impossible to calculate. However, the steady state assumption implies that our knowledge of the system is independent of time. Maximising the entropy of a probability distribution over **C**'s microstates should give us the same result at t_1 as it does for t_0. The generalised entropy of **C** is thus constant over time.

We now treat our knowledge of the state of the system (including the reservoirs) at time t_0 as fixed. We wish to make a maximum entropy prediction of the system's state at time t_1, given this

knowledge. Since $H(t_0)$ is fixed, maximising $H(t_1)$ is equivalent to maximising

$$H(t_1) - H(t_0) = \frac{U_{\mathbf{A}}(t_1) - U_{\mathbf{A}}(t_0)}{T_{\mathbf{A}}} + \frac{U_{\mathbf{B}}(t_1) - U_{\mathbf{B}}(t_0)}{T_{\mathbf{B}}} = Q\left(\frac{1}{T_B} - \frac{1}{T_A}\right)(t_1 - t_0), \tag{3.3}$$

where Q is the steady state rate of heat transport from **A** to **B**. Since $T_{\mathbf{A}}$ and $T_{\mathbf{B}}$ are fixed, this corresponds to maximising Q subject to whatever constraints our knowledge of **C**'s kinetics puts upon it. We have therefore recovered a maximum flow principle from Jaynes' maximum entropy principle.

This reasoning goes through in exactly the same way if the flow is of some other conserved quantity rather than energy. It can also be carried out for multiple flows, where the constraints can include trade-offs between the different steady-state flow rates. In this case the quantity which must be maximised is the steady state rate of entropy production. This is less general than the MEPP as used in climate science, however, since this argument assumed fixed gradient boundary conditions.

3.4 A Possible Solution to the System Boundary Problem

The argument presented in Section 3.3.1 does not in itself solve the System Boundary Problem in a way that justifies the MEPP as it is used in climate science. Instead, it seems to show that only the entropy production of isolated systems is maximised, in the sense that the approach to equilibrium happens as rapidly as possible, subject to kinetic constraints. Planetary atmospheres are not isolated systems, but, to a reasonable degree of approximation, one can consider the Solar system, plus its outgoing radiation field, to be an isolated system. It thus seems at first sight that we should maximise the total entropy production of the Solar system in order to make the best predictions. As shown in Section 3.2.3, this results in a prediction of maximum flow, rather than maximum entropy production, for planetary atmospheres.

It could be that this reasoning is simply correct, and that there is just something special about planetary atmospheres which happens to make their rates of heat flow match the predictions of the MEPP. However, in this section I will try to argue that there is an extra constraint which must be taken into account when performing this procedure on systems with negative feedback boundary conditions. This constraint arises from the observation that the system's behaviour should be invariant with respect to certain details of its environment. I will argue that when this constraint is taken account of, we do indeed end up with something that is equivalent to the MEPP as applied in climate science.

We now wish to apply the argument described in Section 3.3.1 to a situation analogous to the two-box atmosphere model. Consider, then, a system consisting of two finite heat reservoirs, **A** and **B**, coupled by an arbitrarily complex sub-system **C** that can transport heat between them. This time the reservoirs are to be considered small enough that their temperatures can change over the time scale considered. The system as a whole is not to be considered isolated, but instead each reservoir is subjected to an external process which heats or cools it at a rate that depends only on its temperature (this is, of course, an abstraction of the two-box atmospheric model). Let

$T_{\mathbf{A}}$ and $T_{\mathbf{B}}$ stand for the temperatures of the two reservoirs and Q be the rate of the process of heat transport from **A** to **B**. Let $Q_{\mathbf{A}}(T_{\mathbf{A}})$ be the rate of heat flow from an external source into **A**, and $Q_{\mathbf{B}}(T_{\mathbf{B}})$ be the rate of heat flow from B into an external sink. In the case of negative feedback boundary conditions, $Q_{\mathbf{A}}$ is a decreasing function and $Q_{\mathbf{B}}$ increasing.

Suppose that our knowledge of this system consists of the functions $Q_{\mathbf{A}}$ and $Q_{\mathbf{B}}$, together with the conservation of energy, and nothing else: we wish to make the best possible predictions about the rate of heat flow Q without taking into account any specific knowledge of the system's internal kinetics. As in Jaynes' procedure, this prediction is not guaranteed to be correct. If it is then it should imply that the sub-system responsible for the heat transport is so under-constrained that information about the boundary conditions is all that is needed to predict the transport rate.

We cannot directly apply the procedure from Section 3.3.1 because this system is not isolated. We must therefore postulate a larger system to which the system $\mathbf{A}+\mathbf{B}+\mathbf{C}$ is coupled, and treat the extended system as isolated. However, there is some ambiguity as to how this can be done: there are many different extended systems that would produce the same boundary conditions on **C**. In general entropy will be produced in the extended system as well as inside **C**, and the way in which the total entropy production varies with Q will depend on the precise configuration of the extended system.

Figure 3.2 shows two examples of possible extended systems. The first is somewhat analogous to the real situation of the Earth's atmosphere. Heat flows into **A** from a much larger reservoir of constant temperature $\mathbf{A}'$, which is roughly analogous to the solar radiation field, and out of **B** into a large reservoir $\mathbf{B}'$, which can be thought of as representing deep space. The second law implies that the temperature of $\mathbf{A}'$ must be greater than the maximum possible temperature of **A**, and that of $\mathbf{B}'$ must be less than **B**'s minimum temperature. This in turn means that entropy is produced when heat is transferred from $\mathbf{A}'$ to **A** and from **B** to $\mathbf{B}'$. In this scenario, the total entropy production is maximised when Q is maximised.

Figure 3.2(b) shows another, more complicated possibility. In this scenario the functions $Q_{\mathbf{A}}(T_{\mathbf{A}})$ and $Q_{\mathbf{B}}(T_{\mathbf{B}})$ are the same, but the flows are caused by a different mechanism. A heat engine extracts work using the temperature gradient between two external reservoirs (now denoted $\mathbf{A}''$ and $\mathbf{B}''$). This work is used by reversible heat pumps to extract heat from **B** and to add heat to **A**. This is done at a carefully controlled rate so that the rates of flow depend on the temperature of **A** and **B** in exactly the same way as in the other scenario.

Since all processes in this version of the external environment are reversible, the only entropy produced in the extended system is that due to Q, the heat flow across **C**. The boundary conditions on **C** are the same, but now the maximum in the total entropy production occurs when the entropy production within **C** is maximised, at a value for Q somewhere between 0 and its maximum possible value.

Although the first of these scenarios seems quite natural and the second somewhat contrived, it seems odd that the MEPP should make different predictions for each of them. After all, their effect on the system **C** is, by construction, identical. There is no way that system **C** can "know" which of these possible extended systems is the case, and hence no plausible way in which its

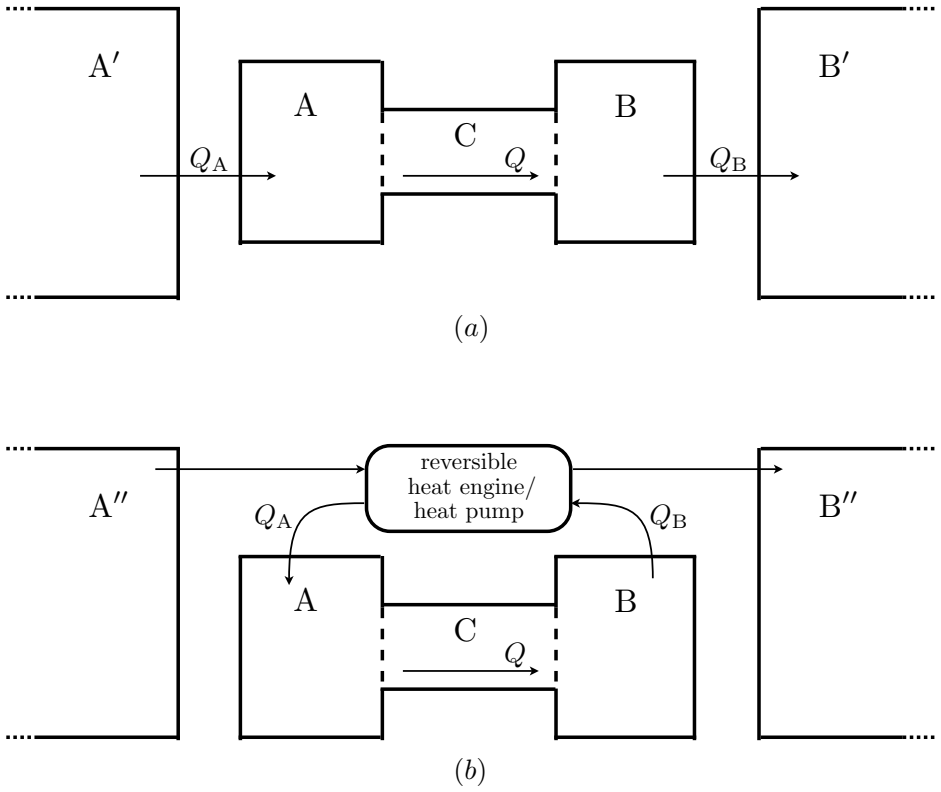


Figure 3.2: Two possible ways in which negative feedback boundary conditions could be realised. (a) Heat flows into **A** from a much larger reservoir $\mathbf{A}'$, and out of **B** into another large reservoir $\mathbf{B}'$. (b) A reversible heat engine is used to extract power by transferring heat between two external reservoirs $\mathbf{A}''$ and $\mathbf{B}''$, and this power is used to transfer heat reversibly out of **B** and into **A**. In this version the only entropy produced is that produced inside **C** due to the heat flow Q.

steady state rate of heat flow could depend on this factor. This observation can be made even more striking by noting that one could construct an extended system for which the total entropy production is maximised for any given value of Q, simply by connecting a device which measures the temperatures of **A** and **B** and, if they are at the appropriate temperatures, triggers some arbitrary process that produces entropy at a high rate, but which does not have any effect upon **C**.

Clearly an extra principle is needed when applying the MEPP to systems of this type: the rate of heat flow Q should be independent of processes outside of the system $\mathbf{A}+\mathbf{B}+\mathbf{C}$ which cannot affect it. This can be thought of as an additional piece of knowledge we have about the extended system, which induces an additional constraint upon the probability distribution over its microstates.

The question now is how to take account of such a constraint when maximising the entropy. It seems that the solution must be to maximise the information entropy of a probability distribution over all possible states of the extended system, subject to knowledge of the system **C**, the functions $Q_{\mathbf{A}}(T_{\mathbf{A}})$ and $Q_{\mathbf{B}}(T_{\mathbf{B}})$, and nothing else. We must ignore any knowledge we may have of processes taking place outside of **A**, **B** and **C**, because we know that those processes cannot affect the operation of **C**.

But this again presents us with a problem that can be solved by maximising an entropy. We know that outside the system $\mathbf{A}+\mathbf{B}+\mathbf{C}$ there is another system, which we might collectively call **E**, which is capable of supplying a constant flow of energy over a long period of time. But this is all we know about it (or at least, we are deliberately ignoring any other knowledge we may have about it). Jaynes' principle tells us that, if we want to make the minimal necessary set of assumptions, we should treat it as though its entropy is maximised subject to constraints formed by this knowledge.

Let us consider again the two scenarios depicted in Figure 3.2. In the first, the temperature of $\mathbf{A}'$ is constrained by the second law to be substantially higher than that of $\mathbf{B}'$. If **E** is to cause a flow of K units of energy through system **C**, its entropy must increase by at least $K(1/T_{\mathbf{B}'}-1/T_{\mathbf{A}'})$, an amount substantially greater than the entropy produced inside **C**. The total entropy of **E** is of the order $U_{\mathbf{A}'}/T_{\mathbf{A}'}+U_{\mathbf{B}'}/T_{\mathbf{B}'}$, where $U_{\mathbf{A}'}$ and $U_{\mathbf{B}'}$ are the (very large) internal energies of $\mathbf{A}'$ and $\mathbf{B}'$.

The situation in the second scenario is different. In this case the temperatures of $\mathbf{A}''$ and $\mathbf{B}''$ can be made arbitrarily close to one another. This suggests that in some sense, the entropy of the external environment, which is of the order $U_{\mathbf{A}''}/T_{\mathbf{A}''}+U_{\mathbf{B}''}/T_{\mathbf{B}''}$, can be higher than in the other scenario. (The machinery required to apply the flow reversibly will presumably have a complex structure, and hence a low entropy. But the size of machinery required does not scale with the amount of energy to be supplied, so we may assume that the size of this term is negligible compared to the entropy of the reservoirs.) In this scenario, if K units of energy are to flow through **C** (assumed for the moment to be in the steady state), the entropy of **E** need only increase by $K(1/T_{\mathbf{B}}-1/T_{\mathbf{A}})$, which is precisely the amount of entropy that is produced inside **C**.

I now wish to claim that, if we maximise the entropy of **E**, we end up with something which behaves like the second of these scenarios, supplying a flow of heat through $\mathbf{A}+\mathbf{B}+\mathbf{C}$ reversibly. To see this, start by assuming, somewhat arbitrarily, that we know the volume, composition and internal energy of **E**, but nothing else. Equilibrium thermodynamics can then (in principle) be used to tell us the macroscopic state of **E** in the usual way, by maximising the entropy subject to these constraints. Let us denote the entropy thus obtained by $S_{\mathbf{E},\max}$.

But we now wish to add one more constraint to **E**, which is that it is capable of supplying a flow of energy through $\mathbf{A}+\mathbf{B}+\mathbf{C}$ over a given time period. Suppose that the total amount of entropy exported from **C** during this time period is s (this does not depend on the configuration of **E**, by assumption). The constraint on **E**, then, is that it is capable of absorbing those s units of entropy. Its entropy must therefore be at most $S_{\mathbf{E},\max}-s$. Again there will be another negative term due to the configurational entropy of any machine required to actually transport energy to and from $\mathbf{A}+\mathbf{B}+\mathbf{C}$, but this should become negligible for large s.

Maximising the entropy of a large environment **E** subject to the boundary conditions of **C** over a long time scale thus gives an entropy of $S_{\mathbf{E},\max}-s$. But if **E** has an entropy of $S_{\mathbf{E},\max}-s$ then it cannot produce any additional entropy in supplying the flow of heat through **C**, because the only entropy it is capable of absorbing is the entropy produced inside **C**. Therefore, if we maximise the entropy of **E** subject to these constraints we must end up with an environment which behaves like the one shown in Figure 3.2(b), supplying the flow of energy reversibly.

This will be true no matter which constraints we initially chose for the volume, internal energy and composition of **E**, so we can now disregard those constraints and claim that in general, if all we know about **E** is that it is capable of supplying a flow of energy through **C**, we should assume that it does so reversibly, producing no additional entropy as it does so. This is essentially because the vast majority of possible environments capable of supplying such a heat flow do so all-but-reversibly.

This result gives us the procedure we need if we want to maximise the total entropy production of the whole system ($\mathbf{A}+\mathbf{B}+\mathbf{C}+\mathbf{E}$), subject to knowledge only of **C** and its boundary conditions, disregarding any prior knowledge of **E**. In this case we should assume that there is no additional contribution to the entropy production due to processes in **E**, and thus count only the entropy production due to the heat flow across **C**.

Therefore, completing the procedure described in Section 3.3.1 subject to the constraint of the invariance of **C** with respect to the details of **E**, we obtain the result that the best possible predictions about **C** should be made by maximising the entropy production of **C** alone. This procedure should be followed even if we have definite *a priori* knowledge about the rate of entropy production in **E**, as long as we are sure that the details of **E** should not affect the rate of entropy production in **C**.

3.4.1 Application to Atmospheres and Other Systems

In this section I have presented an argument for a way in which the system boundary problem might be resolved. If this turns out the be the correct way to resolve the system boundary problem then its application to the Earth's atmosphere is equivalent to a hypothesis that the atmosphere would behave in much the same way if heat were supplied to it reversibly rather than as the result of the absorption of high-temperature electromagnetic radiation. It is this invariance with respect to the way in which the boundary conditions are applied that singles the atmosphere out as the correct system to consider when applying MEPP. However, the argument could also be applied in reverse: the fact that the measured rate of heat flow matches the MEPP prediction in this case could be counted as empirical evidence that the atmosphere's behaviour is independent of the way in which heat is supplied to it.

It is worth mentioning again that although I have presented this argument in terms of heat flows and temperature, it goes through in much the same way if these are replaced by other types of flow and their corresponding thermodynamic forces. Thus the argument sketched here, if correct, should apply quite generally, allowing the same reasoning to be applied to a very wide range of systems, including complex networks of chemical processes driven by a difference in chemical potential, which will be the subject of the next chapter. The method is always to vary some unknown parameters so as to maximise the rate of entropy production of a system. The system should be chosen to be as inclusive as possible, while excluding any processes that cannot affect the parameters' values.

3.5 Discussion

There are two issues which deserve further discussion. The first is the need for experimental study of the MEPP phenomenon, which will help us to judge the validity of theoretical approaches, and the second is the way in which a universally applicable principle of maximum entropy production would affect the way we see the relationship between thermodynamics and kinetics.

3.5.1 The Need for Experimental Study

It is likely to be difficult to determine the correctness of theoretical approaches to the MEPP without experimental validation of the idea. Currently the main empirical evidence comes from measurements of the atmospheres of Earth and other bodies, but this gives us only a small number of data points. There have also been some simulation-based studies such as (Shimokawa & Ozawa, 2005), which focuses on the probability of transitions between distinct steady states of a general circulation model of the Earth system, showing that transitions to steady states with a higher rate of entropy production appear to be more probable.

However, to my knowledge, there is not currently any experimental work which focuses on applying negative feedback boundary conditions to a system. It seems that it should be possible to subject a potentially complex system such as a large volume of fluid to these conditions using an experimental set-up which, in essence, could look like that shown in Figure 3.2(a). This is the type of condition under which we can use the MEPP to make predictions about planetary atmospheres, and if the MEPP is a general principle then we should be able to reproduce this type of result in the lab. In a laboratory set-up, the exact nature of the boundary conditions (the functions $Q_{\mathbf{A}}(T_{\mathbf{A}})$ and $Q_{\mathbf{B}}(T_{\mathbf{B}})$) could be varied, which would give a potentially infinite number of data points, rather than the few we currently have from studies of planetary bodies.

There is of course the problem that, regardless of what theoretical justification one might accept, the MEPP cannot make correct predictions about all systems under negative feedback boundary conditions. In the approach presented here, as well as in Dewar's approach, this is because the entropy production is maximised subject to constraints. Thus if the rate of heat transport in an experimental set-up failed to match the predictions of the MEPP it could be because of additional constraints acting on the system, in addition to the imposed boundary conditions, rather than because the whole theory is wrong.

However, a positive experimental result — some situation in which the MEPP can be used to make accurate predictions about the behaviour of a system under imposed negative feedback boundary conditions — would be of great help to the development of the theory, as well as justifying its use in practice.

One particularly useful experiment would be to vary the experimental setup in various ways until the MEPP predictions based only on the boundary conditions were no longer accurate. This would help to produce a more precise understanding of what is required for a system to be unconstrained enough for the MEPP to apply.

3.5.2 The Relationship Between Thermodynamics and Kinetics

Currently, thermodynamics and kinetics are thought of as somewhat separate subjects. Thermodynamics, in its usual interpretation, tells us that entropy must increase but says nothing about how fast. The study of the rates of chemical reactions and other physical processes is called kinetics, and is thought of as being constrained but not determined by thermodynamics.

If there is a general principle of maximum entropy production which works along the lines described in the present chapter then this relationship can be cast in a rather different light. Now the kinetics act as constraints upon the increase of entropy, which always happens at the fastest possible rate. This change in perspective may have an important practical consequence.

In Section 3.3 I sketched the derivation of the ensembles of equilibrium statistical mechanics as if the possible states of the system could be enumerated and the energy level of each one known *a priori*, and the relationship between energy and temperature then derived from the known values of these energy levels.

But this is not how thermodynamics works in practice. The possible energy levels of a large system are usually unknown, but the relationship between the amount of heat added to a system and its temperature can be measured using calorimetry. From this we work backwards, using Equation 2.19 to make inferences about the system's energy levels, and hence about its microscopic structure in general.

But if the MEPP is a consequence of Jaynes' maximum entropy principle then we should also be able to work backwards from the kinetics of a system to make inferences about its microscopic dynamics.

The methodology would involve something along the lines of a procedure which determines what kind of constraints the observed kinetics put on the underlying microscopic dynamics, and then maximising the entropy production subject to those constraints in order to produce the best possible predictions about the system's behaviour under other circumstances.

The development of a principled algorithm for doing this is a subject for future research, but the idea could have lasting implications for the way in which the microscopic dynamics of physical and chemical systems are studied.

3.6 Conclusion

In this chapter I have discussed the general form of the maximum entropy production principle as it is used in climate science, emphasising the notion of negative feedback boundary conditions, the presence of which allows a unique maximum when the entropy function is varied with a flow rate. I have also emphasised that the principle of maximum entropy production has a serious conceptual difficulty in that it gives different results depending on which part of a physical system it is applied to. I have sketched the beginnings a theoretical approach, based on the MaxEnt principle of Edwin Jaynes, which shows some promise of solving this problem.

Although the theoretical derivation of the MEPP presented here is somewhat tentative, if it or something like it turns out to be correct it will have wide-ranging implications for the study

of non-equilibrium systems. This lends great importance to future theoretical and experimental work.

Chapter 4

Entropy Production in Ecosystems

4.1 Introduction

In the previous chapter I defined the notion of *negative feedback boundary conditions*. In this chapter we will see how such boundary conditions arise naturally in ecological scenarios, in such a way that the negative feedback apples to the ecosystem as a whole. I present a minimal model of a chemical environment which exhibits such conditions, and develop a simple but general model of an evolving population of organisms within it. This puts definite physical bounds on the rates at which matter can flow through the system and paves the way for more detailed models that have thermodynamic principles built in from the start.

It should be noted that the model as presented does not exhibit a maximisation of entropy production. This may be because ecosystems are not unconstrained enough for the MEPP to apply to them, or it may be because the model is too simple, lacking features that would lead to the maximum entropy production state.

One consequence of negative feedback boundary conditions is that, in the absence of organisms, the free-energy density of the environment can become very high, meaning that organisms (or proto-organisms) with very slow and inefficient metabolisms can persist. I argue that this may have been the case on the early Earth, suggesting that early organisms could have had much simpler structures than today's. As the speed and efficiency of metabolisms increases through evolution the environment's energy density drops, so that only the fastest and most efficient metabolisers are able to survive. Thus this negative feedback through the environment provides a ratchet effect whereby organisms can progress from simple, inefficient structures to the complex, highly specified and efficient metabolisms we observe today.

The model is formally analogous to the R. D. Lorenz et al. (2001) two-box model of planetary atmospheres, which suggests that, if there is a general physical principle of maximum entropy production, it may be applicable to ecosystems. This would mean that the MEPP could be used to make numerical predictions about ecosystems, and hence could be tested experimentally. However, the evolutionary model I present does not provide a mechanism by which the peak in entropy

production would be attained, because the evolutionary pressure is always towards in increase in the flow of matter in the system, regardless of whether this results in an increase or a decrease in the entropy production.

There are, however, many processes in real ecosystems, including predation, parasitism, nutrient limitation and higher-level selection that could limit this growth in the rate of flux. The effect of each of these factors will be discussed. It is possible that some of them, or combinations of them, could prevent the flux from increasing once the entropy production has reached its peak, and for this reason the applicability of MEPP to ecosystems must remain an open question.

This chapter is in part based on (Virgo & Harvey, 2007) but has been substantially extended since that publication. In particular the evolutionary model developed in Section 4.3 is published for the first time in this thesis.

4.2 Negative Feedback Boundary Conditions in Ecosystems

In the previous chapter I defined negative feedback boundary conditions in the context of heat transport in the Earth's atmosphere. This heat transport is powered by a difference between two temperatures. The rate of matter flow through a chemotrophic ecosystem is formally analogous to a flow of heat, but is powered by a difference between two chemical potentials rather than temperatures.

Chemical potential is somewhat analogous to temperature, as explained in Chapter 2. We can think of chemotrophic ecosystems as being powered by a difference in this quantity. Just as, in most circumstances, the temperature of a system increases or remains constant as energy is added to it, the chemical potential of a substance within a system usually increases with each mole of that substance added to the system. Concentrations tend to flow from areas of high chemical potential to areas of low chemical potential, so that in thermochemical equilibrium the chemical potentials are equal in all areas of the system.

Strictly speaking, the energy temper $\lambda_U = 1/T$ (see section 2.6) is analogous to the chemical temper of a substance X, $\lambda_X = -\mu_X/T$. We will be able to think of chemical potential and chemical temper as more or less interchangeable in this chapter, because we will be considering systems in which the environment's temperature is held constant (note that this assumption applies only to the environment; the organisms' body temperatures could differ from that of the environment without violating it; see Section 2.6.3).

The chemical potential allows us to predict the direction of flow between areas of differing concentrations, but it can be used to predict the direction of chemical reactions in the same way. If there is a reaction that converts between two substances X and Y, it will take place in the direction that moves the two chemical potentials together, so that at equilibrium they are equal. This corresponds to a maximum of the entropy of the system and its surroundings (or, equivalently, a minimum of the relevant free energy quantity — see Section 2.6.1). If there is a mixture of X and Y in which the chemical potential μ_X of X is greater than that of Y (μ_Y) then X will tend to be spontaneously converted into Y, producing entropy. Chemical work can be done by coupling this spontaneous reaction to another one. (See Section 2.6.2 for the definition of chemical work.)

All chemical reactions can take place in both 'forward' and 'backward' directions in this way but some reactions are experienced as unidirectional because the equilibrium point is such that when the chemical potentials are equal the proportions of reactants are vanishingly small compared to the products.

In an ideal solution the relationship between chemical potential μ_X and molar concentration N_X is given by $\mu_X = \mu_{X0} + RT \ln N_X$, where μ_{X0} is a constant that depends only on the temperature and the nature of X, and $R = N_A k_B = 8.31 \mathrm{JK}^{-1}\mathrm{mol}^{-1}$ is the gas constant. However, the exact form of this relationship does not matter for what follows. The reasoning presented here is quite general in that all it relies upon is that the chemical potential is an increasing function of the concentration. This is always the case in physical systems.

4.2.1 An Ecosystem under a Negative Feedback Constraint

In the two-box version of the atmospheric heat transport model described by R. D. Lorenz et al. (2001, see also Section 3.2.1 of this thesis), the Earth's atmosphere is subjected to boundary conditions in which there is a trade-off between the rate of heat flow and the temperature difference that powers it. The faster the flow, the lower the temperature difference. This trade-off is not a property of the system itself (in this case the turbulent flow of the atmosphere) but is imposed upon it by external forces (in this case, the balance of incoming radiation from the sun and outgoing radiation from the atmosphere).

In this chapter we are interested in ecosystems that are subject to a similar constraint. There are two reasons for wanting to study this possibility: firstly, as discussed in Chapter 3, this is the type of situation in which the maximum entropy production principle makes specific numerical predictions that could be tested experimentally; and secondly, this type of constraint is very common in natural situations, as first pointed out by H. T. Odum and Pinkerton (1955). We proceed by envisaging a highly simplified situation in which a population of organisms, or a whole ecosystem, would be subjected to such a constraint.

The basic components of an ecosystem are a population of organisms and an environment whose entropy can be increased, in this case a chemical mixture held out of equilibrium by some external process. In order for interesting behaviour to take place there must be feedback between the organisms and their environment. One simple way to achieve this, which I will work through in detail, is to imagine that food is transported into the system and waste out by diffusion, as shown in Figure 4.1.

The system contains a population of organisms (which need not be of the same species) whose metabolism is very simple, with individuals maintaining their structure by converting 'food' X into 'waste' Y. The system is coupled via a membrane to a reservoir in which the concentrations of X and Y remain constant, either because the reservoir is very large or because they are held constant by an externally powered process. The concentrations in the reservoir are such that there is a greater proportion of X than there would be at chemical equilibrium. X and Y diffuse through the membrane at rates that are proportional to the difference in their chemical potentials on either side of the membrane.

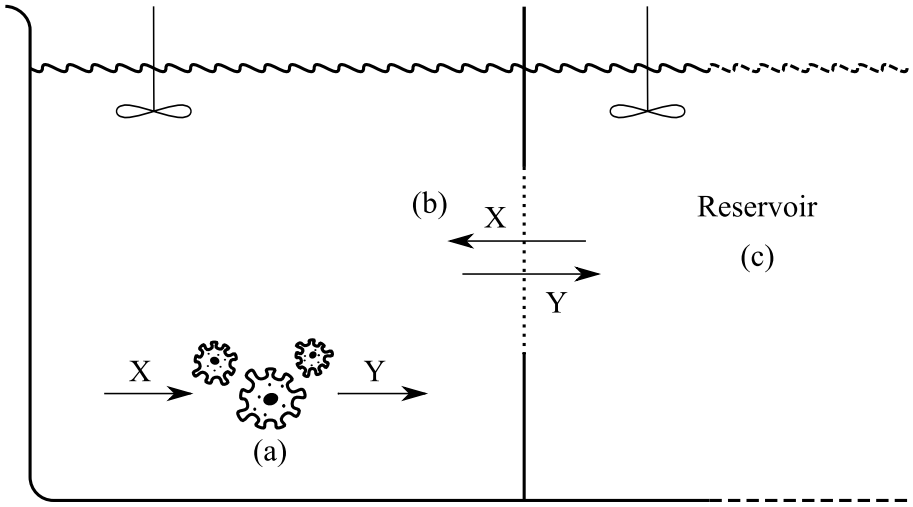


Figure 4.1: Important features of the ecosystem model. (a) The system consists of a chemical mixture containing a population of organisms that metabolise by converting food X into waste Y. (b) A membrane allows X and Y to flow in and out of the system at rates which depend on the difference in chemical potential on either side of the membrane. (c) On the other side of the membrane is a well-mixed reservoir in which the chemical concentrations of 'food' X and 'waste' Y are held at constant non-equilibrium values.

A critical question is which parts of this physical set-up we consider inside "the system" and which outside. This is important because different choices will result in different numerical answers if we apply the maximum entropy production principle.

Unless otherwise stated, in what follows we will take the system boundary to be to the left of the membrane in Figure 4.1, so that it includes the metabolism of all the organisms but not the membrane itself. The exclusion of the entropy produced by diffusion through the membrane is somewhat analogous to the exclusion of entropy produced by absorption of solar radiation in the studies by Paltridge (1979) and R. D. Lorenz et al. (2001). Whether this is justified, and what the justification might be, is an open question for the theory of the MEPP.

If the ideas presented in Chapter 3 are correct then this choice of system boundary can be justified in a similar manner to my proposed justification for excluding the incoming solar radiation from the Earth system when applying the MEPP. The idea is that, for the processes inside the system, i.e. the organisms' metabolisms, there is no way to determine that the changes in concentrations of X and Y are due to diffusion through a membrane, as opposed to some other process that happens to result in similar dynamics for the concentrations. One could imagine replacing the reservoir and membrane with a system analogous to that shown in Figure 3.2b, which achieves the same dynamics reversibly, and in this case the only entropy production would be due to the organisms' metabolism. The arguments in Chapter 3 imply that, because of this, we should choose the system boundary to lie inside the membrane when applying the MEPP.

According to the ideas developed in Section 3.4, choosing the boundary to be inside the membrane would be justified if, from the point of view of processes within the system, it is impossible to tell whether the constraints on the concentrations of X and Y inside the system are caused by the presence of a membrane and reservoir or by some other type process with similar dynamics. This

seems to be the case. However, it should be stressed again that the theory developed in Chapter 3 is somewhat tentative.

The system, defined by the processes that occur to the left membrane in Figure 4.1, is open to matter flow, since matter flows in in the form of X and out in the form of Y. If we assume that the system reaches a steady state, so that the flow in is equal to the flow out, then the system is closely analogous to the two-box atmospheric model used by R. D. Lorenz et al. (2001), as we will see below.

4.2.2 Population Metabolic Rate

In traditional ecosystem models the focus tends to be on the numbers of individuals or the biomass of each species. These are treated as dynamical variables, with rates of change that depend on the population numbers of other species, so for instance the population of a predator species will grow more rapidly if there are more of its prey species present.

We will develop such a population-dynamic model below, but for the time being our focus is on the flow of matter through the system, and the relevant variable to represent this is the *Population Metabolic Rate* M. We define this as the total rate at which food is converted into waste within the system. This takes place due to the combined effect of each individual's metabolism. There will of course be many other processes occurring within each organism's metabolism and within the system as a whole. In this scenario these are all ultimately powered by the conversion of X into Y and we can use the total conversion rate M to summarise the operation of the whole system.

Population metabolic rate is not necessarily directly related to population size. An organism can sustain itself either by having a structure that decays very slowly and using slow, efficient processes to maintain it or by having a rapidly decaying structure which must be renewed at a faster rate, so for a given population metabolic rate the population could consist of a small number of rapidly metabolising individuals or a large number that metabolise slowly; or there could be a small number which metabolise slowly as adults but which have a high rate of turnover. By analogy to the two-box climate model in (R. D. Lorenz et al., 2001) we are not for the moment concerned with what is inside the system, only with the rate at which it operates.

Another difference between this model and traditional models in ecology is that the population metabolic rate is treated not as a dynamical variable but as a parameter of the system. This gives the system a range of possible steady states (one for each value of M) which enables the use of the Maximum Entropy Production principle to choose between them.

4.2.3 Entropy Production in the Steady State

Because the concentrations of X and Y are held constant on the reservoir side of the membrane, their chemical potentials remain constant in the reservoir also. Their values μ_X^{res} and μ_Y^{res} are parameters in the model. Since the proportion of 'food' X compared to 'waste' Y is assumed to be higher than it would be in equilibrium, $\mu_X^{\text{res}} > \mu_Y^{\text{res}}$. The chemical potentials μ_X and μ_Y within the system are variables that depend on the population metabolic rate and the rate of diffusion through the membrane.

The rate at which a fluid flows through a membrane is in general a function of the differences in concentrations on either side of the membrane. For the time being we will make the (unrealistic but thermodynamically reasonable) simplifying assumption that the rate of flow of each substance across the membrane are proportional to the difference in tempers of the substance on either side of the membrane, i.e.

$$\begin{aligned} \dot{N}_X &= D_X \left(\mu_X^{\text{res}} - \mu_X\right)/T - M \\ \dot{N}_Y &= D_Y \left(\mu_Y^{\text{res}} - \mu_Y\right)/T + M, \end{aligned} \tag{4.1}$$

where D_X and D_Y are diffusion coefficients that depend on the properties of X and Y and the membrane[1]. The rate of diffusion through the membrane is positive for X and negative for Y. In the steady state $\dot{N}_X = \dot{N}_Y = 0$ and the diffusion term will be balanced by the population metabolic rate M, from which we can obtain the chemical potentials as functions of M: $\mu_X = \mu_X^{\text{res}} - MT/D_X$ and $\mu_Y = \mu_Y^{\text{res}} + MT/D_Y$. As M increases the two potentials move closer together, and if no metabolism takes place then the system will be in equilibrium with its surroundings ($\mu_X = \mu_X^{\text{res}}$, $\mu_Y = \mu_Y^{\text{res}}$).

The total entropy produced per mole of X converted to Y is given by $\Sigma = \lambda_X - \lambda_Y = (\mu_X - \mu_Y)/T$, which as a function of M is

$$\Sigma = (\mu_X^{\text{res}} - \mu_Y^{\text{res}})/T - kM, \qquad \text{where } k = \frac{1}{D_X} + \frac{1}{D_Y}. \tag{4.2}$$

Σ is closely analogous to the inverse temperature difference $(1/T_{\text{C}} - 1/T_{\text{H}})$ between the two heat baths in a classical heat engine or in a two-box atmospheric heat transport model. It represents an upper limit on the amount by which an organism can reduce its structural entropy by metabolising one mole of X into Y. Equivalently, the upper limit on the amount of chemical work that can be performed per mole metabolised is $T\Sigma$. Σ decreases with M (figure 4.2a), so that a slow population metabolic rate will result in a greater ability to do work per mole metabolised. M has a maximum value of $M_{\text{max}} = (\mu_X^{\text{res}} - \mu_Y^{\text{res}})/kT$ at which Σ becomes zero and no work can be done. Converting X to Y faster than this rate would require work to be done rather than being a source of work.

The total entropy production due to metabolism is given by

$$\sigma = M\Sigma = \frac{\mu_X^{\text{res}} - \mu_Y^{\text{res}}}{T} M - kM^2. \tag{4.3}$$

This function is zero at $M = 0$ and $M = M_{\text{max}}$, with a maximum in between at $M_{\text{MEP}} = \frac{1}{2}(\mu_X^{\text{res}} - \mu_Y^{\text{res}})/kT$. (MEP here stands for "maximum entropy production," indicating the value of M that maximises σ.) From a global point of view there is therefore a tradeoff between slow and fast population metabolic rates (figure 4.2b). A slow population metabolism leaves a large difference

[1] Empirically, a much better formula for the rate of diffusion across a membrane is given by Fick's law, which says that the rate of diffusion is proportional to the difference in concentration rather than chemical potential. Making this assumption complicates the formalism of this model quite substantially, in part because it requires us to give explicit formulae for the relationship between concentration and chemical potential in the reservoir and inside the system. However, it leads to qualitatively similar results in which $\mu_X - \mu_Y$ is a decreasing (though nonlinear) function of the population metabolic rate and the entropy production has a single peak at an intermediate value of M.

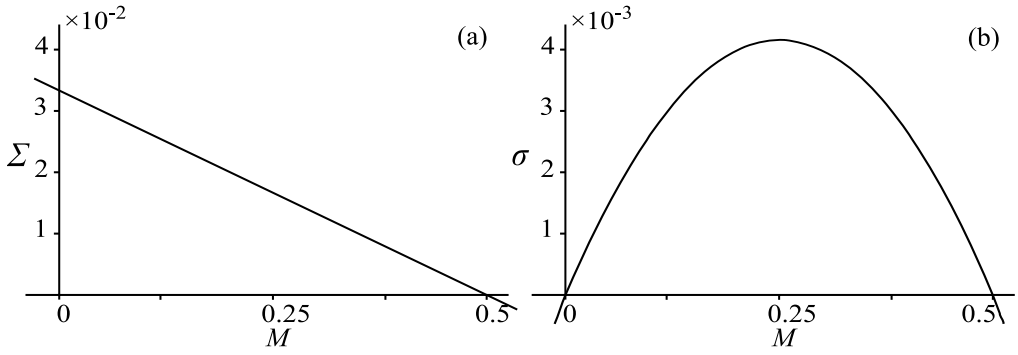


Figure 4.2: (a) $\Sigma = (\mu_X - \mu_Y)/T$ tails off with increasing metabolic rate M. Σ is proportional to the difference in chemical potential between X and Y, which is equal to the maximum amount of work that can be done by converting one mole of X into Y. With $M > M_{\max} = 0.5$, work would have to be done to perform the conversion since $\Sigma < 0$. (b) The total entropy production $\sigma = M\Sigma$ rises to a peak and then falls off with increasing M. The (arbitrary) values used for these plots are $\mu_X^{\text{res}} - \mu_Y^{\text{res}} = 3000$, $D_X = D_Y = 0.1$ and $T = 300$.

in potential between food and waste but produces a sub-optimal amount of work because it occurs slowly, whereas a fast population metabolic rate leaves too small a difference in potential to perform as much work as the optimal rate.

It should be stressed that the curves in Figure 4.2 are a property not of the population of organisms but of their physical environment. In deriving the expressions for Σ and σ as functions of M we did not make any assumptions about what was causing the conversion of X into Y. These relationships are therefore purely properties of the physical set-up in figure 4.1, with the membrane separating the system from a reservoir. In particular, the reason that $\Sigma(M)$ is a linear function is due to the slightly unrealistic kinetics we have assumed for the transport across the membrane. Different choices for the membrane kinetics would lead to a non-linear curve, but it would still be a decreasing function.

However, it is easy to see that the general situation in which Σ is a decreasing function of M, and thus $\sigma(M)$ has a peak, must be common in many natural situations (see H. T. Odum and Pinkerton (1955) for a number of examples). I will use the phrase "negative feedback boundary conditions" to refer to this type of situation. The "negative feedback" part of this phrase refers to the fact that the amount of entropy reduction (or, equivalently, chemical work) that can be achieved per mole of reaction decreases with increasing activity in the system (M), and the term "boundary conditions" specifies that this feedback is a property of the system's physical environment, rather than of the system itself. Negative feedback boundary conditions can be thought of as somewhere between the two types of boundary condition usually considered in non-equilibrium thermodynamics, namely constant gradient (i.e. the chemical potentials μ_X and μ_Y are held constant) and constant flow, which in this case would correspond to adjusting the potentials in such a way that M would attain a particular constant value.

Another, perhaps more realistic, scenario in which negative feedback boundary conditions can occur is in a flow reactor. This is a system of constant volume, into which X is fed a constant rate, which we denote f. The mixture of X and Y inside the system is removed at the same rate,

maintaining the constant volume (we assume that the organisms are separate from the mixture and are not removed by this process). The equivalent of Equations 4.1 for this system is

$$\begin{aligned} \dot{N}_X &= f(1 - N_X/N_{\text{total}}) - M \\ \dot{N}_Y &= M - fN_Y/N_{\text{total}}, \end{aligned} \tag{4.4}$$

where $N_{\text{total}} = N_X + N_Y$. In the steady state, this leads to $N_X = N_{\text{total}}(1 - M/f)$ and $N_Y = N_{\text{total}}M/f$. Assuming an ideal solution, this leads to

$$\Sigma = (\mu_{X0} - \mu_{Y0})/T + \log(f/M - 1). \tag{4.5}$$

As in Equation 4.2, Σ is a decreasing function of M in the steady state. The two expressions also share the feature that Σ becomes negative for values of M greater than some value M_{max} (it can be seen that M_{max} must be less than f, as one might expect). However, in Equation 4.5, $\Sigma(M)$ is no longer a linear function (in particular, Σ becomes infinite as M approaches zero). The function $\sigma(M) = M\Sigma(M)$ still has a single peak but is no longer parabolic in form, and M_{MEP} is no longer equal to $M_{\text{max}}/2$. This non-linearity does not cause any difficulties for what follows. It is important to bear in mind that $\Sigma(M)$ may have a variety of different forms, depending on the exact nature of the system.

The MEPP suggests a hypothesis that, under negative feedback boundary conditions, real ecosystems would tend to have a population metabolic rate close to M_{MEP}. However, the MEPP does not provide an explanation for this in terms of mechanisms that take place within the system, just as MEPP based atmospheric models predict rates of heat transfer without specifying the mechanisms by which the heat transfer takes place. R. D. Lorenz et al. (2001) give an interesting discussion of this in relation to the Martian atmosphere: the atmospheres of Earth and Mars both transport heat at a rate close to the value predicted by MEPP, but on Mars the heat is transported largely in the form of latent heat, due to the freezing and thawing of carbon dioxide, whereas on Earth convection carries sensible heat directly from the equator to the poles. The MEPP makes predictions about the overall rate of heat transport, but its mechanism is different in the two different cases.

In the following section I will develop a population-dynamic model, which allows us to model changes in the Population Metabolic Rate over time, both on the time scale of population dynamics and on evolutionary time scales. However, the mechanisms included in this model do not result in a maximisation of σ on either time scale. There is evolutionary pressure to increase M even if it means a decrease in the population and of the entropy production. If the theory of MEPP does apply to real ecosystems then it must be due to additional mechanisms which are not included in the model developed below. Identifying these mechanisms, if they exist, is a problem for future work.

4.3 Organisms as Engines: an Evolutionary Model

In the previous section we defined a model in which the population metabolic rate was a parameter. In effect we defined only the boundary conditions of the system, with the single parameter M representing the overall effect of all the processes that take place within the system.

In this section we will develop a specific model of the population, which allows us to calculate the value of M and to study how M changes over evolutionary time. We will find that in this very simple model there is always evolutionary pressure towards an increase in M, regardless of whether M is greater or less than M_{MEP}, meaning that in general, in this model there is no dynamical trend towards maximising the entropy production. A steady state in which $M = M_{\mathrm{MEP}}$ is optimal for the population as a whole (it represents not only a maximum in the rate at which chemical work is extracted but also a maximum in the population itself) but evolutionary competition drives the system towards a faster, less optimal, population metabolic rate.

There are many factors that can limit the increase of M in natural systems, including physical constraints on the efficiency of individual metabolisms, altruistic restraint (which could evolve in a variety of ways) predation and spatial effects. We will discuss some of these possibilities in Section 4.4.

4.3.1 A Heat Engine Metaphor

In order to motivate this model it is instructive to consider an analogy in terms of heat engines. Imagine that we have two heat baths **A** and **B**, with temperatures $T_{\mathbf{A}} > T_{\mathbf{B}}$. We could position a (not necessarily reversible) heat engine between the two heat baths, taking heat from **A** at a rate Q_{in} and adding heat to **B** at a rate Q_{out}, producing work at a rate $W = Q_{\mathrm{in}} - Q_{\mathrm{out}}$.

If this heat engine is to be in place for a long time then its parts will suffer wear and tear and need replacing. Let us assume that the appropriate raw materials needed to achieve this are readily available. Work will need to be done at a certain rate, W_{repair} (constant over a long enough time scale) to transform these raw materials into replacement parts for the heat engine. But if $W \geq W_{\mathrm{repair}}$ then we can simply use some of the engine's output to perform this repair process. The worn-out parts are returned to the pool of raw materials. The actual value of W_{repair} depends on many factors, including the nature of the heat engine and of the available raw materials, and on the way in which the repairs are carried out.

Heat is produced as an effect of producing the spare parts. We assume this heat is returned to the cold reservoir **B**. If the total energy content of the raw material pool is not changing over time then by the conservation of energy this heat must be produced at a rate $Q_{\mathrm{waste}} = W_{\mathrm{repair}}$.

This leaves a rate $W_{\mathrm{excess}} = W - W_{\mathrm{repair}}$ of work to be put to other uses. One use that this work could be put to is building another heat engine, similar to the original. This is analogous to biological reproduction, while the use of work to repair the engine is similar to biological metabolism[2] (it could be thought of as a literal interpretation of Kauffman's (2000) idea of a

[2] An important difference is that in this metaphor I have described the repair and reproduction as being done by us, whereas in a biological situation it is achieved by the dynamics of the engines/organisms' structures. This doesn't make any difference for the energetics of the present model, but it is important for a proper understanding of biological

"work-constraint cycle").

We may now consider a population of n such engines. If the temperatures of **A** and **B** are fixed then each engine can produce work at the same rate, regardless of the size of n, and the population can grow without limit, unless the available space or the size of the raw material pool becomes a factor.

However, we can consider the heat baths' temperatures to be functions of the rate at which heat is added or removed from them, in a similar manner to the equatorial and polar regions in the two-box atmosphere model, or to the chemical potentials in the ecosystem model above. In line with both these examples, we assume that the difference between the two temperatures narrows as the overall rate of heat flow from one bath to the other increases.

If $W > W_{\text{repair}}$ then $W_{\text{excess}} > 0$ and it is possible for new engines to be built; if $W_{\text{excess}} < 0$ then then not enough work is being produced to maintain all the engines and some will have to be decommissioned. But W depends on the difference between the two temperatures. If we assume that each engine removes heat from **A** at the same rate, regardless of the temperature difference, then a greater population means a greater total heat flow between the two heat baths, and hence a lower difference in temperatures. The second law implies a maximum possible value for W of $Q_{\text{in}}(1 - T_{\mathbf{B}}/T_{\mathbf{A}})$, which decreases as the population increases and the values of $T_{\mathbf{A}}$ and $T_{\mathbf{B}}$ become closer. A balance is therefore reached, where $W_{\text{excess}} = 0$ and the population neither grows nor shrinks.

By reasoning along these lines we can create a model of the population dynamics of these engines. We can then go on to consider the effects of evolution, in the sense that a new, slightly more efficient or faster-running type of engine might be introduced to compete with the older model. Before we do this I will complete the metaphor by translating the argument into the domain of chemical rather than heat engines.

4.3.2 Organisms as Chemical Engines

For the most part the translation is straightforward. As explained in detail in Chapter 2, the quantity $\lambda_U = \partial S/\partial U = 1/T$ has a close formal relationship to $\lambda_{N_X} = \partial S/\partial N_X = \mu/T$. Rather than thinking about engines that transport heat from **A** to **B** we can switch to thinking about engines that convert a reactant or set of reactants X into products Y, and most of the formal details will stay the same.

The one thing that needs a little consideration when making this switch is the concept of work. In the heat engine analogy we were using the classical thermodynamic idea of work as energy concentrated into a single macroscopic degree of freedom, so that it doesn't contribute to the calculation of the entropy. Effectively, work in this sense can be thought of as energy with an infinite temperature. It is not clear whether work in this form plays any role in a biological systems (though there are certainly energy quantities in biology that can be thought of as having extremely high temperatures; see Jaynes (1989) for an example). It is therefore better to use the concept of *entropy reduction* in our model rather than work.

Like heat engines, organisms have a low entropy structure. Entropy is produced within this

phenomena. This concept will be more fully explored in Chapter 5.

structure at some rate as a result of various processes including the decay of the structure, and this production of entropy must be countered in order to maintain it. This can only be achieved by increasing the entropy of the environment at a greater or equal rate. Thus, instead of thinking about how much work can be produced by moving heat from a high to a low temperature region, we can think about how much the organism's entropy can be reduced by converting matter in its environment from a high chemical potential to a low chemical potential form.

One could, if one wished, divide this entropy reduction by a temperature to put it into energy units, in the same way that the entropy of a system plus that of its surroundings is usually expressed as a free energy. The quantity thus formed might reasonably be referred to as "chemical work," although it doesn't behave in exactly the same way as the mechanical concept of work (see Section 2.6.2). However, this chemical work can only be defined if the temperature is well-defined and constant over space, which is not necessarily the case inside a living cell (Jaynes, 1989) or in an ecosystem at large. It is more general to keep the quantity in entropy rather than energy units, and in my opinion it is more informative to think of it in this way.

In the previous section we defined the molar entropy production Σ as the increase in the entropy of the organisms' chemical environment per mole of reactants X converted to products Y. Σ is also the upper bound on the amount by which the organisms' entropy can be reduced per mole of reaction. We are assuming here that the organisms' metabolism proceeds by coupling the $X \rightarrow Y$ reaction to another reaction $Z \rightarrow \mathit{biomass}$, where Z is an abundant material resource whose concentration does not significantly change and which is returned as the organisms' structure decays (i.e. the decay process can be summarised as $\mathit{biomass} \rightarrow Z$). In other words, we assume that the organisms' structure is made of matter other than that which makes up X and Y. This assumption will simplify our calculations but should have little effect on the end result.

If the population is at a steady state value then the total rate of entropy reduction occurring in all the organisms must be balanced by the total rate of entropy produced within their structures. Thus the overall entropy production is still equal to $M\Sigma$ as calculated in the previous section. This is equivalent to the assumption in the heat engine metaphor that all the work performed by the engines eventually degrades to heat and flows into the cold reservoir.

4.3.3 The Population Dynamics of Engines

All the conceptual apparatus is now in place to make a population dynamic model where the organisms are modelled as chemical engines. We will have to make a few additional assumptions in order to do so.

Let us suppose that each individual organism's structure's entropy must be reduced at a constant rate s, which is independent of μ_X and μ_Y. Any additional entropy reduction that the organism's metabolism can achieve will be put into reproduction.

Let us also assume that each organism converts X into Y at a constant rate m, which is also independent of μ_X and μ_Y.[3] This results in a rate of entropy production in the environment of

[3]There is a slight element of unrealism in this assumption, because it means that even if the difference between μ_X and μ_Y is arbitrarily small the organisms are still capable of converting X to Y at the full rate, even though the conversion process is powered by the difference in chemical potential. This is not thermodynamically impossible —

$m\Sigma = m(\mu_X - \mu_Y)/T$ for each organism. This is also the maximum possible value for the entropy reduction each individual's metabolism can achieve. We will assume that the actual rate of entropy reduction achieved is given by $\alpha m(\mu_X - \mu_Y)/T$, where $\alpha < 1$ is a constant proportion representing the irreversibility of the organisms' metabolism. If there is turnover in the population then the losses involved in death and the reproduction required to offset it can be incorporated into α, so we do not need to model this explicitly.

We can see that if $\alpha m(\mu_X - \mu_Y)/T > s$ then there is excess entropy reduction available, which can be put to use for the creation of new individuals, whereas if $\alpha m(\mu_X - \mu_Y)/T < s$ then $\mu_X - \mu_Y$ is too small to support the organisms' metabolism indefinitely and they will die at some rate.

The population metabolic rate M is given by nm. By the assumptions we have made in analogy to two-box atmospheric models, $\Sigma = (\mu_X - \mu_Y)/T$ is a decreasing function of M. The population growth rate has the same sign as $\alpha m\Sigma(nm) - s$. Since Σ is a decreasing function an initially small population will grow or an initially large population shrink until a stable steady state is reached where $\Sigma = s/\alpha m$.

Note that the only assumption we have made about the organisms' growth rate is that it is positive when there is enough energy available for it to be so. This model is very general in that it does not depend on any other specific features of the growth dynamics. One could add additional assumptions in order to specify the dynamics of n, but we will not do this because it is unnecessary, and the steady state condition is the value we are most interested in.

The result that $\Sigma = s/\alpha m$ in the steady state is quite general in that it holds for any type of boundary conditions where Σ is a decreasing function of M. In the particular case of our membrane-bound ecosystem model, $\Sigma(nm) = (\mu_X^{\text{res}} - \mu_Y^{\text{res}})/T - knm$. Setting this equal to $s/\alpha m$ gives

$$n = \frac{1}{k}\left(\frac{\mu_X^{\text{res}} - \mu_Y^{\text{res}}}{Tm} - \frac{s}{\alpha m^2}\right) \tag{4.6}$$

as the steady state value of n with these particular boundary conditions.

Note that this grows with increasing α and decreasing s, meaning that the more efficient the organisms' metabolism, and the less entropy reduction required to maintain their structure, the larger the population that can be maintained. If s/α is too large then $n < 0$, indicating that no population of these organisms can be supported by this environment.

The dependence of the population n on the individuals' metabolic rate m is slightly more complex. $n = 0$ when $m = sT/\alpha(\mu_X^{\text{res}} - \mu_Y^{\text{res}})$, indicating a minimum value of m for which a population can be sustained. n increases with m up to a maximum when $m = 2sT/\alpha(\mu_X^{\text{res}} - \mu_Y^{\text{res}})$, and then drops asymptotically to 0 as $m \to \infty$.

The total entropy production rate at the steady state in this example is given by

$$\sigma = \frac{1}{k}\left(\frac{\mu_X^{\text{res}} - \mu_Y^{\text{res}}}{T}\frac{s}{\alpha m} - \frac{s^2}{\alpha^2 m^2}\right). \tag{4.7}$$

in the heat engine analogy it corresponds to heat being pumped almost-reversibly by almost-frictionless flywheels that require very little temperature difference to maintain their angular momentum — but it is biologically implausible. A more realistic choice would be to let m be proportional to $\mu_X - \mu_Y$. This turns out not to affect the conclusions of this section, so I have presented the simpler assumption.

We can consider this a function of s, m or α. In each of these cases it exhibits a peak, with the entropy production maximised for a particular value of the parameter (though in the case of α this peak might occur at a value of α greater than one, which cannot be achieved physically).

We could therefore use the maximum entropy production principle to predict values for these parameters. However, we can also consider the effect of evolution by natural selection upon these parameters. One might hope that by doing this we could show that the values predicted by MEPP are the same as the values that evolution would select for, but this turns out not to be the case, as shown in the next section.

4.3.4 Evolutionary Dynamics in the Chemical Engine Model

We will now consider the dynamics of evolution in this model. We will find that under all circumstances, a mutant population with greater m, greater α or lower s will be able to establish itself and out-compete the established population. In the case where m is allowed to evolve without constraint, this leads to an ever-decreasing population of ever faster-metabolising organisms.

First we must consider what happens if there are multiple populations occupying the same ecological environment. We assume that each of these populations consists of organisms with a metabolism that is powered by converting X into Y, but we will allow each population to have different values for s, α and m.

Let n_i be the number of individuals of the i^{th} type, whose metabolic parameters are given by s_i, α_i and m_i. The Population Metabolic Rate $M = m_1 n_1 + m_2 n_2 + \dots$ now includes contributions from each type of organism. In our membrane-bound environment, $\Sigma(M)$ is now given by

$$\Sigma = (\mu_X^{\text{res}} - \mu_Y^{\text{res}})/T - k(m_1 n_1 + m_2 n_2 + \dots). \tag{4.8}$$

The i^{th} sub-population can achieve excess entropy reduction, and hence grow in numbers, only if $\Sigma > s_i/\alpha_i m_i$. But Σ is now determined by the total metabolic rate of all the sub-populations. The sub-populations compete with each other indirectly, since they rely on the same resource.

Coexistence between two populations i and j is only possible if $s_i/\alpha_i m_i = s_j/\alpha_j m_j$. If $s_i/\alpha_i m_i < s_j/\alpha_j m_j$ then there are three possibilities: either $\Sigma < s_i/\alpha_i m_i$, in which case both populations shrink (and thus Σ grows), or $\Sigma > s_j/\alpha_j m_j$ in which case both populations grow and Σ shrinks, or it is somewhere in between. In this latter case, population i grows while j shrinks. This process must continue until population j becomes zero and $\Sigma = s_i/\alpha_i m_i$. The only other steady state in the system is where population i is zero and $\Sigma = s_j/\alpha_j m_j$, but this is locally asymptotically unstable because any positive value for population i will result in its growth.

Note that again this reasoning is very general in that it does not depend on any assumptions about the species' growth rates beyond their sign. One could therefore say that the impossibility of coexistence in this model is a consequence of thermodynamic constraints rather than specific assumptions about population dynamics. In order to allow coexistence in this model, new physical processes would have to be added.

As a consequence of the impossibility of coexistence, if we assume that mutations occur only rarely then we can expect to find only one type of organism occupying the system at any given

time. If a mutation occurs, introducing a small new sub-population j to the established population i then there are three possibilities: (1) $s_j/\alpha_j m_j < s_i/\alpha_i m_i$, in which case the mutants cannot reduce their structures' entropy fast enough to support themselves, so their population must die out; (2) $s_j/\alpha_j m_j = s_i/\alpha_i m_i$ which is a neutral mutation. If this occurs we can expect the mutants to take over the population stochastically some of the time; or (3) $s_j/\alpha_j m_j > s_i/\alpha_i m_i$, in which case the mutant population's growth rate is positive, and as the mutant population grows it reduces Σ, so it will eventually out-compete the original population and become the new established species.

Therefore evolutionary pressure in this model always favours an increase in $s/\alpha m$. If there are no trade-offs between these characteristics of an organism then we can expect that the metabolic efficiency α will evolve towards its maximum possible value (its theoretical maximum is 1 but constraints on the possible architecture of living cells will certainly mean its real maximum is somewhat lower), and metabolic rates m will also reach the maximum value permitted by the organisms' architecture. We can also expect the required entropy reduction s to evolve to as low a value as possible. This corresponds to minimising the rate of entropy production within the organisms' structures. This is another form of metabolic efficiency that involves the reduction of unnecessary decay processes.

Note that changes in the individuals' body size have no effect unless they change the value $s/\alpha m$. In general a change in body size could make this quantity larger or smaller. A smaller body size gives a larger surface area to volume ratio, which might increase m relative to s by allowing a faster flow across an external membrane. On the other hand a larger body size might decrease s relative to m, analogously to the way a mammal's large body size makes it easier to maintain a constant body temperature without losing heat. This model therefore does not predict any particular direction to evolutionary changes in body size and one would expect it to be determined by a trade-off between these two types of effect. A similar argument can be made for the organisms' complexity, which we will come back to later.

The evolution of the individuals' metabolic rate m is interesting and deserves a little examination. Assume for the moment that s and α are fixed and consider a population with an initially very low value for m. This value increases over evolutionary time until it reaches a limit determined by the organisms' architecture. However, initially this increases both the population n and the entropy production σ, which is also equal to the total rate of entropy production occurring in the cells.

Both n and σ reach a maximum when $m = 2sT/\alpha(\mu_X^{\mathrm{res}} - \mu_Y^{\mathrm{res}})$. However, there is still an evolutionary pressure to increase m when it is greater than this value: mutants with an increased metabolic rate can still invade and take over the system, but this results in a reduction of the overall population and of the entropy production. This is an example of a well-known phenomenon in evolutionary biology where "selfish" mutation can become established even though its effects are deleterious to the population as a whole. The population metabolic rate M grows without limit as m increases, and the molar entropy production Σ decreases indefinitely toward 0. The end result is an ever-decreasing population of ever-faster metabolising organisms, living in an environment that is able to supply less and less capacity for entropy reduction.

4.4 Possible Constraints on M

In the previous sections I developed a simple evolutionary model which, with a minimum of assumptions, showed that evolutionary pressure is towards an increase in population metabolic rate M, rather than towards an increase in the entropy production $\sigma = M\Sigma$. However, if M were to increase without limit the resulting situation would be a very small population of extremely rapidly metabolising individuals. Additionally, the model does not permit the coexistence of more than one species. Neither of these phenomena correspond what we observe in nature, and it is therefore worthwhile to discuss the various constraints and additional processes which could occur to prevent this from happening.

Another reason for discussing constraints on M is that we might hope to find one that would cause the value of M to stop increasing at M_{MEP}, where the entropy production curve is at its peak. Such a result would provide a tangible explanation for the ecological maximisation principles of H. T. Odum and Pinkerton (1955) and later authors, lending such principles a far greater degree of credibility than they can currently attain.

The ideas in this section should be seen as suggestions for future extensions of the model, and further work will be required to verify their consequences.

4.4.1 Physical constraints on metabolism

By far the most obvious type of constraint on M is formed by physical limits on the values of s, m and α. These limits (maxima for m and α, and a minimum for s) would be determined by the basic architecture of the organisms. For instance, for single-celled organisms, m is constrained by the rate at which nutrients can flow across the cell membrane and by the diffusion rate of nutrients in the environment. Evolutionary innovations such as mobility (allowing the organism to move to areas of higher food concentration) or multicellularity may be able to increase this to some extent, but one would expect there to be values of m which cannot be reached through evolution. Thus one would expect $\alpha m/s$ to increase over evolutionary time until it reaches a value of $\alpha_{\max} m_{\max}/s_{\min}$.

Constraints on metabolism need not take the form of maximum values for α, m and $-s$. In general, very efficient machines tend to operate slowly, and this probably applies to metabolisms as well. So there might be a trade-off between α and m such that organisms with high α must have low m and vice versa. In this case one would still expect $\alpha m/s$ to increase over evolutionary time until it reaches a maximum value.

Putting physical constraints on metabolism into the model prevents the unrealistic indefinite growth of M, but it does not necessarily cause it to converge to a value anywhere near M_{MEP}. The value attained by M is given by solving $\Sigma(M) = (s/\alpha m)_{\max}$, whereas the MEPP predicts $\Sigma(M) = M\frac{\mathrm{d}\Sigma}{\mathrm{d}M}$. There is no reason to expect that the maximum value of $s/\alpha m$ would be equal to $M\frac{\mathrm{d}\Sigma}{\mathrm{d}M}$, since the latter is a property of the population's physical environment whereas the former is a property of an individual's metabolism.

4.4.2 Limited Nutrients or Space

In this model we assumed that the organisms' structures were constructed of a material Z which is available in high abundance. However it would be a relatively simple matter to change this so that the availability of this resource were limited. Thus, in order to reproduce an individual would not only need to be able to extract an excess of free energy from the environment, but would also have to be able to extract a sufficient quantity of the material resource Z. This would become more difficult with a larger population as the concentration of Z would be lower, lowering its chemical potential. There would thus be an additional factor limiting the population size, in addition to the availability of X.

One important effect this might have is that an increase in metabolic efficiency would not necessarily increase the population metabolic rate. Without a limiting resource, a faster or more efficient metabolism always leads to a higher population, but if metabolic efficiency is not the factor limiting the population then it may be possible for the population to remain constant while the organisms' metabolic efficiency decreases.

Another effect of the addition of a limiting resource is that it might enable the coexistence of more than one species. When the only competition is for food, the only viable strategy for an organism is to metabolise food and reproduce as rapidly as possible. When there is competition for two independent resources (X and Z) it might be possible for two populations to coexist by adopting different strategies.

Competition for space would have a broadly similar effect, with space playing the role of a resource that becomes less available as the population increases. A simple way to model this is to say that the environment is big enough to support no more than $n_{\max}$ individuals. In this case we would expect the attained value of M to depend on $n_{\max}$ as well as upon any physical constraints on the organisms' metabolism.

Modelling limited space or nutrients would require substantial extra assumptions to be added to the model. However, in both cases it seems unlikely that the final value of M would match the MEPP prediction, because it seems that its value would depend upon a number of specific parameters, rather than on $\mathrm{d}\Sigma/\mathrm{d}M$.

It may be worthwhile to add multiple limited nutrients, as well as multiple sources of chemical work to the model. Real ecosystems contain species with many different types of metabolism. Real metabolisms are also more complex, performing multiple reactions at rates that can vary depending on circumstances. More complex reactions result in more complex biochemical feedbacks, including nutrient cycling. Adding such features to the model would allow it be used as a tool to study the interaction between metabolism and environment at a very fundamental physical level.

4.4.3 Predation and Parasites

Another possible addition to the model that would reduce the attained value of M is the addition of a predator species which feeds upon the organisms, keeping their population down. An interesting property of this addition is that we can see it as the removal, rather than the addition, of a con-

straint: the model described in this chapter concerns an ecosystem that is constrained to have only one type of organism, with only one type of metabolism. Adding the possibility of a predatory organism could be seen as the removal of that constraint.

However, as with the addition of limited resources, the addition of predation requires the addition of numerous extra assumptions in the model. There seems little reason to think that any of these assumptions would result in a maximisation of the entropy production, for the same reason as with resource limitation: there is no process whose rate depends upon $\mathrm{d}\Sigma/\mathrm{d}M$ rather than on Σ or the population n. In general the presence of a predator would decrease the population of its prey, resulting in a decrease in M, even if $M < M_{\mathrm{MEP}}$.

Another possibility is that, rather than adding a population of predators which feed on the prey population as a whole, one could add parasites, where each parasite feeds only on a single prey, effectively stealing some of the work extracted by the prey so that it goes into the maintenance and reproduction of the parasite rather than of its prey. One reason this possibility is interesting is that, if each parasite individually steals work at the optimum rate (as fast as possible, but not so fast that its host is no longer able to maintain its own structure and dies) then it might lead to the rate of work extraction by parasites being maximised overall, which would correspond to a maximum in entropy production. Constructing such a model and testing this idea is a task for future work. If a model along these lines did result in a maximisation of σ then on the one hand it would be a case of "optimisation in, optimisation out" — the maximisation of σ would be a result of the assumption of optimal harvesting by parasites — but on the other hand the idea that parasites would evolve to maximise the free energy available for their reproduction seems fairly reasonable.

4.4.4 Altruistic Restraint via Kin Selection

In the model as presented, the individuals' metabolic rate m is always under evolutionary pressure, because an increase in m always benefits the individual even though it is deleterious to every other organism. Thus, if an organism were to restrain its metabolic rate this would benefit the other individuals at a cost to itself, and could thus be seen as a form of altruism.

There are many ways in which altruism is thought to be able to evolve, the most well-known being kin selection (Hamilton, 1964), whereby altruistic acts performed towards an organism's close relatives are selected for because it increases the frequency of the genes they share. In order for this to occur it must be possible for an organism's behaviour to have an effect on its relatives separately from less closely related individuals.

In the model as presented this is not possible, because the chemical environment is assumed to be thoroughly mixed, and hence an individual's metabolism affects all organisms equally. However, it would be possible and not unrealistic for the mixing assumption to be relaxed. Kin selection could then take place via a "viscous population" mechanism, where the rate of an individual's metabolism has a greater effect on nearby individuals, which are more likely to be related.

As for many of the other effects in this section, however, there is no reason to believe that this factor alone would result in M becoming equal to M_{MEP}, since kin selection would provide evolutionary pressure to reduce M regardless of whether $M > M_{\mathrm{MEP}}$.

4.4.5 Ecosystem-Level Selection

Another mechanism by which altruism can evolve is selection at the level of large groups of individuals. The notion of group selection was widely criticised during the latter half of the 20$^{\text{th}}$ century but is currently somewhat resurgent (Penn, 2005).

This idea requires us to imagine a scenario in which there are many separate environments, with negative feedback boundary conditions applied to each one individually, some of which are unpopulated. Periodically, some process extracts individuals or collections of individuals from populated environments, which are selected according to some criteria, and puts them into empty ones. This "reproduction" of ecosystems is balanced by some other process which empties populated environments. Depending on a number of factors, including the size of the populations and the relative scales of within-group reproduction versus the group reproduction process, this can lead to the selection of organisms with traits that increase the group's probability of being selected.

One plausible criterion for the selection process is that individuals are more likely to be taken from groups with higher populations. This could occur if, for instance, the selection process takes only a small amount of matter from a randomly chosen group; this volume would be more likely to contain an organism if the population were higher. Unlike many of the other constraints on M discussed in this section, this ecosystem selection process could act to decrease M only when $M < M_{\text{MEP}}$ and increase it otherwise. This is because the population is at its maximum when $M = M_{\text{MEP}}$. This scenario may seem somewhat contrived, but it is not implausible that situations analogous to it could apply to ecosystems on the microbial scale.

4.4.6 Summary

I have presented several possible mechanisms by which the increase in M presented by the simple model could be constrained. The addition of some of these additional factors, such as predation and parasitism, can be seen as the relaxation of constraints on σ, since they allow additional processes to occur that were previously prevented.

All but two of these additional factors (parasitis m and group selection) seem unlikely to lead to a value of M that would maximise the entropy production. Parasitism in particular is ubiquitous in natural ecosystems, but further work is needed to model its effect.

One further possibility is that maximisation of entropy production could arise from the complex interaction of many such effects. This seems to be to some extent implied by the evidence from climate science, in which MEPP works as a predictive principle only for highly complex, turbulent atmospheric systems, and by the statistical argument in the preceding chapter, whereby a system must be unconstrained in order for MEPP to be applicable, implying that it must have many macroscopic degrees of freedom (R. D. Lorenz et al., 2001). However, this idea is highly speculative and it is difficult to say any more about it until ecosystem models of the required complexity can be constructed.

4.5 Discussion

4.5.1 An Important Implication: Bootstrapping Complexity

A potentially important implication of this model is what I call the *bootstrapping* of organisms' metabolism over evolutionary time. This theme will recur in Chapter 5. The idea is that the negative feedback in this model provides a mechanism by which Σ reduces over evolutionary time, in step with the emergence of evolutionary innovations that allow metabolisms to become faster and more efficient, enabling organisms to persist in environments with lower values of Σ. Applying this idea to the Earth as a whole, it could easily be the case that the chemical environments of the early Earth had a much higher Σ than the majority of chemical environments that exist on Earth today, for the simple reason that suitable environments on the present-day Earth with high Σ tend to be rapidly colonised by modern organisms, whose metabolism reduces Σ to a low level.

Entropy reduction in chemistry is often thought of in terms of quantities measured in energy rather than entropy units. In these terms, $\Sigma/T = \mu_X - \mu_Y$ is the amount of energy available to perform work per mole in the chemical environment. This bootstrapping argument is then a claim that the available chemical free energy on the early Earth could have been much higher than it is today.

There are many hypotheses about the source of energy that powered early life (to be discussed briefly in the next chapter). One possibility is a scenario for the early Earth which is similar to the situation that might exist on Titan today (e.g. Schulze-Makuch & Grinspoon, 2005), whereby free energy is produced via photochemistry in the atmosphere. The products of photochemical reactions would have rained onto the surface and oceans of the early Earth, increasing in concentration over long periods of time. If the populations of early organisms or proto-organisms were low and their metabolic rates low then the steady-state concentrations of their photochemically produced food molecules would be high. As faster and more efficient metabolisms evolved and populations increased, this concentration would have dropped, which suggests that its value today should be much lower than during the time of life's origins.

This may be too crude a sketch to apply to the Earth as a whole, but might apply instead to specific locations where stable food molecules tend to build up. One can easily imagine a lake or shallow sea with a much higher molar free energy than would be found in a sterile location on Earth today, and it is in these locations that we might most expect to find the kind of complex autocatalytic networks of chemicals and spatial structures from which early life might have arisen.

It is worth mentioning the "great oxidation" event in this context. The oxygenation of the Earth's atmosphere might seem a counterexample to the idea that successful metabolic innovations tend to decrease the free-energy density of the environment. The great oxidation event greatly increased the availability of chemical energy to non-photosynthesising organisms, making possible the great proliferation of animal life as we know it today (though multicellular anaerobic animal life has recently been discovered, see Danovaro et al., 2010). However, it also decreased the efficiency with which chemical work can be performed by oxygenic photosynthesis, by increasing the chemical potential of oxygen. Thus the oxygenation of the atmosphere can be seen as causing a drop in Σ as far as plants and algae are concerned, but it had the side-effect of introducing a new

chemical gradient that other types of organism are able to feed off.

It seems reasonable to think that the types of structure which can persist and reproduce in environments with low Σ would generally be those that are in some sense "harder" to produce, requiring more evolutionary time to appear. A negative feedback of the type found in our model would provide a ratchet or bootstrapping effect, whereby the environment is always able to support only the most efficient and fastest-type of organism that exists in a given niche.

In today's world, living cells are highly non-trivial structures. Consequently their origin seems something of a mystery, and it can therefore seem that a specific sequence of unlikely events must have occurred in order for it to be brought about. But this bootstrapping effect suggests an alternative explanation: the early chemical environments on Earth might have been able to support much simpler types of metabolism, which could have spontaneously formed much more readily. Moreover, due to the lack of competition these early metabolisms need not have operated at anywhere near the rate or the efficiency that modern organisms can achieve. The bootstrapping effect would then have operated over early evolutionary time to weed out all but the most efficient and fastest-metabolising of these forms, eventually resulting in the highly complex and specific mechanisms present in extant cells.

It is worth being careful about what we mean by "simple" and "complex" in this context. It is unlikely that early life would have been simpler than modern life in terms of the basic molecular constituents (e.g. monomers) of which it is composed: experiments such as that of Miller (1953) and theoretical work by authors such as Morowitz (1968) or Kauffman (2000) have repeatedly shown that complex molecules and biological monomers form readily in chemical systems held out of thermal equilibrium. Modern life is composed of a relatively small number of basic molecular constituents, but this might not necessarily have been the case for early forms of metabolism: the bootstrapping effect could have acted to pare down an initially much more chemically diverse metabolic network until only the specific very efficient type of molecular architecture we find today was left. This idea is elucidated by Cairns-Smith (1985), who refers to it as "scaffolding."

Neither must early life's spatial structure necessarily have been simpler than that of modern cells, since spatially complex structures also form readily in non-equilibrium conditions. Examples such as hurricanes are abundant in the Earth's atmosphere, and there are known examples in chemistry, including the reaction-diffusion patterns that are studied in the next chapter.

However, it does seem reasonable to think that modern organisms' structures and molecular components are much more *specific* than those of early proto-organisms. That is to say, they are probably much further from the types of structure that can form spontaneously without being produced by an equally specific molecular machine. Thus the negative feedback process can more strictly be seen as bootstrapping the *specificity* of the organisms' physical and chemical structure.

4.5.2 Experimental Testing

A model along these lines could be used to test the application of the MEPP to ecosystems experimentally. In our model food and waste enter and exit the system via diffusion through a membrane but a similar calculation can be performed for a chemostat-like bioreactor in which a constant in-

flow of food is balanced by a constant outflow of the system's contents (a mixture of food and waste). This leads to a nonlinear decline in Σ with M but the analysis is qualitatively the same and one can find a value M_{MEP} for which σ is maximised. It should therefore be possible to perform a bioreactor experiment in which a measured value of M, which can be calculated from the amount of unused food in the system's outflow, is compared against the value predicted using the MEP principle.

In order for the MEPP to apply the organisms' growth must be constrained by the rate of food availability and/or waste disposal, and not significantly constrained by other factors. I suspect that this is not normally the case in a bioreactor since the aim is usually to produce as high a growth or reaction rate as possible so high concentrations of food are used, leading to a population metabolic rate that is constrained only by physiological factors. In order to test the applicability of MEP to biological populations it will probably be necessary to perform a specialised bioreactor experiment in which the nutrient inflow is very dilute and the system run until the population reaches a steady state. It may also be important to use species for which the bioreactor is a close approximation to their natural environment because an environment to which the organisms are not well adapted could induce additional constraints on the system.

4.5.3 Economic Implications

Although this thesis is primarily concerned with the application of thermodynamic methods to ecosystems and the origins of life, it seems worth commenting that the reasoning in this chapter could be applied to some economic situations as well. In particular, the result that increases in efficiency can lead to greater overall resource use, even in the case where this is detrimental to the population as a whole, seems applicable to increases in technological as well as metabolic efficiency. This has obvious implications for the popular idea that the invention of more efficient technology can solve problems such as climate change that result from excessive resource consumption.

This effect is doubtless already well known to economists. It seems worth mentioning here because it makes for an interesting analogy between economic and ecological systems. Both are non-equilibrium physical systems at root, both consisting of many interacting individuals, all ultimately relying on common resources. As a result it seems that some of their dynamics can be described in broadly similar terms, suggesting a generality to the arguments in this chapter that goes beyond the specific situation of a biological ecosystem.

As an example, imagine that, in an economic society similar to our own, a more efficient car engine is invented. If the rate at which cars are used remains the same this will result in a decrease in the overall rate of fuel usage. But this will not necessarily be the case: it could be that the resulting lower cost per mile of travel would encourage greater use of cars, which could increase the overall use of fuel in a manner analogous to the evolutionary tendency to decrease α described in Section 4.3.4.

Of course, in economics as in ecology there are many factors which could limit this effect: as an extreme example, nobody can drive a car for more than 24 hours per day regardless of how

cheap the fuel is. One can therefore expect a saturation point to be reached in this example, beyond which an increase in engine efficiency would reduce overall fuel consumption after all, because fuel cost is not the only factor that can limit transport use. (Although, conceivably, the reduced cost of transport could support a greater human population through an increase in the ability to distribute other resources such as food; in this case the saturation point would occur at a higher level of overall resource increase.)

Nevertheless it seems as if this effect — of an increase in efficiency leading to an increased uptake of a technology, and hence greater resource use overall — might be common in human economic development. Indeed, the industrial revolution can be seen as being the result of, or at least closely connected to, the invention of a series of more efficient technologies, and its result was to greatly increase the global use of resources by human activity.

4.6 Conclusion

I have presented a simple model that illustrates the relationship between the rate of chemical processes in an ecosystem and its rate of entropy production. This shows that the applicability of a principle of Maximum Entropy Production to living systems could be tested and opens up a range of possibilities for new research directions.

I have also developed a way to explicitly model the evolution of species' metabolisms, with thermodynamic constraints built in. My hope is that such models will form a bridge between the physics-based methodology of systems ecology and the more dynamics-oriented modelling methods of population ecology. Ultimately the two methods must be combined for a full understanding of natural ecosystems.

I have developed perhaps the simplest such model possible. In this scenario, evolution tends to increase the overall rate of resource flow (population metabolic rate), regardless of whether this results in an increase or a decrease in the rate of entropy production. The resolution of this apparent conflict should be a clear priority for future researchers who wish to apply the principle of maximum entropy production to ecological systems.

I have suggested a number of mechanisms that could serve to limit the rate of resource flow. It remains to be seen whether any of them, alone or in combination, could have the property of limiting the population metabolic rate to the value that maximises entropy production.

There are many other possible ways in which this model could be extended, to better represent the complexities of biological evolution and its interaction with its abiotic environment. The model should generalise readily to multiple species with more complex metabolisms, giving rise to important features such as nutrient cycling. In this respect, future developments of this model could resemble the abstract models of Downing and Zvirinsky (1999) and Williams and Lenton (2007), with the added feature of increased thermodynamic realism; alternatively, the parameters of the model could be determined by the thermodynamic properties of real environments, allowing the model to be used as a tool to investigate real systems. In any case, thermodynamically realistic population models along the lines of this one should represent a fertile avenue of future research.

Chapter 5

A Model of Biological Individuation and its Origins

In the previous chapter I presented a population-dynamic model of thermodynamically self-maintaining and reproducing individuals. However, in that model such individuals' existence was assumed *a priori*: the model did not include the mechanisms by which the individuals were formed. This chapter is concerned with a lower-level description: we will be interested in precisely how such individuals can arise and persist in physical systems under non-equilibrium conditions.

The central concept of this chapter is the phenomenon of *individuation*, the splitting up of living matter into individual cells, organisms and colonies. When an external energy gradient is applied, some non-living physical systems exhibit an analogous phenomenon. I argue that studying the process of individuation in dissipative structures can give us useful insights into the nature of living systems. I present simulation results suggesting that feedbacks analogous to those found in ecosystems can help form and maintain the conditions under which individuals occur.

These simulations use a type of chemical model known as a reaction-diffusion system, which can form patterns in which there are blurred but spatially distinct "spots" of an autocatalytic substance, separated by regions in which no autocatalyst is present. These spots maintain their individuality through a balance between ongoing processes. In this respect they are similar to living organisms, especially when seen from the perspective of the theory of autopoiesis.

Living organisms exist within ecosystems, in which the supply of energy and nutrients is restricted, limiting their populations. I show that applying the same conditions to reaction-diffusion systems makes the formation of individuated patterns more likely: the system's parameters become "tuned" to the values where individuation occurs.

I demonstrate that, in reaction-diffusion systems, this phenomenon can produce individuals with a more complex structure than a single spot, and I argue that it is likely to occur in non-equilibrium physical systems of many kinds rather than just reaction-diffusion systems. Based on this result, I present a hypothesis about the role that this kind of ecosystem-level negative feedback might have played in the origin of biological individuation. In addition, I show that simple dissipative structures can exhibit a very limited form of heredity, adding weight to the idea that metabolisms arose before genetic material in the origins of life.

5.1 Introduction

In a previous publication (McGregor & Virgo, 2009), McGregor and I argued that, besides "life as it could be" (Langton, 1989), Artificial Life researchers should also consider "near-life as it is": phenomena in the physical world that exhibit life-like properties. This chapter follows this methodology, studying the commonalities between living organisms and the patterns that arise in a type of non-equilibrium physical system known as a reaction-diffusion system.

The models used in this paper share a particular set of properties with known living organisms: they are *precarious*, *individuated dissipative structures*. Briefly, dissipative structures are those which arise in physical systems out of thermal equilibrium, the word "precarious" implies the possibility of death – a precarious dissipative structure can be irreversibly destroyed – and individuation means the separation of dissipative structures into distinct individuals, spatially differentiated from one another and from the environment in which they exist, analogously to the separation of living matter into individual cells and organisms. This definition is similar in spirit to Maturana and Varela's (1980) notion of autopoiesis.

Many types of non-living individuated, precarious dissipative structure can be found in the physical world. One example is a hurricane, and another is the spot patterns which can form in reaction-diffusion systems, which form the basis of the experimental work in this chapter. This inclusiveness allows us to study the phenomenon of individuation in simple inanimate systems, allowing us to study it in isolation from the contingent complexities of life on Earth.

The experimental results of this chapter show that, in reaction-diffusion systems at least, individuation can occur more readily if the rate at which energy is supplied to the system is limited, or if the system contains a limited amount of some essential material or nutrient required for the formation of structures. Such situations are common in ecological scenarios.

This limitation of energy or nutrients induces a system-wide negative feedback effect which can maintain conditions conducive to the formation of distinct individuals. I will show that this effect can occur in reaction-diffusion systems with various types of autocatalytic chemical reaction, and that it can result in the production of complex individuals with multiple functionally differentiated parts. I argue that it is also likely to occur in systems where processes other than reaction and diffusion can occur, and thus could plausibly have played a role in the origin of life.

In order for such individuated dissipative structures to have been life's ancestors they must have been capable of reproduction with heredity. Full, unlimited heredity would almost certainly require complex molecules, but I will demonstrate that a very limited form of heredity is possible even in simple dissipative dissipative structures that contain only a small number of types of molecule.

The remainder of this introduction introduces the notions of dissipative structures, precariousness and individuation and argues that they are desirable qualities for a model of living phenomena. Section 5.2 introduces reaction-diffusion systems and shows that these three desiderata are met by the "spot" patterns they can exhibit and develops the analogy between spots and living organisms. The relationship of this approach to the theory of autopoiesis is discussed in Section 5.3.

The experimental results concerning individuation under nutrient and input limitation are

shown in Section 5.4. Section 5.6 discusses the implications of these results and develops a speculative picture of how this effect could have played a role in the origins of life.

5.1.1 Organisms as Chemical Engines

The second law of thermodynamics states that the entropy of an isolated system must always increase until it reaches its maximum possible value. Entropy is a property of a physical system that is greater the closer the system is to the chemical and thermal equilibrium state that its constituent matter would eventually reach if it were left in a sealed container for a long enough period of time. The second law thus says that all isolated systems eventually approach such an inert equilibrium state. The structures of living organisms are far from this equilibrium state, and hence organisms have low entropy.

An organism can survive for a time in a sealed chamber, as long as there are sufficient resources, such as food and oxygen in the case of an animal. The entropy of the chamber as a whole must increase, so in order for the organism's entropy to decrease the entropy of the matter in the rest of the chamber must increase at a greater rate. This occurs as the animal consumes the food and oxygen and excretes waste and carbon dioxide (in the case of a plant, it feeds on the high-energy photons in sunlight, which has a low entropy, and gives out heat, which has a higher entropy). One can think of entropy as being continually generated within the organism's tissue due to processes of decay, and then expelled into the environment as the organism repairs and maintains its structure. This was first pointed out by Schrödinger (1944), who wrote of organisms "feeding on negative entropy".

Another way to think of this is to picture the organism as an engine, burning fuel in order to extract work. In the case of an engine built by humans, this work might be used to pump water or move a car, and the engine's parts would gradually wear out over time until the engine no longer functioned. For this reason we build engines out of parts that wear out as slowly as possible. In the case of a living organism the work it generates is primarily used to continually repair and replace its components, which on the whole decay very rapidly. In this way the organism's structure is maintained, even though all the chemical components of an organism (with the exception of inert structures such as shells in some species) are replaced many times over its lifetime. If there is any surplus work produced, beyond that needed to repair the organism's structure, it can be used for reproduction — the construction of another, similar engine. This picture of living organisms was described by Kauffman' (2000) as a "work-constraint cycle", whereby work is used to generate thermodynamic constraints (in the form of the organism's structure), which then channel energy into work.

A key desideratum of our model organism is therefore that its structure is maintained in this way, by degrading its environment. Phenomena with this property are known as *dissipative structures* (Prigogine, 1978). We will add two further desiderata below.

Such self-producing engines are easier to build than one might at first imagine. A natural example is a hurricane. This is a heat engine rather than a chemical engine, but the principle is similar. It increases the entropy of its environment by transporting heat from the warm ocean

surface to the cold upper atmosphere. As it does so it extracts work, which it uses to maintain its far-from-trivial structure. The specific mechanism by which this occurs is complicated, but a key process involves the Coriolis effect, whereby air that gets sucked inwards towards the rising column in the eye wall gains rotational momentum. This in turn increases the efficiency of the convective processes that drives the rising of warm, moist air at the centre. There is thus a positive feedback, which is presumably balanced by a negative feedback once the hurricane reaches a certain size. In this way, the hurricane's structure remains stable despite the fact that the material from which it is constructed is continually replaced.

Hurricanes require quite specific conditions in order to form, but once one has formed it is able to persist in a much wider range of environments, essentially anywhere with a large enough vertical temperature gradient and low enough wind shear. Thus there is a sense in which, after its creation due to external factors, the hurricane itself is responsible for its own persistence. If it is perturbed sufficiently its processes will cease irrecoverably in a manner analogous to biological death, and its structure will cease to exist. The possibility of death is of course an important property of life, and has been discussed under the name "precariousness" in the autopoietic literature (Weber & Varela, 2002; Di Paolo, 2005, 2009), the concept being due originally to Jonas (1968). For a precarious dissipative structure, disrupting the structure can also disrupt processes that contribute to its production, such that the structure is no longer maintained and decays over time. This kind of precariousness will be another desideratum for our model.

Not all dissipative structures are precarious. An example of a non-precarious dissipative structure is a lenticular cloud, a cloud formation caused by a standing wave in the air pressure downwind of a mountain range. Like an organism or a hurricane, the cloud appears to maintain a constant form despite a continual flow of matter through it, but unlike an organism, if one were to destroy its structure (by removing all the water vapour of which it is composed, for instance) it would quickly be replaced as more moist air blew into the low pressure region. There is nothing in the lenticular cloud that is responsible for the cloud's own maintenance; the cloud is entirely maintained by external processes.

These two desiderata — feeding on negative entropy and precariousness — can be fulfilled by a simple chemical reaction. For instance, consider the following reaction, in which A, B and C stand for arbitrary (fictional) chemical species:

$$A + 2B \rightarrow 3B + C, \tag{5.1}$$

together with the decay process

$$B \rightarrow D. \tag{5.2}$$

We imagine these reactions taking place in a well-mixed flow reactor to which A is added at a constant rate, and the inert products C and D are continually removed. If B has a lower entropy (higher free energy) than A at the concentrations involved in the experiment, and C and D have a higher entropy, then one can see the reaction as maintaining a low-entropy concentration of B at the cost of raising the entropy of the rest of the system, by converting A into C and D. A positive

concentration of B in this system is a non-equilibrium phenomenon, since if we stop feeding A into the system and removing the products then it will progress to a chemical equilibrium in which all (or almost all) of the matter is composed of C and D.

Reaction 5.1 is autocatalytic in B, which leads to a form of precariousness. The decay process occurs at a rate proportional to b, the concentration of B, whereas B is produced by autocatalysis at a rate proportional to b^2. Thus a sufficient concentration of B is required in order for the reaction to take place, and if the concentration drops too low then it will decay toward zero, even if A is still being fed in to the system.

A lot of work in artificial chemistry, notably that of Kauffman (e.g. 2000), has focused on the origins of such autocatalytic reactions in the chemical environment of the early Earth. This is an important field of study, since biological metabolism consists essentially of a large autocatalytic network of chemical reactions. However, a reaction taking place in a homogeneous, well-mixed reactor lacks an important feature that is shared by both living organisms and hurricanes, namely the property of being an individual, spatially distinct from its environment.

In this chapter we will add this property of *individuation* as a third desideratum for our model. By individuation I mean an on-going, continuous process which produces and maintains the spatial localisation of the individual, so that the individual remains identifiable as a particular individual rather than spreading out and becoming indistinguishable from its environment. This on-goingness is what separates the individuation of organisms and hurricanes from that of, say, a metal coin, which gets stamped out in the factory and maintains its identity as a coin only in virtue of being very slow to wear.

The concept of individuation that I will develop in this chapter is perhaps in many ways similar to French philosopher Gilbert Simondon's concept of biological individuation (Simondon, 1964/1992). Hans Jonas' concept of biological individuality (Jonas, 1968) also has much in common with this approach, and another similar version of the concept has been discussed in the context of autopoiesis by McMullin (2000).

We thus have the three desiderata that our model should exhibit structures that are *dissipative*, *precarious* and *individuated*. These form a loose hierarchy: not all dissipative structures are precarious, and not all precarious dissipative structures exhibit individuation. There are other properties which are shared by all or most known living things (perhaps the most general being the presence of genetic machinery, allowing unlimited heredity), and so life exists at a deeper level on this hierarchy. Studying this more general case of individuated, precarious dissipative structures should nevertheless be instructive, since the properties of such structures are shared by all living things.

It should be noted that the notions of precariousness and dissipative structures seem fairly straightforward and easy to define, whereas individuation seems difficult if not impossible to pin down formally. It may or may not be possible to formalise the concept in the future, but the current lack of a formal definition does not prevent its study as a phenomenon, and there may be some advantage to not trying to define it prematurely. The fuzziness of the concept will be discussed in Section 5.2.2.

These desiderata bear a close relationship to Maturana and Varela's (1980) notion of *autopoiesis*. This will be discussed in Section 5.3, but first I will introduce reaction-diffusion systems and the spot-patterns that form within them as a model that satisfies these three desiderata.

5.2 Reaction-Diffusion Systems

A reaction-diffusion system is a simple spatial model of a chemical system, in which reactions take place locally, and reactants diffuse across space. This can give rise to the formation of a variety of different types of pattern. Such patterns can be directly observed in physical experiments, for example in the much studied Belousov-Zhabotinsky reaction. In addition there are many physical processes up to the scale of ecosystems whose behaviour can be more or less well approximated by reaction-diffusion equations (Cantrell & Cosner, 2003).

We use reaction-diffusion systems because they are perhaps the simplest explicitly spatial model of a physical system with realistic thermodynamic constraints that can be held out of equilibrium. Their simplicity makes them easy to simulate, yet their dynamics are rich enough to exhibit phenomena which meet our criteria. Diffusion is also a very important transport process on the scale of a biological cell.

We use a system based upon Reactions 5.1 and 5.2, which has identical dynamics to the Gray-Scott system (Gray & Scott, 1983), which was first studied in a two-dimensional reaction-diffusion context by Pearson (1993). The reactions take place on a two-dimensional surface. Every point on the surface has two values $a(x,y)$ and $b(x,y)$ representing the concentration of A and B at that point.

We imagine that the surface is immersed in a reservoir consisting of a solution of A at the concentration a_{res}. Molecules of A can exist either in solution in the reservoir or attached to the surface. If the concentration in the reservoir is higher than the concentration a at some particular point on the surface then molecules of A become attached to the surface at a rate proportional to $a_{\text{res}} - a$. The inert products C and D dissolve into the solution and do not need to be modelled. This gives rise to the equations

$$\frac{\partial a}{\partial t} = D_A \nabla^2 a - ab^2 + r(a_{\text{res}} - a); \tag{5.3}$$

$$\frac{\partial b}{\partial t} = D_B \nabla^2 b + ab^2 - kb, \tag{5.4}$$

where standard "mass action" assumptions have been made about the reactions' kinetics. D_A, D_B, r and k are parameters; the rate constant of Reaction 5.1 has been set to 1 without loss of generality by scaling the other rate constants relative to it. Initially, following Pearson, we hold a_{res} constant at the value 1, though later we will add a process that allows it to vary.

The expression $\nabla^2 a$ represents $\partial^2 a/\partial x^2 + \partial^2 a/\partial y^2$, the operation summing the second derivatives of a with respect to the two spatial directions, which defines the dynamics of the diffusion process. Pearson sets the diffusion rate parameters $D_A = 2 \times 10^{-5}$ and $D_B = 10^{-5}$, and I will use these values throughout this chapter unless otherwise mentioned.

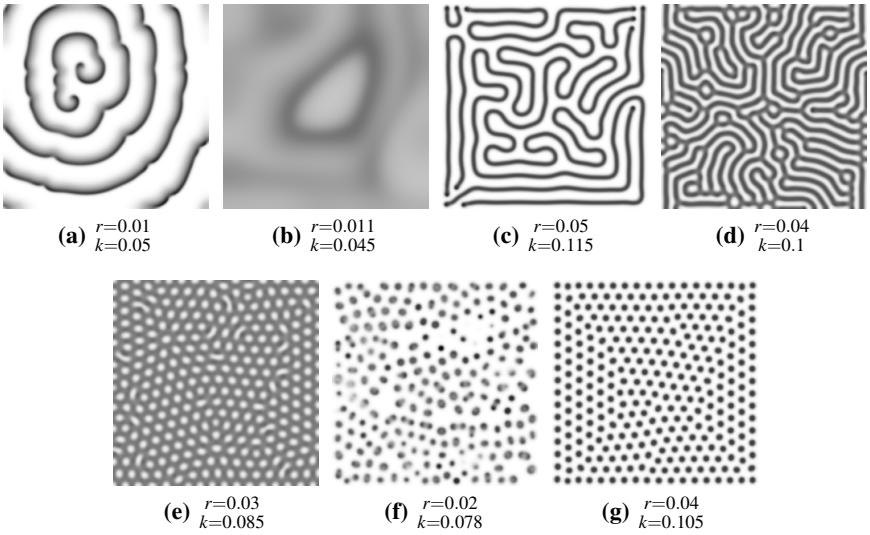


Figure 5.1: Examples showing the range patterns exhibited by the Gray-Scott system with various parameters. Integration method and initial conditions similar to (Pearson, 1993). Patterns are chosen as exemplars of various phenomena; see (Pearson, 1993) for a more systematic classification. (a) A spiral pattern; (b) A chaotic pattern of travelling waves; (c) A line pattern. Lines grow at the ends and then bend to fill space in a process reminiscent of a river meandering; (d) A labyrinth pattern; (e) A hole pattern; (f) A pattern of unstable spots, whose population is maintained by a balance between reproduction and natural disintegration; (g) A stable spot pattern. Spots reproduce to fill the space and then slowly migrate into the more-or-less organised pattern shown. With a different choice of parameters, spots can be produced that are stable but cannot reproduce.

A straightforward way to integrate such a system of partial differential equations is to discretise space into a grid and then treat the resulting system as a set of ordinary differential equations with a very large number of variables: two for each grid cell. Following Pearson, we integrate these equations using the forward Euler method and use grid cells of size 0.01×0.01 units, so that a system of size 2 by 2 is integrated using a grid of 200 by 200 cells. The resulting algorithm resembles a cellular automaton, with the important difference that each cell contains a continuously variable concentration of each of the two chemical species. These levels can be thought of as changing as a result of reactions occurring within the cells, along with diffusion between the cells.

In modelling the system using partial differential equations, a mean-field approximation has been taken, where the amount of each substance is considered a continuous quantity rather than a discrete number of molecules. However, it will still sometimes be useful to describe the equations' dynamics in terms of the creation and destruction of molecules.

Pearson found that a wide variety of spatial patterns can be formed in this system, depending on the values of the parameters r and k. Figure 5.1 shows some examples. Of particular interest are the patterns in sub-plots 5.1f and 5.1g, which consist of spatially distinct "spots" of autocatalyst. These spots satisfy the desiderata enumerated above for a model organism: they are dissipative structures, maintaining their structure by converting A into C and D, increasing the entropy of the reservoir; they are precarious, which is a consequence of the autocatalysis that underlies their

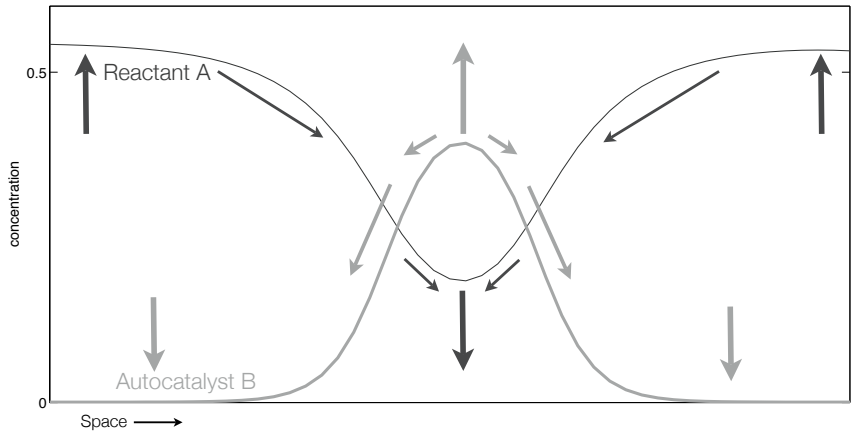


Figure 5.2: Concentration profile across a single spot in a one-dimensional version of the Gray-Scott system. This spatial structure remains constant over time due to a balance of ongoing dynamic processes. The arrows represent the direction in which these various processes occur, as described in the text.

construction (the holes in plot 5.1e are arguably individuated but they are not precarious: if a hole is filled in it will rapidly re-form); and finally the spot patterns are individuated, consisting of individually identifiable spots.

Pearson found that, depending on the values of the parameters, these spots are often capable of a form of replication: if there is an empty space nearby, the spots divide in two in a process which superficially resembles cell division. Also depending on the parameters, these patterns either approach a static configuration in which each spot persists indefinitely, or there is a continual turnover, with some spots disappearing while others replicate to fill the resulting space.

5.2.1 The Anatomy of a Spot

In order to understand the analogy between reaction-diffusion spots and living organisms we must understand the mechanism by which the spots' structure is maintained over time. The idea of this section is not to give a detailed mathematical treatment but simply to given an overview of the important features of a spot's organisation.

Figure 5.2 shows the equivalent to a single spot in a one-dimensional version of the Gray-Scott system. A cross-section through the centre of a single two-dimensional spot would look similar. The spatial pattern remains constant over time, even though the processes of reaction between the two chemical species, diffusion through space, and exchange with the external reservoir take place continually. In order to understand why the spot persists over time, we must examine the ways in which these processes are balanced.

Towards the centre of the spot there is a large amount of catalyst compared to the amount of reactant. This means that the rate of autocatalyst production is faster than the rate at which the exchange process removes it from the system. However, this excess production is balanced by diffusive flow towards the edges of the spot. Similarly, towards the centre of the spot, the reactant is used up faster than it can be transported in from the reservoir, but this is balanced by the inflow

of food from outside the spot.

Outside of the spot, towards the edges of Figure 5.2, the situation is reversed. There is a small amount of autocatalyst compared to reactant, and the removal of catalyst from the system occurs faster than it is produced (this is because of the nonlinear dynamics of the reaction: the growth rate of the autocatalyst grows with the square of its concentration, whereas its rate of removal is linear in the concentration). This rate of removal is balanced towards the edges of the spot by outflow of catalyst from the centre. Likewise, there is a net flow of reactant into the system towards the edge of the spot, which is balanced by flow towards the centre.

If the parameters of the system are within the appropriate ranges, this spot pattern is stable, in the sense that the pattern persists if small perturbations are made to the spatial distributions of the two chemical species. (We consider for the moment only parameter settings where spots do not spontaneously reproduce but remain stable indefinitely.)

However, it is important to understand the difference between the stability of a spot's structure in this sense and the presence of a locally stable attractor in the system defined by Equations 5.3 and 5.4. The system as a whole can be thought of as a very high-dimensional dynamical system, with a variable for the concentration of each chemical species at each point on the plane. This system has many attractors, corresponding to the presence of varying numbers of spots, located at varying positions within the system. After a perturbation, the spot will usually return to its previous shape, but it will not necessarily be in the same place. Thus the spot can persist while the underlying system shifts to a different attractor. This is possible because of the spatial symmetry of the underlying system.

The notion of space as defined by (at least approximate) symmetry in this way is central to the notion of individuation that I wish to express. It is the translational symmetry of space that allows two spots of the same kind to exist in the same system, and it is what allows us to say that the same spot has moved from one place to another.

A spot remains stable due to a dynamical balance between variables such that the rate at which autocatalyst is produced near its centre is equal to the rate at which this autocatalyst is transported to the outside. However, these variables are distinguished on a higher level of description than the dynamics of Equations 5.3 and 5.4. They are properties of a spot, and if the spot is perturbed so much that it disintegrates (i.e. the concentration of autocatalyst everywhere drops below the level at which it can be maintained) then these variables cease to be applicable at all: it makes no sense to ask how fast autocatalyst is transported out of the spot if there is no longer a spot.

This is also the case for the variables we use to describe living organisms. In order for an organism to be created its physiological variables must not only be given the correct values but the structure that allows us to determine such values must be brought into existence in the first place. This is to be contrasted with approaches to artificial life where the existence of physiological variables is assumed. For instance, in evolutionary robotics as described in (Beer, 2004), death is thought of as being the moment when a number of pre-defined physiological variables pass a certain "boundary of viability", a notion first proposed by Ashby (1960). Such a concept is also inherent in protocell models such as Gánti's Chemoton model (e.g. Gánti, 2003), where

the protocell is assumed to burst if its dynamical parameters exceed pre-specified bounds. In the present approach, because the variables are distinguished on a higher level than the underlying dynamics, we can say that death occurs not when the variables pass out of bounds, but when they cease to have distinguishable values altogether.

5.2.2 Individuation in the Living World

Individuation — the splitting up of living matter into distinct individuals — occurs on multiple spatial scales in the living world. On the smallest scale there is the individuation of cells. As either a single-celled organism or as part of a multicellular organism, a cell is spatially differentiated from other cells and from its surroundings. Since cells are such a universal feature of life, understanding the origin of cellular individuality is a key part of understanding the origins of life. The results in this chapter suggest a different picture from the usual one: instead of asking how the first cell arose, we might ask how an autocatalytic network that was initially homogeneously spread throughout space could have become separated into multiple individual cells. We will discuss this further in Section 5.6.2.

On a wider scale we can consider the individuation of multicellular organisms, as well as colonies of various species, which can be individuated to a greater or lesser degree. The concept of individuation can thus be linked to some of the major transitions in evolution, as identified by Maynard-Smith and Szathmáry (1997), such as the origin of multicellular life or of insect colonies. A similar point can be made about these transitions. Instead of picturing the formation of the first colony of, say, ants, we can imagine instead a gradual process where a population of solitary insects might have become concentrated into more and more distinct and cooperative colonies, with the particular population structure of modern ant colonies arising contemporaneously with or after the individuation step itself. A similar picture can be envisaged, involving a gradual transition from more-or-less homogeneous, unindividuated biofilm-like colonies to populations of modern multicellular organisms.

There are exceptions to all of these forms of individuation in the natural world. Fungi of the species *Armarilla bulbosa*, for instance, can in some situations grow indefinitely to cover many square kilometres of land (Smith, Bruhn, & Anderson, 1992), effectively losing the property of individuation and producing a structure that resembles Figure 5.1d more than 5.1f or 5.1g. Even cellular individuation can be temporarily lost in so-called plasmoidial slime moulds, species in which individual cells can fuse together, losing their separating membranes and forming a macroscopic bag of cytoplasm and nuclei.

Individuals cannot always be identified unambiguously in nature, and it is this that leads me to suspect that it will not be possible to give a formal definition of individuation, or of an individual. At what point during division does one cell become two? Does removing a part of a plant that could be made into a cutting immediately create a new individual? These ambiguities are equally present during the division of one reaction-diffusion spot into two, which adds support to the idea that the individuation of reaction-diffusion spots is a suitable model for the individuation of living organisms. If individuation is something that can arise gradually over time, it stands to reason that

it should be something that is able to happen to a greater or lesser extent.

5.3 Autopoiesis?

It is worth commenting on the relationship between the ideas expressed here and Maturana and Varela's theory of *autopoiesis* (Maturana & Varela, 1980). Maturana and Varela's work was a considerable inspiration for this research, as was Beer's analysis of Conway's game of Life in autopoietic terms (Beer, 2004) (the relationship between this work and Beer's is discussed below). However, the concept of autopoiesis can be understood in a variety of ways, and thus there is bound to be some controversy over whether the word can, strictly speaking, be used to describe the models in this chapter.

Literally meaning self-creation, autopoiesis is an etymologically appropriate word to describe the concept of an individuated dissipative structure. The definition of autopoiesis given in (Maturana & Varela, 1980) (one of the earliest works on the subject in English) seems to be compatible with the spirit of this research:

> An autopoietic machine is a machine organized (defined as a unity) as a network of processes of production (transformation and destruction) of components that produces the components which: (i) through their interactions and transformations continuously regenerate and realize the network of processes (relations) that produced them; and (ii) constitute it (the machine) as a concrete unity in the space in which they (the components) exist by specifying the topological domain of its realization as such a network.

However imperfectly expressed (key terms such as component, process, space and unity are never satisfactorily defined), this description certainly seems as if it could apply to a reaction-diffusion spot. The spot can be seen as a network of (reaction and diffusion) processes which continually regenerates itself through the balance between the autocatalytic reaction and the decay and diffusion processes, while also maintaining itself as a spatially distinct individual, or unity.

Maturana and Varela write that, as a consequence of this definition, "an autopoietic machine is an homeostatic (or rather relationstatic) system which has its own organization (defining network of relations) as the fundamental variable which it maintains constant." (Maturana & Varela, 1980) Reaction-diffusion spots also seem to fulfil this description, if one considers the organisation of a spot to be its general individuated shape, together with the relationships between processes that this entails (as in Figure 5.2). This organisation is conserved under a variety of perturbations, meeting the authors' definition of an homeostatic machine. This idea of homeostasis of organisation is a powerful concept and neatly expresses an important feature that is shared between living organisms and other individuating dissipative structures.

Unfortunately, however, the definition of autopoiesis given in (Maturana & Varela, 1980) is contradicted by definitions given in some other works by Maturana and Varela, which turn on the production of a membrane which separates the organism from its environment, rather than on the more general "constitution of a unity". (This definition is given explicitly in (Varela, Maturana, & Uribe, 1974) and is implicit in (Maturana & Varela, 1987) and much of the authors' later

work.) Reaction-diffusion spots trivially fail such definitions, since there can be no membranes in a reaction-diffusion system. A lot of work on autopoiesis implicitly or explicitly assumes this version of the definition, including most previous simulation-based models (e.g. (Varela et al., 1974), see (McMullin, 2004) for a review).

My colleagues and I have argued in the past that the importance of the cell membrane in autopoiesis has been overstated: it plays an important role in maintaining a spatially distinct individual (in autopoietic terms, constituting it as a unity) but it should not be seen as delimiting the network of processes that contribute towards the system's autopoiesis, which should be seen as extending outside of the organism's spatial boundary and into its physical environment (Virgo et al., 2009; McGregor & Virgo, 2009). The research in this chapter continues this theme: the reaction-diffusion spot example demonstrates that individuation is possible in the absence of membranes.

To my mind the presence of a membrane should be seen as a somewhat tangential concept to that of a unity or individual. Not only can a unity be constituted without being surrounded by a membrane, but a membrane does not necessarily have to surround anything. I therefore see the cell membrane as part of the mechanism by which cells are individuated, but not as a logically necessary part of the concept of their individuation.

However, because of the controversy about the role of the membrane I have met a certain amount of resistance from researchers in the field to the idea that reaction-diffusion spots can be seen as a model of autopoiesis. Perhaps another reason for this resistance is that Maturana and Varela originally set out to show that autopoiesis is both necessary and sufficient for life. If one's goal is to define life then one's instinct will tend to be to refine the definition rather than allow inanimate structures such as hurricanes and reaction-diffusion spots to meet it. My instinct differs because I am more interested in the similarities between living and inanimate dissipative structures than in the differences. I must therefore allow the reader to decide whether autopoiesis is an appropriate word.

Another difference between the theory of autopoiesis and the view taken in this thesis is that Maturana and Varela consistently described autopoiesis as being an all or nothing property: a system is either autopoietic or it is not, with nothing in between, in contrast to the inherently fuzzy view of individuation described in the previous section.

Even if one doesn't consider them to be strictly autopoietic in their own right, however, it is instructive to consider reaction-diffusion spots in autopoietic terms. This is the approach taken by Beer (2004), who strives to make concepts from the theory of autopoiesis more concrete by showing how they might apply to gliders in Conway's Game of Life. Beer raises a number of challenges in this work, regarding the development and analysis of better models of autopoietic systems, in order to further clarify the concepts.

I believe the present work answers some of those challenges. Reaction-diffusion spots seem to have several advantages over gliders, which make them more suited to this kind of study. One such advantage is the physical interpretation of reaction-diffusion models. Beer questions in his paper whether turning cells on and off really counts as production of components, but in the case of reaction-diffusion systems it seems rather more clear-cut: we can identify the physical processes

of reaction and diffusion and we can see clearly that the spot's molecular components are produced through the autocatalytic reaction.

Additionally, because reaction-diffusion systems are physical models, the processes that occur within them are constrained by the second law of thermodynamics, an aspect of living systems that has generated a lot of recent discussion in the autopoietic literature, e.g. (Moreno & Ruiz-Mirazo, 1999).

However, a more significant advantage of reaction-diffusion spots over gliders stems from the continuous nature of the model. Beer writes that gliders have a very limited "cognitive domain", meaning that there are very few perturbations that can be applied to a glider without destroying it, and few of those cause any change in its state, making for a very limited behavioural repertoire. This stems from the discrete nature of the Game of Life, and from the simplicity of gliders. Because of this, Beer switches to a higher level of abstraction in the latter part of his paper, whereby a model agent's internal dynamics are specified arbitrarily rather than arising from an underlying model of autopoiesis.

In contrast, reaction-diffusion spots can be perturbed in an infinite number of ways. Any temporary change in the concentration pattern of food or of catalyst will not destroy the spot, as long as it is sufficiently small. However, the spot's shape will be temporarily changed, and it may well move to a slightly different spatial position as a result of the perturbation. Spots therefore have a rich cognitive domain, resisting a wide variety of perturbations, and exhibiting behavioural responses to them.

We will see shortly (in Figure 5.4) that spots exhibit a form of "chemotaxis", moving along concentration gradients of the food chemical. This is closely related to their individuation (their tendency to move away from low food concentrations is a mechanism that keeps them apart from one another) and seems related also to their ability to resist perturbations. Perhaps this represents evidence of the deep relationship between autopoiesis and cognition that Maturana and Varela (1980) proposed. Perhaps it would be meaningful to say that the spot is actively resisting the perturbation induced by the gradient by moving away from it. The philosophical issues here are complicated, however, and sorting them out is a task for future work.

5.4 Individuation under Negative Feedback

As Figure 5.1 shows, there are many types of pattern which can be exhibited by the Gray-Scott system. In addition to the patterns shown in Figure 5.1, there are regions of parameter space where nothing can persist, and regions where any autocatalyst spreads out homogeneously across space. Individuated spots only form within quite a specific region of parameter space (i.e. a specific range of values for f and r). In order for them to form, the correct balance is required between the rate at which food is input and the other parameters of the system. Too little food availability and it becomes impossible for spots to persist; too much and they spread out into a homogeneous region in which the concentration of autocatalyst is constant, or into other, less individuated patterns such as spirals or labyrinths.

At first sight this seems like it might be a problem for our model. Individuation is a fairly

ubiquitous feature of living organisms, whereas the reaction-diffusion model has to be finely tuned in order for it to occur. However, in this section we will demonstrate that when the system is changed so that its capacity to support autocatalyst is limited, individuation often occurs as a result.

Limiting the system's capacity introduces a *system-wide negative feedback*, whereby an increase in activity in one part of the system has a negative effect on the activity in all other parts of the system. This feedback effectively does the tuning for us, keeping the system in a state where individuation occurs. Such feedbacks are extremely common in ecological scenarios, making it plausible that this kind of regulation played a role in the origin of biological individuation.

Two important ways in which such a system-wide feedback can occur are *input limitation* and *nutrient limitation*. I have borrowed these ideas from ecology, where they represent processes that can limit the growth of a natural population or ecosystem. In this context, the difference between an input (sometimes called a "resource" in the ecological literature) and a nutrient is that nutrients can be recycled within the system whereas an input is something that must be continually supplied to the system.

Input limitation means that there is a negative feedback in the amount of energy or some other input supplied to the system: the greater the total rate at which energy is used up in the system, the less is available on average at each point in the system. In an ecological context, such a system might occur naturally in a system such as a lake, where energy-providing molecules are washed into the system at a constant rate by rivers. If these food molecules are being used up only slowly then their concentration will build up and there will be a high availability of food in all parts of the lake. If there is a large population of organisms using up the food then the concentration will drop and there will be a smaller amount available per unit time to each individual. Assuming that the population does not explode to the point where it uses up too much of the food and subsequently crashes to zero, the system will converge to an attractor such that the population's average size is just right to use up the food at the rate it enters the lake.

Nutrient limitation, on the other hand, occurs when there is a limited supply of a substance which is needed for the construction of an organism but which does not supply energy and is returned to the system when the organism dies (think nitrogen or phosphorous for biological organisms). The finite amount of nutrient available to be recycled in this way can mean that the system is able to support only a limited population.

5.4.1 Results: Input Limitation and Nutrient Limitation in Reaction-Diffusion Systems

Figure 5.3 shows a physical setup that results in input limitation in a reaction-diffusion model. This model closely resembles the lake scenario described above. The surface on which reactions take place is immersed in a well-mixed reservoir as described in Section 5.2, except that the concentration of the food molecule A in the reservoir, a_{res}, is no longer assumed to be constant. Instead it is added to the reservoir at a constant rate λ. Its concentration then depends upon the rate at which it is used up by the autocatalytic reaction, and we expect an equilibrium to be reached where the amount and configuration of autocatalyst on the surface is such that the inflow λ of

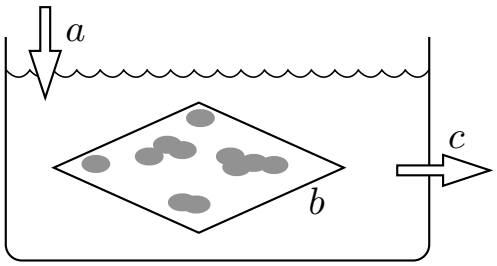


Figure 5.3: Diagram showing a model of input limitation. (a) the "food" resource A is fed into a finite sized reservoir at a constant rate; (b) the reactions $A+2B \rightarrow 3B+C$ and $B \rightarrow D$ take place on a surface immersed in the reservoir; (c) the inert products C and D are removed from the reservoir.

reactant into the reservoir is balanced by its flow onto the surface and ultimately out of the system.

The dynamics of this new system are again given by Equations 5.1 and 5.2, except that a_{res} is now considered a variable rather than a parameter. Its dynamics are given by

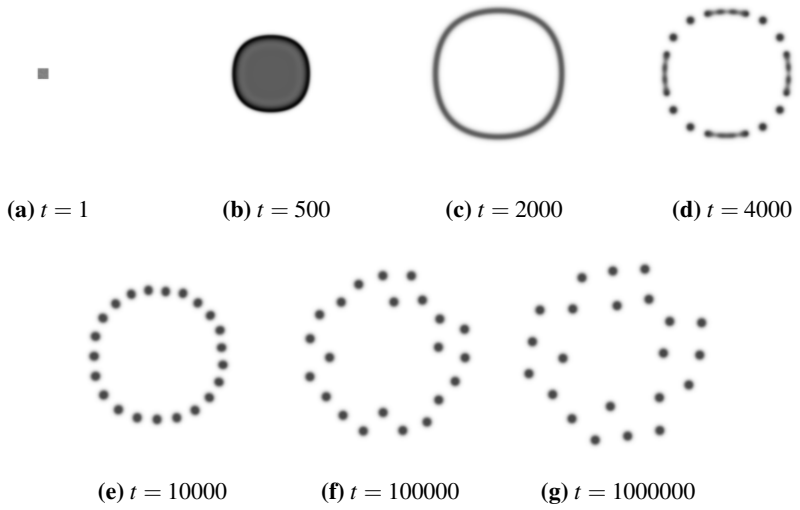
$$\frac{da_{\text{res}}}{dt} = \frac{1}{w}\left(\lambda - r\iint\Big(a_{\text{res}}(t) - a(x,y,t)\Big)\,dy\,dx\right), \tag{5.5}$$

where the integral is taken over the whole area on which the reactions take place, and w is a time constant proportional to the size of the reservoir. Unlike a and b, a_{res} is not a function of space, since the reservoir is well-mixed.

Figures 5.4 and 5.5 show an example of the dynamics of these equations, for a particular choice of parameters. The details of the dynamics differ according to the choice of parameters but the behaviour seen in Figures 5.4 and 5.5 is a common result across large regions of the parameter space: the growth of the autocatalyst is slowed by a reduction in the food concentration a_{res} and the system converges to a state in which there are a number of individuated spots.

This is because the total input of the food chemical to the system is limited to λ, which means that only a certain amount of autocatalyst can be supported in the system. Individuation is not the only possible outcome of this setup, however. With different parameter settings (particularly when k is low) the autocatalyst can tend to remain in one large area resembling Figure 5.4b. Such a configuration is stable and will tend to return to a roughly circular shape after a perturbation. Similarly, for a narrow range of parameter values, the system can become stuck in a line pattern similar to Figure 5.4c. However, the addition of this kind of negative feedback does seem to dramatically increase the range of parameters for which individuation occurs (see Figure 5.6 below).

In order for individuation to occur the area of the surface on which the reactions take place must be sufficiently large. If it is not then space, rather than resource input, can become the limiting factor. The parameter w, proportional to the size of the reservoir, must also be chosen appropriately, because it is possible for a situation to arise where the concentration of A in the reservoir initially grows very high, leading to an overshoot in the concentration of autocatalyst on the surface, which causes the food concentration to subsequently be depleted to the point where the autocatalyst can no longer persist, and the system decays to the homogeneous equilibrium where no autocatalyst is present. This tends to happen when w is too large.



(a) $t = 1$ **(b)** $t = 500$ **(c)** $t = 2000$ **(d)** $t = 4000$

(e) $t = 10000$ **(f)** $t = 100000$ **(g)** $t = 1000000$

Figure 5.4: A typical sequence of snapshots showing individuation occurring in response to input limitation. Equations 5.3, 5.4 and 5.5 were integrated in a space of size 2 units by 2 units. The parameters used were $r = 0.04$, $k = 0.09$, $L = 50$ and $w = 10^6$. The concentration of autocatalyst B over space is shown, with darker shades representing higher concentration. (a) The initial conditions were a small slightly randomised square of autocatalyst in the centre of the system, with the rest of the system containing no autocatalyst and a concentration of reactant A of 1.0. The initial value of a_{res} was 1.0. (b) Initially the autocatalyst populates a growing area of the system. The concentration of autocatalyst is higher at the edges of this region because of reactant flowing in from the region of high food concentration outside it. (c) This growth eventually reduces the amount of reactant in the reservoir to the point where the autocatalyst can no longer persist in the interior of the region. The circular line pattern thus produced persists for around 1500 time units. (d) Eventually it becomes unstable, splitting up into individual spots. This is the point at which individuation occurs. (e)–(f) the spots slowly migrate in response to concentration gradients induced by the other spots.

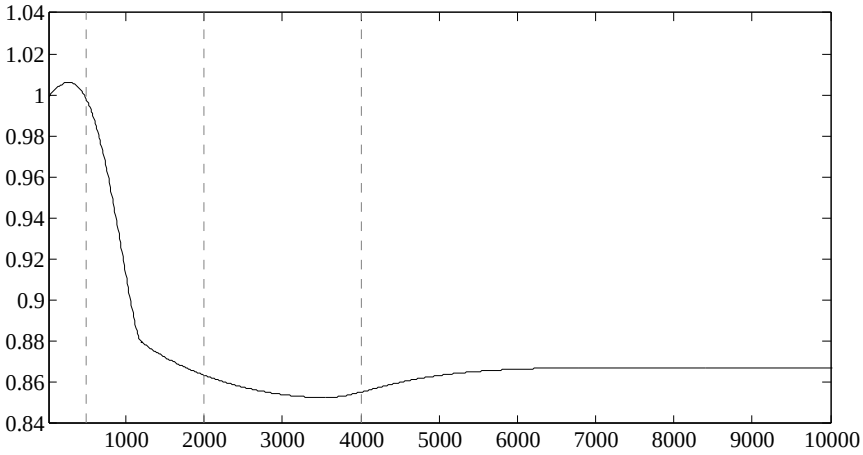


Figure 5.5: A graph of the concentration a_{res} of A in the reservoir over time for the configuration described in the caption to Figure 5.4. The dashed lines indicate the times at which the snapshots in Figures 5.4b, 5.4c and 5.4d were taken. The dying away of the central area of autocatalyst concentration occurred at around time 1100–1200, corresponding to a change in the rate of decline of a_{res}, and individuation occurred from time 3000–4000. The slow migration of the spots after time 10000 seen in Figures 5.4e–5.4g is accompanied by a very slow reduction of a_{res} towards the value 0.865.

The results for nutrient limitation are similar to those for input limitation. In order to implement nutrient limitation in a reaction-diffusion system we change reactions 5.1 and 5.2 to

$$A + 2B + E \rightarrow 3B + C, \tag{5.6}$$

$$B \rightarrow D + E. \tag{5.7}$$

Species E plays the role of the nutrient. It is used up along with A in order to create molecules of B, and is returned when B decays. We assume the molecules of E inhabit the reservoir rather than the surface, and that E cannot enter or leave the reservoir except through these reactions. This implies that the total amount of E in the reservoir, plus the amount of B on the surface, is a constant, denoted E_{total}.

The dynamics then become

$$\frac{\partial a}{\partial t} = D_A \nabla^2 a - e_{\text{res}} a b^2 + r(a_{\text{res}} - a); \tag{5.8}$$

$$\frac{\partial b}{\partial t} = D_B \nabla^2 b + e_{\text{res}} a b^2 - k b, \tag{5.9}$$

$$\text{where} \quad e_{\text{res}} = \frac{1}{w_e}\left(E_{\text{total}} - \iint b(x,y,t)\,dy\,dx\right). \tag{5.10}$$

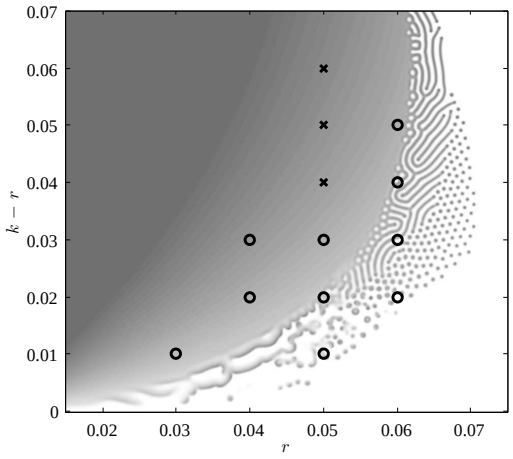
a_{res} is again considered a constant parameter in these equations, since input limitation is now not taking place. The other parameters are r, k, the diffusion rates, and w_e, which is proportional to the volume of the reservoir. The system-wide nature of the feedback comes from the dependence of e_{res} on the behaviour of the whole surface, as indicated by the integrals. When there is a large amount of autocatalyst B in the system its growth rate is effectively reduced due to the low value of e_{res}.

The dynamics of these equations are broadly similar to those of the input limitation equations, with individuation often taking place in a manner that closely resembles the sequence of configurations shown in Figure 5.4. Similarly to the input limitation case, the area of the surface must be large enough compared to E_{total}/w_e in order to precent space from being the limiting factor. Unlike the input limitation case, it is not possible for the amount of autocatalyst to overshoot the concentration of nutrient and subsequently crash. This is due to the lack of a time lag in the nutrient limitation equations.

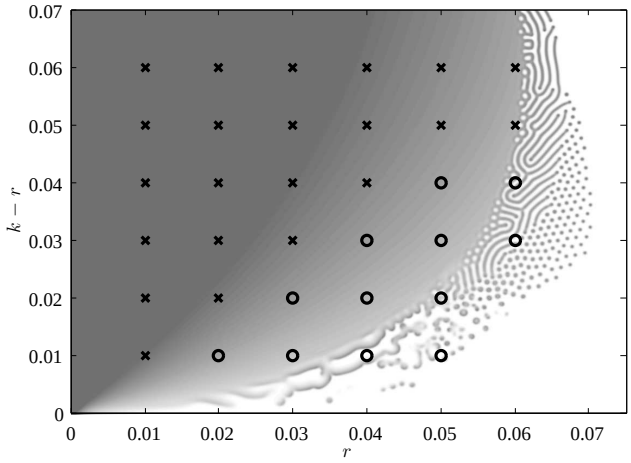
Figure 5.6 summarises the results of adding either input or nutrient limitation to the system, showing that either form of negative feedback increases the range of parameters for which individuation occurs.

5.4.2 More Complex Structure

One obvious question which arises from the results of the previous section is whether this phenomenon of individuation occurring as a response to negative feedback is a general phenomenon which could apply to a general class of physical systems, or whether it is a contingent feature of the Gray-Scott system. This is an important question because if this phenomenon were only a



(a) Input Limitation



(b) Nutrient Limitation

Figure 5.6: Summary of results from experiments using input and nutrient limitation. r is plotted against $k-r$ in order to better fit the results into a rectangular plot. All combinations of r between 0.01 and 0.06 and $k-r$ between 0.01 and 0.07 were tried, each in steps of 0.01. If the system eventually converged to an individuated spot pattern the corresponding point is marked **O**, whereas if after 100000 time units the pattern was not individuated it is marked **×**. The absence of a symbol indicates that all autocatalyst died out and the system converged to the homogeneous equilibrium. The regions in which individuation occurs in the original system can be clearly seen in the background pattern, which is generated by continually varying the parameters of the original system (without negative feedback) over space. The parameters in (a) are the same as in Figures 5.4 and 5.5 (except for the variation of r and k), and in (b) the values $C_{\text{total}} = 10^4$ and $w = 10^4$ were used.

feature of the Gray-Scott system then it would be unlikely to be relevant to the origins of life.

There is some evidence that it might be a general phenomenon. Kerner and Osipov (1994) study reaction-diffusion spots as an example of a general class of phenomena called autosolitons, claiming that phenomena with similar dynamics occur in a wide class of physical substrates including semiconductors, plasmas and hydrodynamic systems. We might thus expect individuation to occur under negative feedback in these systems as well. I will argue further for the generality of this phenomenon in the discussion section.

The question of how general the phenomenon of individuation under negative feedback is cannot be answered fully without performing a wide range of experiments across many types of non-equilibrium physical system. However, we can go some way towards an answer by constructing reaction-diffusion systems with different reaction schemes. This should add plausibility to the idea that the phenomenon occurs generally in systems where diffusion and autocatalysis are important factors in the dynamics.

One way to do this is to replace the autocatalytic reaction step $A+2B \rightarrow 3B+C$ with a multi-step scheme such as $A+B \rightarrow E$, $E+B \rightarrow 3B+C$ (which has the same stoichiometry), or with a more complicated autocatalytic set such as $A+2F \rightarrow 2F+G+C$, $A+2G \rightarrow 2G+F+C$. Nutrient or input limitation can then be added as before. Doing this results in individuated spots with properties similar to those found in the Gray-Scott system, but without any qualitatively different behaviour or structure: the spots are simply composed of a mixture of the species involved in the autocatalytic reaction. Since these results are similar to those of the previous section they are not shown here.

However with an appropriate reaction scheme we were able to produce individuated patterns which differ from the simple spots found in these systems. The individuated structures thus produced are more complex than a single spot and consist of multiple spatially differentiated parts. The reaction scheme is

$$A+2U+X \rightarrow 3U+Y+C, \tag{5.11}$$

$$A+2V+Y \rightarrow 3V+X+C, \tag{5.12}$$

$$U \rightarrow D_1, \tag{5.13}$$

$$V \rightarrow D_2. \tag{5.14}$$

Here U and V are autocatalysts which operate independently, feeding on the same food A. X and Y can be seen as nutrients (though unlike the nutrient in section 5.4.1, X and Y inhabit the surface rather than the reservoir). The important thing to note is that the nutrient required by U is produced by V and *vice versa*. Also note that X is converted into Y by reaction 5.11 and Y is converted back into X by reaction 5.12. X and Y are assumed never to enter or leave the surface, so the total amount of X plus Y is conserved.

This means that although the two autocatalysts U and V compete for the same food resource they also rely on each others' operation in order to convert the nutrient back into the form they require. In this respect this scheme is similar to the concept of the hypercycle (developed in the

context of the RNA world hypothesis (Eigen & Schuster, 1979) but applicable to many more general situations in biology, see (Maynard-Smith & Szathmáry, 1997, pp. 51–58)).

One might expect that under the right conditions the two autocatalysts might form separate spots (with competition for food preventing them from occupying the same space) but that they would tend to migrate to positions close to spots of the other autocatalyst because that is where the concentration of their required nutrient is highest. We will see shortly that this can indeed happen.

In order to find the region of parameter space in which the patterns in this system are individuated, we use the input limitation scheme described in section 5.4.1. This is a useful technique for finding individuating regions of parameter space, because the tuning effect lessens the need for searching, and also because the initially high rate of resource input means that a wider range of initial conditions are able to persist.

Assuming mass action kinetics as usual, reactions 5.11–5.14 result in the following equations:

$$\frac{\partial a}{\partial t} = D_A \nabla^2 a - \alpha_u axu^2 - \alpha_v ayv^2 + r(a_{\text{res}} - a); \tag{5.15}$$

$$\frac{\partial u}{\partial t} = D_U \nabla^2 u + \alpha_u axu^2 - k_u u; \tag{5.16}$$

$$\frac{\partial v}{\partial t} = D_V \nabla^2 v + \alpha_v ayv^2 - k_v v; \tag{5.17}$$

$$\frac{\partial x}{\partial t} = D_X \nabla^2 x - \alpha_u axu^2 + \alpha_v ayv^2; \tag{5.18}$$

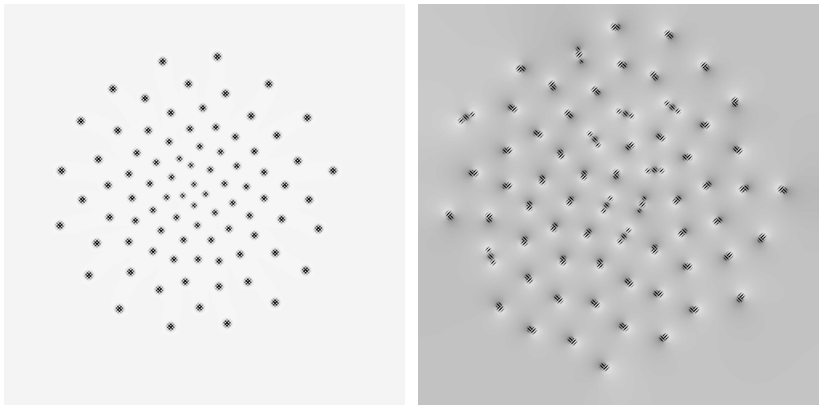
$$\frac{\partial y}{\partial t} = D_Y \nabla^2 y + \alpha_u axu^2 - \alpha_v ayv^2. \tag{5.19}$$

The dynamics of a_{res} are given by equation 5.5. One set of parameter values that produces the desired result is as follows: $D_A = 10^{-4}$, $D_U = D_V = 0.5 \times 10^{-5}$, $D_X = D_Y = 0.3 \times 10^{-4}$, $\alpha_u = \alpha_v = 3.0$, $k_u = k_v = 0.1$, $r = 0.04$, $w = 10^4$ and $\lambda = 100$. We used a 2.5×2.5 surface and integrated with a time step $\Delta t = 0.15$.

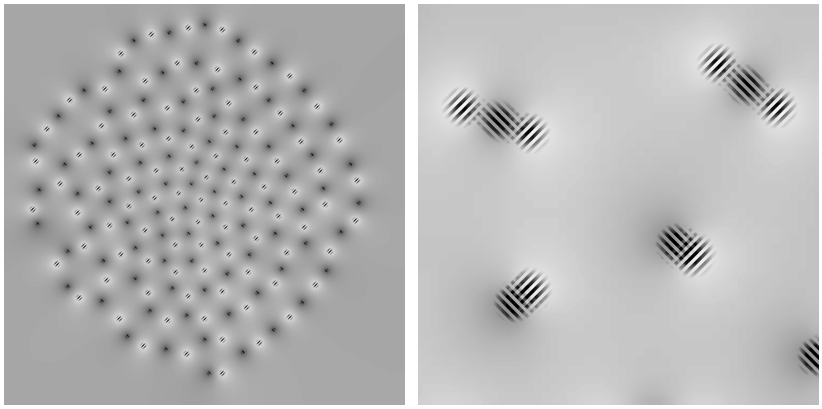
Since the total amount of X plus Y is conserved, the behaviour of the system depends on the amount of X and Y initially present in the system. The initial conditions we used were an initial reservoir concentration $a_{\text{res}} = 0$; $a = 1.0$, $u = v = 0$ everywhere in the system except for a 0.125×0.125 square in which $a = 0.5$ and $u = v = 0.7$, $\pm 10\%$ random noise; and x and y set everywhere to the same value, which we varied.

These results are shown in Figure 5.7. With an initial concentration of x and y equal to 0.9, the system individuates into structures consisting of one or two spots of U, closely associated with one or two spots of V. These clusters are spatially separated from one another and can be seen as individuals in their own right, in the same sense that the spots in the Gray-Scott system are individuals.

The differentiated parts of these multispots are somewhat analogous to the parts of an organism's anatomy: they each perform different functions and cannot persist without the rest of the organism (a spot of U on its own would use up all the available X by converting it to Y unless there were a nearby spot of V to convert it back again). In Figure 5.7c the relationship looks more like a symbiosis between individuals of separate species, since the spots of U and V are not paired



(a) $u_0 = v_0 = 0.5$ **(b)** $u_0 = v_0 = 0.9$



(c) $u_0 = v_0 = 1.5$ **(d)** a magnified section of plot b.

Figure 5.7: Three snapshots of a system with containing two autocatalysts which are interdependent yet compete with each other for food. Depending on the amount of nutrients, the system can form spots that contain a mixture of both autocatalysts (a); "complex" spots with spatially differentiated parts (b, d); or separate spots of each autocatalyst (c). The system is described by equations 5.18–5.19 and 5.5, with parameters as described in the text. The system size is 2.5×2.5 units and each image was taken after 105000 time units. In order to display the data clearly in black and white we have linearly interpolated between the pixels and superimposed cross-hatch patterns in opposite directions over the concentrations of the autocatalysts U and V. The concentration of nutrient X is indicated in grey, although the contrast has been increased for clarity, so the darkness of each pixel is not proportional to the concentration. The concentration pattern of nutrient Y can be inferred, since where there is less X there is more Y and vice versa. The concentration of food A is not shown.

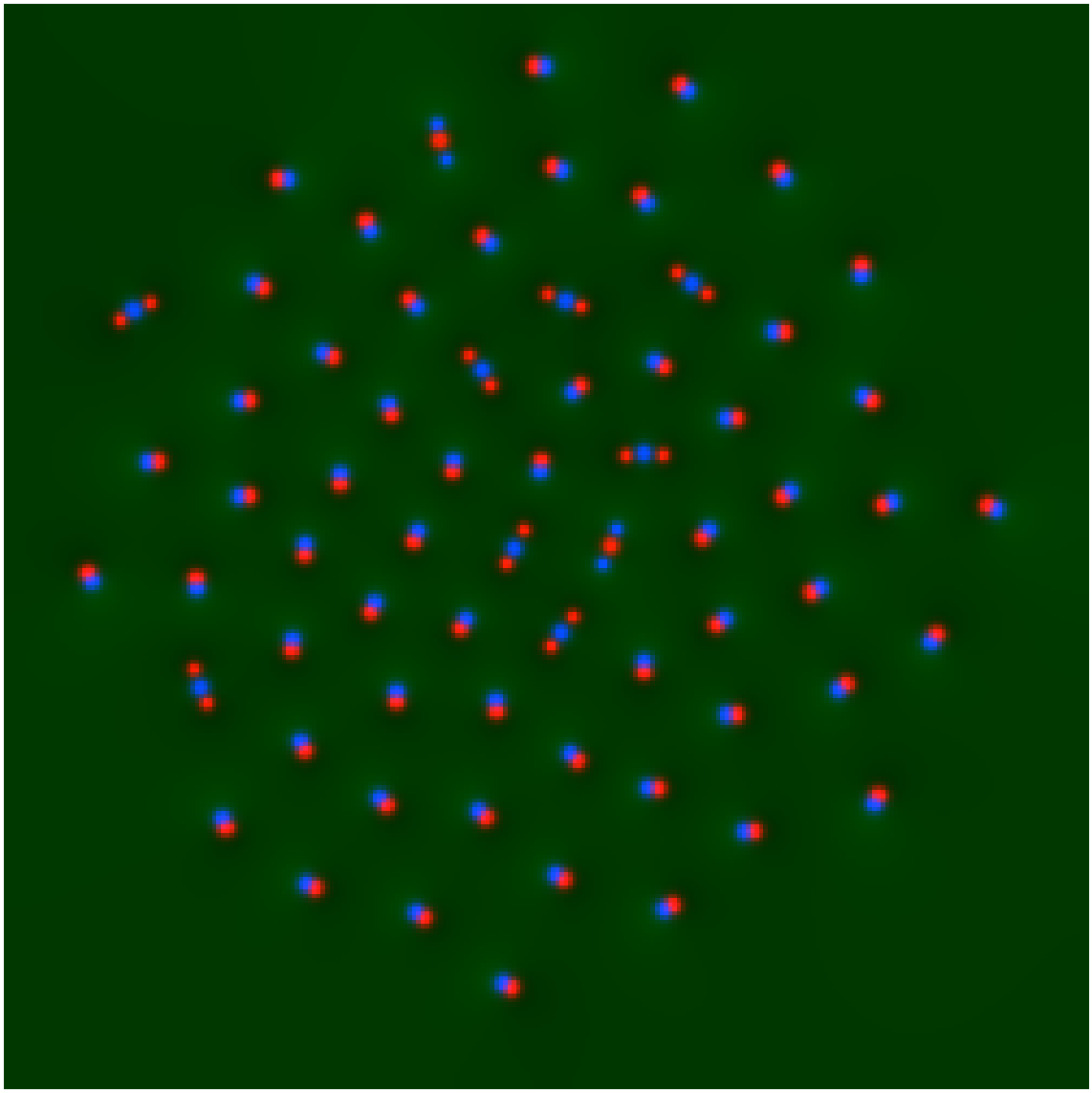
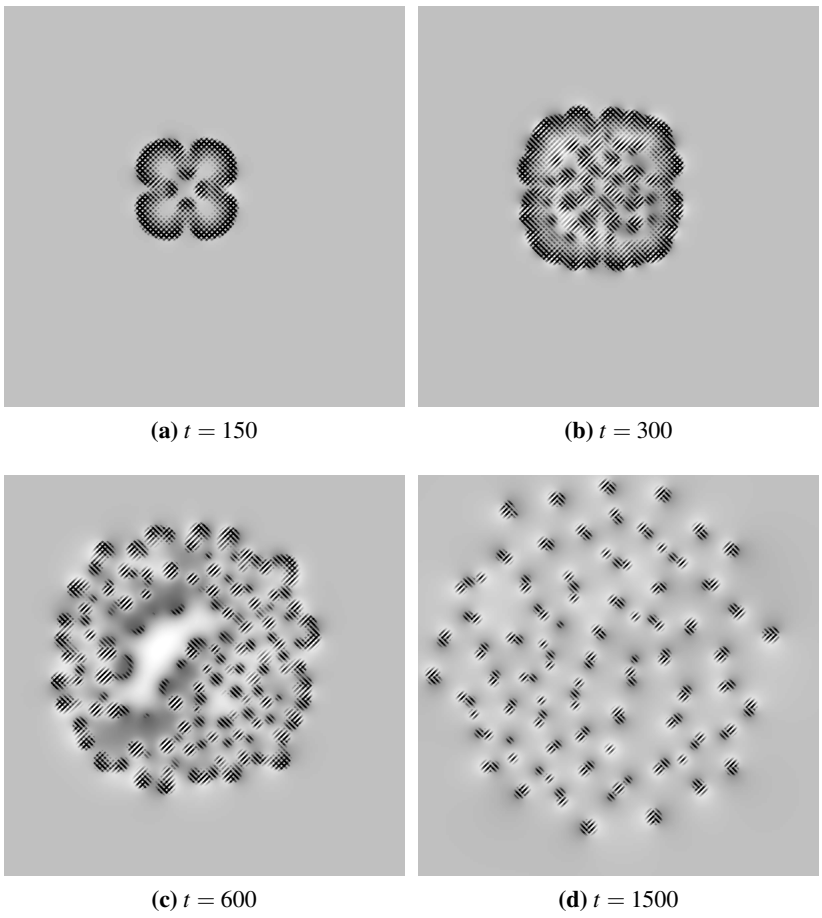


Figure 5.8: A colour version of Figure 5.7b, showing the symbiotic coexistence of spots of different autocatalysts. The concentrations of U, V and X are indicated by the intensity of the blue, red and green component of each pixel respectively. The concentrations of Y and A are not shown.



(a) $t = 150$ **(b)** $t = 300$

(c) $t = 600$ **(d)** $t = 1500$

Figure 5.9: Some snapshots from the individuation process leading to figure 5.7b (magnified 160%). Initially the autocatalytic species are homogeneously mixed (a) but the decreasing availability of food *A* leads to a bifurcation, causing the species to become spatially separated (b). Distinct spots of each autocatalyst form (c) but their configuration is not entirely stable, which leads to oscillations in their size, which can sometimes lead to a region of spots dying out (this is the cause of the white region towards the centre of plot c). As the pattern slowly spreads to fill more of the space, each spot migrates towards a spot of the opposite autocatalyst (d). Both spots benefit from this arrangement because of the exchange of the two nutrients. These "symbiotic" pairings of spots are stable and do not oscillate.

up together, but nevertheless each type of spot relies on individuals of the other type to recycle their waste products. Varying the initial concentrations between 0.9 and 1.5 produces a continuous range of behaviours, with the spots of U and V becoming less strongly associated with one another until they appear intermixed as in Figure 5.7c. This demonstrates another way in which individuation is not an all-or-nothing phenomenon.

Figure 5.9 shows the progression from the initial state (a homogeneous square of U and V, with a little noise added to break the symmetry) towards the individuated pattern of multispots in Figure 5.7b. Initially the high availability of food allows this autocatalytic mixture to spread homogeneously, but as the food availability declines this becomes impossible, and the final pattern of clustered spots gradually arises. This is interesting because it shows an example of a simple, disorganised structure giving rise to a more complex and more finely organised one, without the need for evolution by natural selection. Perhaps this can give us some insight into how complex living structures first arose in nature; we will return to this theme in the discussion section.

Although the results presented in this section depend fairly heavily upon the parameters of the system, they do at least indicate that individuation under negative feedback is not only a property of the Gray-Scott system. It can also occur in reaction-diffusion systems with more complex reaction schemes, and can result in individuated structures more complex than a single spot.

5.5 Heredity in Reaction-Diffusion Spots

The simulations in the previous sections exhibit precarious, individuated dissipative structures, but modern organisms have an additional important property that is not shared by these structures: the capacity for evolution by natural selection. In order for the structures to exhibit the full "unlimited heredity" of modern organisms (*sensu* Maynard-Smith & Szathmáry, 1997) the system would almost certainly have to be given the capacity to form complex information-carrying molecules, which would require entirely different modelling techniques. However, it is possible to give spots a limited capacity for heredity with variation simply by choosing an appropriate reaction-scheme. In this section I will propose two ways in which this can be done, including one example in which spots come in two different types, which exhibit different behaviour and are better able to survive in different types of environment.

One can easily produce a system in which there are two distinct types of spot, by using two distinct autocatalysts, with a reaction scheme along the lines of

$$A + 2R \rightarrow 3R + C; \tag{5.20}$$

$$A + 2S \rightarrow 3S + C, \tag{5.21}$$

plus a decay reaction for each autocatalyst. If these reactions have similar rates the two chemical species will behave in similar ways, and, with the appropriate choice of r, will form spots that can reproduce independently from one another. In this case the offspring of a spot of species R will always be composed of R, and likewise for S. However, while it exhibits two different types of spot this scheme lacks the possibility of mutation between them, and hence it cannot be called

heredity with variation. (Spots composed of a mixture of R and S cannot persist due to competitive exclusion.)

One can improve this situation somewhat by adding the reactions

$$A + R + S \rightarrow R + 2S + C \qquad \text{and} \tag{5.22}$$

$$A + R + S \rightarrow S + 2R + C. \tag{5.23}$$

If the rates of all four catalytic reactions are the same then spots composed of a mixture of R and S (in any proportion) can be supported because there is now no functional difference between a molecule of R and a molecule of S. Assuming that the spatial distributions of R and S are not identical, mixed spots' offspring tend to be composed of the same proportions of each species as their parents, so that if any process changes the concentration of spot, this will be inherited. An example set of parameters for which this occurs is $a_{\text{res}} = 1$, $r = 0.02$, both decay rates equal to 0.078 and all four catalytic reaction rates equal to 1. This system is not very stable however, in the sense that if the catalytic reaction rates differ only slightly from each other then the relative concentrations of the two autocatalysts will tend towards a particular value.

A more satisfying example of limited heredity involves an extension to the Gray-Scott system in which two distinct kinds of self-replicating spot can be formed: one with and one without a "tail" formed from a second autocatalyst that feeds on the first. Perhaps surprisingly this tail does not destroy the spot it is attached to but rather changes its behaviour, making it move constantly away from its tail. When a spot with a tail fissions it usually results in two spots which also have tails. Spots with tails generally reproduce faster than those without, and in some environments the tail provides a selective advantage over tail-less spots.

This system exhibits limited heredity with variation, because the tails occasionally become detached from the spots and then die due to the lack of food, leaving a spot of the tail-less form, which can then reproduce on its own to produce tail-less offspring. Conversely, it is possible for a tail-less spot to be "infected" with tail material from a nearby tailed spot, although with the parameters quoted below this happens rarely if at all. Nevertheless, there are two distinct types of reproducing individual, with the possibility for one of them to "mutate" into the other, making this an example of perhaps the most limited possible form of heredity.

These spots with tails can be produced by adding an additional species to Reactions 5.1 and 5.2. This new species has the same kind of autocatalytic dynamics as species B, except that it feeds upon species B rather than species A:

$$A + 2B \rightarrow 3B + C_1, \tag{5.24}$$

$$B + 2E \rightarrow 3E + C_2, \tag{5.25}$$

$$B \rightarrow D_1, \tag{5.26}$$

$$E \rightarrow D_2, \tag{5.27}$$

which gives rise to the equations

$$\frac{\partial a}{\partial t} = D_A \nabla^2 a - \alpha_a a b^2 + r(a_{\text{res}} - a); \tag{5.28}$$

$$\frac{\partial b}{\partial t} = D_B \nabla^2 b + \alpha_a a b^2 - \alpha_e b e^2 - k_b b; \tag{5.29}$$

$$\frac{\partial e}{\partial t} = D_D \nabla^2 b + \alpha_e b e^2 - k_e e, \tag{5.30}$$

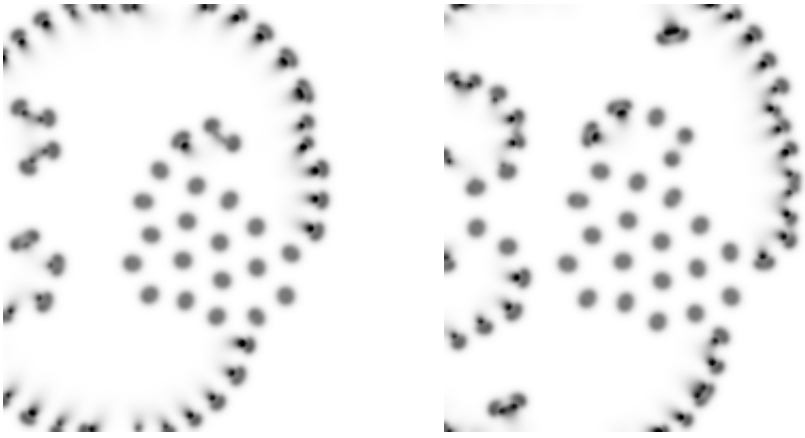
where α_a, α_e, k_b and k_e are the rate constants of reactions 5.24, 5.25, 5.26 and 5.27 respectively. (System-wide negative feedback is not used in this example.) With the appropriate choice of parameters, species E forms into "tail" structures that trail behind individual spots. These tails are parasitic on the spots, in the sense that they make no contribution to their metabolism. Species E only serves to deplete species B and makes no direct contribution to increasing it. However, with the correct choice of parameters the tail does not destroy the spot but rather changes its behaviour, causing it to continually move away from its tail.

Figure 5.10 shows some snapshots from such a system. The parameter values used for this example are $D_A = 2 \times 10^{-5}$, $D_B = 1 \times 10^{-5}$, $D_E = 1 \times 10^{-6}$, $a_{\text{res}} = 1$, $r = 0.0347$, $\alpha_a = 1$, $\alpha_e = 0.8$, $k_b = 0.2$ and $k_e = 0.005$. These parameters were chosen according to two criteria: firstly the behaviour exhibited permits limited heredity as described above, and secondly the tails are neatly separated from the spots, which facilitates visualisation of the results in a black and white image.

A note on the initial conditions: it can be quite tricky to find a set of initial conditions for this system in which spots with tails will form, because the concentration of E tends to drop too low before it can become established. For this reason I started this example with a_{res} at 1.02 and gradually reduced it to 1.0 over 2000 time units. The higher food availability makes it easier for concentrations of B and E to persist initially, and the slow change tends to result in individuated structures that are persistent in the new regime. Such recipes for baking patterns are an invaluable tool for exploring the parameter space of these systems. This technique was used along with an initial square of size 0.1×0.1 in which the concentration of A and B are both 0.5, $\pm 10\%$ random noise, (A is 1.0 everywhere else) adjacent to a square of size 0.07×0.07 in which the concentration of E is 1.5, to produce Figure 5.10.

As Figure 5.10 shows, with these parameter settings, although spots with tails are faster at colonising empty parts of the system, spots without tails are in some sense more stable and eventually out-compete them. However, we can modify this example to allow both types of spot to persist by occasionally clearing an area of the system. In Figure 5.11 this is accomplished by setting the concentration of A in a randomly chosen 0.5×0.5 area of the system to zero every 1000 time units in the right-hand part of the system. The concentration of A recovers rapidly, but not before any spots within the area (with or without tails) have been destroyed. This creates a niche in which the spots with tails have an advantage, because there is always a freshly cleared area into which they can spread.

Although the circumstances are somewhat contrived, the significance of this is that the system contains all of the ingredients for the most basic form of evolution by natural selection in a simple



(a) $t = 9060$ **(b)** $t = 9949$

Figure 5.10: Two snapshots of the system resulting from equations 5.28–5.29, integrated on a surface of size 2×2 with the parameters described in the text. The colours are adjusted so that the parasitic autocatalyst D, which forms into tails behind some of the spots, appears as a darker shade of grey than the primary autocatalyst B. A group of spots with tails can be seen on the left hand side of Plot (a), whereas in Plot (b) some tail-less spots can be seen in the same place, due to tails being lost during reproduction. The spots with tails move constantly away from their tails at a rate of approximately 4×10^{-4} distance units per time unit, which results in their colonising the empty part of the space more rapidly than the tail-less spots. However, with this choice of parameters, spots with tails cannot invade areas colonised by tail-less spots, and the spots with tails are eventually crowded out and become extinct.

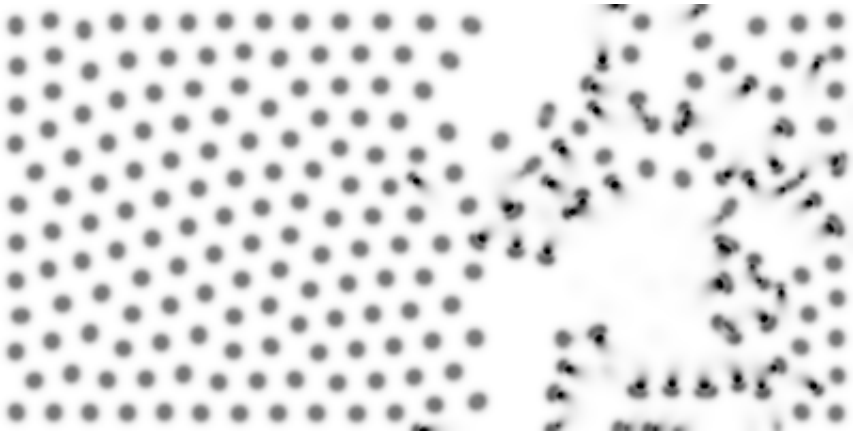


Figure 5.11: A snapshot from the same system as shown in Figure 5.10, except that random areas in the right-hand side of the figure are cleared by an occasional externally-induced cataclysm. The spots with tails' ability to colonise these cleared areas more rapidly than the tail-less spots enables them to persist in this region, whereas they are out-competed by tail-less spots on the more stable left-hand side.

dissipative system that lacks any information-carrying macromolecules. There is heredity (albeit highly limited, with only two different types of individual) with variation, in the sense that the one of the two types of individual can mutate into the other, and there is selection, in the sense that one of the two types of individual is better fit to a given niche than the other.

5.6 Discussion

In this chapter I have explored the notion of *individuated, precarious dissipative structures*, arguing that the study of such systems can enhance our understanding of biology. The idea is to study this type of system in the manner of a biologist, by observing them in their natural habitat, by studying their behaviour, and by trying to see how their anatomy functions (McGregor & Virgo, 2009), allowing us to understand the mechanisms behind their life-like features in a system that lacks much of the contingent complexity of Earth biology. I have made a start on such a project in this chapter.

I have argued that the notion of individuated dissipative structures has much in common with the idea of autopoiesis, although the two approaches differ in emphasis if not in content. On the one hand, the idea of an individuated dissipative structure is more explicitly based in physical reality and, one hopes, is less open to multiple interpretations. On the other hand I have not given a formal definition of individuation, and it may be that one cannot be given. However, this fuzziness can be seen a strength rather than a weakness, since individuation does not appear to be an all-or-nothing property in the living world.

The experimental results in Section 5.4 show that under input or nutrient limitation — conditions which are common in naturally arising situations — the system's parameters can become "tuned" into the region where individuation occurs. This occurs in both the basic Gray-Scott type system and in reaction-diffusion systems with more complex reaction networks. If this result carries over to yet more complicated types of dissipative system (as I will argue it will) then this suggests a possible role of negative feedback in the origins of biological individuation.

5.6.1 What causes individuation as a response to negative feedback?

The stability of spot patterns in the Gray-Scott system can doubtless be understood in formal mathematical terms, but in order to relate these results to the full complexity of a living organism it is better to aim for a more conceptual level of explanation.

Individuation can be thought of as arising from a balance between economies of scale and transport costs. To make this idea clear, let us consider an economic metaphor. Suppose that we are a company planning to manufacture a product. The manufacturing could be done in a distributed manner, with many small workshops spread throughout the country (we assume, unrealistically, that this country contains a more-or-less even distribution of the raw materials needed to make the product, and customers willing to buy it). However, because of the nature of the manufacturing process, we find that the cost of manufacture reduces as we increase the size of the workshops. At the other extreme we could build one giant factory in the centre of the country — but then the cost

of transporting the raw materials to the factory, and finished products to the customers, becomes significant.

In many cases the most profitable course of action would be to build a series of factories of intermediate size, evenly distributed across the country (if the distribution of resources or customers were uneven then it would make sense to put the factories close to them in order to further minimise transport costs). The size of each factory might be expected to reach an optimum, where either expansion or scaling back would reduce profits.

Although reaction-diffusion spots are not trying to maximise any analogue of profit, the explanation for the regulation of their size is similar. The economies of scale in the factory metaphor are akin to the autocatalytic nature of the Gray-Scott reaction: the more autocatalyst is present in a spot, the faster the reaction proceeds, as long as a similar amount of food is present. However, a faster reaction uses up food at a faster rate, so it must be then transported via diffusion from a wider area. This is analogous to the transport costs in the factory example, and the balance between these two factors appears to be a main factor underlying the individuation process.

With this picture in mind we can begin to understand why a system-wide negative feedback in the form of nutrient or input limitation might increase the tendency for individuated systems to become stable. Individuation doesn't occur if the advantage in expansion always outweighs any transport costs because in this case the autocatalyst will spread out indefinitely. A system-wide negative feedback prevents this indefinite expansion, cutting back the advantage of an increase in size for each individual, making the situation where the advantage of a larger size is balanced by transport costs for a given individual more likely.

One advantage of this more general way of understanding the phenomenon is that it suggests that individuation under negative feedback might be an equally general phenomenon. Transport in reaction-diffusion systems is via diffusion, whereas the transport in the factory example might be by road and rail, yet it seems at least plausible that this phenomenon could take place in both examples. This suggests that the phenomenon of individuation under system-wide negative feedback does not depend on the specific dynamics of reaction-diffusion systems and might apply quite generally to many different types of non-equilibrium system. Demonstrating this conclusively is a task for future work, however.

In modern organisms, particularly complex multicellular organisms, individuation doesn't seem to occur purely due to the balance between economies of scale and transport costs, and body size is usually partially determined genetically, although of course the availability of energy usually also plays a role. I would speculate, however, that individuation could have occurred by this method originally, which would have provided a unit on which natural selection could act, leading eventually to a more genetically controlled form of individuation.

In the case of genetically controlled individuation the explanation in terms of cost again plays a role, this time in terms of evolutionary fitness. It seems reasonable to think that grazing animals, for instance, evolve a body size which balances economies of scale in their metabolism (such as the reduced rate of heat loss for a large animal) against the need to move around in order to find fresh pastures. The interplay between individuation and genetic evolution is an interesting

question for future work.

5.6.2 Individuation and the Origins of Life

It seems reasonable to think that system-wide negative feedback, in the form of both nutrient and input limitation, would have been present on the early Earth, applying both to the planet as a whole and to more local systems such as lakes or hydrothermal vents. Theories of the origin of life differ in what they propose as the energy supply for the first ecosystems: it could have come from geothermal sources, similar to the chemical energy that powers today's "black smoker" ecosystems (Wächtershäuser, 1988), or it could have been in the form of molecules formed by photochemical reactions in the atmosphere (Oparin, 1952), or from molecules generated in lightning strikes or meteorite impacts. In all of these cases, however, the energy would have been in the form of relatively stable molecules. One would expect the most stable of them to build up to high concentrations in seas and oceans in the absence of any process to consume them, leading to a global situation similar to the input limitation setup used in Section 5.4.1. Could the phenomenon of individuation under negative feedback then have played a role in the origin of cellular life?

The picture presented here is somewhat speculative but worth exploring. It seems plausible to consider a scenario where a complex autocatalytic network that was initially spread more-or-less homogeneously throughout some part of the pre-biotic Earth. The work of Kauffman (e.g. 2000) shows that large, complex autocatalytic sets are quite likely to form in non-equilibrium chemical systems. It doesn't matter whether you picture this network as composed primarily of RNA, or of proteins and short chains of nucleotides, or as primarily of simpler or more exotic molecules. It could have existed globally in the oceans or been concentrated in shallow pools, or near hydrothermal vents. The point is that, to begin with, there would have been a lot of available energy in the form of stable chemicals that had built up over time and so the concentration of molecules comprising the autocatalytic set would increase.

Due to a limit in either the availability of nutrients or the input rate of energy, this could easily result in a situation where a homogeneous concentration of the autocatalytic set could no longer be supported. The results of this chapter suggest that one possible response to this could be the emergence of localised regions of high concentration separated by regions of low concentration. These localised regions would have been autocatalytic and able to reproduce, and could even have exhibited non-trivial spatial organisation within themselves, as the results of Section 5.4.2 show. It is these localised structures which I suggest might have gone on to become life's ancestors.

Of course, modern cells have two important features which I have not yet discussed in this picture: firstly they contain nucleic acids which store information, allowing for heredity and thus evolution by natural selection, and secondly, modern cells are surrounded by complicated membranes.

There are at least two ways in which the origin of nucleotides could fit into this picture. One possibility is that template replicators such as RNA could have existed as part of the initial spatially homogeneous autocatalytic set (although the complexity of such molecules and the very specific conditions required in order for them to be replicated count against this possibility (Cairns-Smith,

1985)). In this case it would be unsurprising that such molecules would end up as part of the individuated regions, as these would consist of the same chemicals that comprised the initial homogeneous mixture, and this would have given the first individuated proto-organisms a limited capacity for heredity. Views on the origin of life are often divided into *replicator-first* and *metabolism-first* theories; perhaps this constitutes a third option, since the information-carrying molecules would pre-date the emergence of individuals but, like modern nucleotides, would be unable to reproduce except as components of a large, complex autocatalytic set.

A second, perhaps somewhat more plausible, possibility is that the individuation step could have occurred before the emergence of specific information-carrying molecules, which would make this a metabolism-first theory. The results in Section 5.5 demonstrate that limited forms of heredity are possible in simple dissipative systems without the need for complex molecules. More complex heredity could have arisen through the occasional creation of new molecules that modify the topology of the autocatalytic set, as demonstrated in the model of Fernando and Rowe (2008). It would be interesting to implement such a scenario in a simulation model, as one might expect it to result in much more complex structures than have been seen in this chapter.

Similarly, the origins of the cell membrane could have occurred after individuation, since the possession of a membrane would confer an evolutionary advantage over other individuals that lacked one. Another alternative is that the membranes arose contemporaneously with the first individuals: if we imagine that the initial autocatalytic set happened to include some lipid-like molecules then these would have spontaneously formed into membranes. Initially these membranes would not necessarily have surrounded regions where the concentration of the autocatalytic set was high, but during the onset of individuation, any regions of autocatalysis that did happen to be surrounded by a membrane may have had an advantage, since the membrane would have kept their concentration high. Perhaps this type of process could lead to the formation of membrane-bound proto-cells. Further simulation or experimental work is required in order to show how plausible this idea is.

It is interesting to note that, in the experiment of Section 5.4.2, if the initial conditions are a square containing the components of the autocatalytic set in low concentrations (e.g. $a = 0.5$, $u = v = 0.05$) individuation will still occur, but after some time the availability of food in the system drops to a point where such an initial configuration would be unable to persist. After 10^5 time units, when the system is in a state similar to that shown in Figure 5.7b, regions of autocatalyst can only persist if they are configured in a specific way — but such specifically configured regions exist in the system because they arose from less ordered structures at earlier points in time, when the available energy was higher (Figure 5.9). The presence of the negative feedback leads to a kind of boot-strapping, whereby the structures in the system can become more and more specifically organised over time, without the availability of energy dropping so fast that none of them are able to persist.

This adds plausibility to the notion that the early Earth's environment could have been such that the spontaneous generation of life was possible, but that the presence of life itself has changed the conditions so that this is no longer the case. More specifically, in the scenario presented here,

the available chemical energy in the environments where life first arose was much higher than it is anywhere on the modern Earth, allowing much less structured organisms or organism-like structures to persist than is possible today. As more efficient metabolisms arose the available energy reduced, so that the less structured versions could no longer persist.

Although speculative, I believe these ideas are worth pursuing because they provide a concrete mechanism by which first organisms or proto-organisms could have arisen, which does not rely on any specific assumptions about the chemical environment of the early Earth, beyond the high availability of chemical energy. Although somewhat abstract this theory is amenable to empirical testing.

5.6.3 Future Work

In this chapter I have introduced a methodology of studying simple individuated dissipative structures in order to understand how processes such as individuation, reproduction and adaptivity might work in organisms, which are effectively much more complex individuated dissipative structures. I have made a start on such a project by examining these structures in perhaps the simplest type of system that can support them.

There are many questions which this research has raised, however: does the phenomenon of individuation under negative feedback occur with other types of dissipative structure than reaction-diffusion patterns? What role could the generation of membranes play in this kind of individuation process? How might such structures behave with much more complex reaction networks? Can individuated dissipative structures be found that exhibit a greater degree of heredity than the examples in Section 5.5, leading to evolution by natural selection?

One approach to answering these questions is to proceed with more complex simulation models. The addition of membranes or genetic material to this type of model would require more advanced methods of simulation but could produce impressive results. It would also be worth investigating whether the presence of negative feedback encourages the formation of individuated structures such as vortices or convection cells in fluid dynamic systems, since these incorporate quite different transport processes than diffusion, so such a result would add weight to the idea that individuation under negative feedback is a general phenomenon.

Aside from more advanced simulations, another possibility would be to apply this methodology to physical experiments. Perhaps the addition of system-wide negative feedback would encourage the formation of protocell-like structures in "wet A-Life" experiments. Indeed, if such experiments are successful, the use of negative feedback in this way could prove a useful trick by which the formation of complex individuated dissipative structures could be encouraged.

5.7 Conclusion

In this chapter I have argued for the study of individuated, precarious dissipative structures as a close analogy to living organisms. Whether or not such structures can properly be considered autopoietic, the relationship between them and living systems is certainly worthy of study.

Perhaps the most important contribution of this chapter is the idea that the origin of organisms could have taken place via a splitting-up of an initially homogeneous autocatalytic substrate, rather than being built up around initially naked replicating molecules, or else descending from a single individual that arose by chance. The experimental results of this chapter provide a general mechanism by which this splitting-up could happen, and suggest that the global ecological context in which organisms exist is not only important to modern life but also played a fundamental role in its origins.

Chapter 6

More Complex Feedback Conditions: A Model Inspired by Pask's Ear

6.1 Introduction

Most of this thesis has been concerned with the response of physical systems to a particular type of non-equilibrium boundary condition characterised by negative feedback: the system transports some quantity from a reservoir of low temper (e.g. high temperature or high chemical potential) to one of high temper, and the boundary conditions are such that the higher the rate of flow, the lower the difference between the two tempers. In this chapter I will consider the response of systems to boundary conditions where the relationship between the system's behaviour and the externally applied temper gradient is more complex.

In particular I wish to consider situations which might be termed *positive reinforcement boundary conditions*, whereby the system is "rewarded" with an increase in the flow (or the gradient) for behaving in a particular pre-specified way. I argue that under some circumstances, some systems will become organised in such a way as to cause an increase in the reward signal.

A striking example of such a set-up can be found in the electrochemical experiments of Gordon Pask (1958, 1960, 1961). Although the details of these experiments are somewhat obscure, it seems that by "rewarding" an electrochemical system in such a way, Pask was able to "grow" a complex mechanism that was capable of distinguishing between two sounds.

This chapter is based on the text of (Virgo & Harvey, 2008a). The research presented is somewhat preliminary, but I believe its presence in this thesis is justified because it concerns a mechanism by which non-equilibrium physical systems can spontaneously develop into complex functional structures, which is relevant to this thesis' theme of investigating the ways in which life or life-like phenomena can arise from purely physical systems.

In the late 1950s Pask (1958, 1960, 1961) was able to construct a device whereby a complex functional structure would emerge within a physical medium (a solution of ferrous sulphate), simply by increasing its supply of electrical current according to its performance at a given task. The

task Pask set the device was to react to sound. The structure it produced resembled an ear, with an array of resonating metal threads which was able to distinguish between two different sounds. He was also able to get the same device to respond to changes in a magnetic field.

Unfortunately Pask did not record all the details of his setup and the experiment has never been repeated. Nothing quite like it has been achieved before or since. However, if the result can be repeated and generalised its potential would be enormous. It would mean that complex structures with functional components could be grown *in situ* without the need for an evolutionary population. For instance, one could imagine Pask's ear being used as a sensor in an adaptive robot. If the robot found itself in a situation where magnetic fields were relevant it could find itself adapting its ear to respond to them. If Pask's result is sufficiently general one could go further and imagine a neural controller that is able to grow adaptively, creating new neural structures in response to novel challenges. It is clear from (Pask, 1960) that Pask thought along similar lines.

The mechanism I propose to explain Pask's result is quite simple. The idea is that the structure which initially grows in the solution as a result of the electrical current continually undergoes small fluctuations, with some new filaments growing and others being dissolved. If it happens that such a change in configuration results in an increase in the reward signal (i.e. more electric current) then these fluctuations will tend to become canalised, with the thin new threads becoming thicker and thus less likely to fluctuate out of existence in the future. In this way the system performs a kind of hill-climbing in its space of possible configurations, with the fluctuations allowing it to "sample" nearby configurations and the canalisation process allowing "successful" adaptations (i.e. those that increase the reward signal) to be maintained. I call this type of phenomenon an *adaptive growth process*.

This idea of adaptive growth resulting from the canalisation of small fluctuations is comparable to the process of evolution by natural selection. The main difference is that while natural selection requires a population of individuals, each with the capacity for reproduction with heritable variation, adaptive growth requires only a single system with the capacity for fixation of fluctuations. Adaptive growth can produce complex, functional structures in a similar manner to natural selection. As an algorithm it is less efficient, sampling variations sequentially rather than in parallel and lacking any analogue of recombination, but it can occur in simpler physical systems that need not be biological in nature.

I back up this idea of adaptive growth with a simplified computational model of such a process which, although it does not perform as impressive a task as Pask's ear, works in what I believe is a similar way. The system I will present resembles a model of water droplets running down a smooth inclined plane. Each droplet leaves a trail of moisture, which later droplets are more likely to follow. The task we wish this system to perform is to produce a pattern of trails that funnel the water droplets toward a particular "target" area at the bottom of the plane.

This has some resemblance to an ant colony model, whereby ants leave trails of pheromones that later ants can follow. However, the important difference is that we do not selectively reinforce the trail of a successful water droplet. Instead the reward for reaching the target is applied to the system as a whole, in the form of an increase in the rate at which droplets enter the system. This

stems from a difference in focus between the water drop model and ant or swarm based models. Our model is intended to illustrate a plausible mechanism behind the success of Pask's experiment, in which individual threads could not have been reinforced. There is no mechanism by which this could be achieved in Pask's electrochemical system, but more importantly, in Pask's system the solution was formed by a network of many interacting threads, so the identification of a single "successful" thread to reward is impossible.

This chapter also explores the possible biological implications of this kind of emergent adaptive structure.

6.1.1 Pask's experiments

The electrochemical experiment in which the ear was produced is mentioned in (Pask, 1958, 1960, 1961) but is always presented as an example to back up a philosophical point rather than as an experimental result in itself and it is difficult to decipher the exact experimental conditions which were used. There is also an eyewitness account given by Stafford Beer, described by Bird and Di Paolo (2008), although this was given many years after the event and does not give a complete description. Further descriptions of Pask's experiment and the history behind it can also be found in (Bird & Di Paolo, 2008) and (Cariani, 1993). There is a photograph of the resulting "ear" structure in (Pask, 1958), which is reproduced in (Cariani, 1993).

Although the details are obscure the basic idea is clear. Pask was experimenting with passing an electric current through a solution of ferrous sulphate. This causes a thin metal wire to be deposited along the path of maximum current. Since these threads have a very low resistance they affect the electric field surrounding them and thus the growth of further threads. Such systems were a popular way to model neural growth at the time, since digital computers did not yet have the required capacity.

By varying the current flow through an array of electrodes Pask was able to affect the growth of the threads. For instance, when activating one negative and two positive electrodes a wire forms that starts at the negative electrode and branches toward the positive ones. If one of the positive electrodes is switched off the wire moves so that both branches point towards the remaining positive electrode, with the branching point remaining stable. If part of the wire is then removed the gap gradually moves toward the positive electrode, with the wire ahead of the gap being dissolved by the acidity of the solution but the wire behind growing back in its place. The branching pattern is reproduced by the new wire. Details of this can be found in (Pask, 1960).

Pask then took his electrochemical system and subjected it to sound. At this point the details become less clear but the system was rewarded in some way by an increase in available current whenever it responded to the sound, presumably by forming connections between a particular set of the electrodes. After about half a day a structure was formed which was able to perform this task, and Pask then went on to train it to distinguish between two tones, one about an octave above the other.

The interesting thing is that this structure is fairly complex, with functionally differentiated parts that are not specified by the experimenter. In Pask's words, "the ear, by the way, looks rather

like an ear. It is a gap in the thread structure in which you have fibrils which resonate with the excitation frequency" (Pask, 1960). This solution is truly system-level. It is not determined simply by the fittest individual thread but by a combination of many threads playing several different roles. Some act as vibrating fibrils while others are part of the supporting structure. Presumably the fibrils can become further specialised to vibrate when exposed to sound in a particular frequency range. This spontaneous division of labour is not a feature of most learning algorithms.

6.1.2 Pask's Ear as a Dissipative Structure

The system of metallic threads that forms in Pask's ear is a dissipative structure. That is to say, it is a kind of structure that exists as the result of an externally imposed flow of a conserved quantity (in this case electrons) through a system. The threads can only form and persist when an electrical current is passed through the medium, and if the current stops the acidity of the ferrous sulphate solution will gradually dissolve them. The structure maintains its form through a balance of creative and destructive processes.

This seems to me to lie at the heart of its operation. The system is subject to continual fluctuations, in which filaments are lengthened or shortened, or new ones grown and old ones destroyed. But the threads are in competition with each other: a given amount of electrical current can only support a certain amount of filament, so any new structure which increases the amount of metal thread can only become stable and be maintained if it causes a sufficient increase in the current flow.

6.1.3 Adaptive Growth Processes

I will use the term *adaptive growth process* to refer to a system which operates in this way, where fluctuating structures compete for some resource and those that contribute towards maintaining an increased supply of that resource tend to be more stable over time.

This definition may be refined at a later date but some general preconditions for an adaptive growth process are that there is a substrate and a resource whose availability depends in some way on the state of the system. The substrate is such that without any inflow of the resource it will decay to a homogeneous state, but that an inflow of resource enables structures to persist in the system. The dynamics of the system must be such that these structures compete for the resource and are continually subject to fluctuations, and there must be some way in which these fluctuations can be canalised when the amount of resource is increased, resulting in structures that contribute to an increased resource flow becoming more likely to persist over time than those that do not.

Of course, this does not happen in all circumstances in all possible substrates. This research is a first step towards identifying the specific requirements for an adaptive growth process to take place. Some insights gained from the model developed herein can be found in the discussion section.

One interesting feature of adaptive growth processes is that they can in some circumstances result in an increase in complexity: Pask's ear presumably starts with a fairly simple network of threads and ends up with a relatively complex arrangement of resonating fibres. This increase in

complexity occurs when an increase in resource flow can be achieved by a more complex structure. Since an increase in complexity will usually require an increase in resource use (to maintain a larger amount of metal thread, for instance) the structure will not generally become more complex than it needs to be.

6.1.4 Relationship to Reinforcement Learning and the Credit Assignment Problem

Reinforcement learning is a general term for a learning algorithm which can learn a task (usually a classification task) by being given a "reward signal" whenever it behaves in the desired way. Pask's ear can be seen as an example of this, with the supply of electric current acting as the reward signal. In general the supply of resource to an adaptive growth process acts as a reward signal.

From a reinforcement learning point of view, an interesting feature of Pask's ear and adaptive growth processes in general is that they avoid the so-called Credit Assignment problem. This is simply the problem of deciding which part of the system to reward for a successful behaviour. Since behaviour results from the dynamics of the system as a whole it can be hard or impossible to know which parts contribute to a given behaviour and which detract from it. Pask's ear solves this problem in a very simple way: the reward is applied to the whole system in the form of an increase in resource availability. Since the system's parts are in competition with each other it is only the ones which contribute to this increased resource availability that will remain stable in the long run.

6.1.5 Relationship to Evolution by Natural Selection

Our proposed mechanism for adaptive growth bears some resemblance to the process of natural selection. In both cases a system is subject to small random variations which are more likely to persist if they increase a certain target property of the system (its fitness to its environment in the case of natural selection, or the rate of resource input from the environment in the case of an adaptive growth process). In both cases it is the behaviour of the system as a whole, rather than its individual components, that determines which variations are selected.

The primary difference is that in natural selection the selection takes place between a number of similar systems, whereas in an adaptive growth process there is only one system and the selection occurs over time. Or, if one prefers to think of a system like Pask's ear as being a population of threads then the selection takes place according to a population-level property rather than to individual fitness; but again there is only one population.

Both processes could be simultaneously relevant in living systems. Adaptive growth processes within individual organisms could be honed by natural selection operating on a longer time scale, for instance.

6.1.6 Implications for Biological Development

All the structures that occur within living organisms are also maintained by a flow of energy and/or matter (ultimately provided by the organism's food) and will decay if that flow ceases.

Perhaps some of the complexity of biological structures is formed and maintained by what I have termed adaptive growth processes. If this is the case then research into these principles could vastly increase our understanding of biological growth and development processes. Pask saw this potential, writing in (Pask, 1960) that "the natural history of this network [of metallic threads] presents an over-all appearance akin to that of a developing embryo or that of certain ecological systems."

It seems quite possible to me that nature would take advantage of this "design for free" whenever possible. For instance, an organism's genes would not need to specify every component of a particular organ, but simply to arrange for circumstances in which the organ will emerge via adaptive growth in response to stimuli from the environment and a modulated supply of energy. Of course there will also be a strong element of genetic design in nature, but an element of adaptive growth could perhaps explain why organs can atrophy or fail to develop properly when not used. Conceivably it could also be a factor in the enormous plasticity of biological development (and, in particular, the brain's ability to re-organise its physical structure in response to the need to perform a particular task, e.g. Maguire et al. (2000)).

Another possibility is that adaptive growth in this form is possible during an organism's development but is less efficient than more genetically determined growth. In this case one might still expect it to play a role in evolution via the Baldwin effect.

One could also imagine adaptive growth processes occurring at a larger scale in the development of an ecosystem. If this turns out to be the case, it might suggest a mechanism by which a maximisation of entropy production (in the sense defined in Chapters 3 and 4) might occur in ecosystems.

6.2 An Adaptive Growth Process in a Model

Rather than trying to simulate the complex physics of electrochemical deposition we have developed a very simple computational model in which we have tried to capture the conditions required for adaptive growth to take place: there is a balance between creative and destructive processes and structures compete for a resource whose supply is globally changed in response to the system's performance at a task.

In Pask's experiments metallic threads were deposited by a flow of electrons between electrodes. We have abstracted this to something which resembles a system of moisture trails laid down by water droplets which move down a pane of glass. We shall use the metaphor of droplets and moisture trails rather than electrons and filaments to describe our model in order to emphasise that it is not meant to be a physical simulation of Pask's electrochemical experiment. However, we consider the moisture trails and the rate of arrival of droplets to be analogous to the metallic threads in Pask's ear and to the electric current respectively.

It is important to note that we do not selectively reinforce the trail left by a successful water droplet. Instead the 'reward' is applied to the system as a whole in the form of an increase in the rate at which droplets arrive.

6.2.1 Specification of the Model

The system is divided into a grid 50 cells wide by 500 high. Each cell contains a floating point number representing the amount of moisture present. These are initialised to zero.

The following two processes are then repeated a large number of times: first a water droplet enters at the top of the grid and moves towards the bottom, tending to follow any moisture trails that are present according to the rules given below, and laying down a trail of its own. This is followed by a period of time during which all moisture trails decay due to evaporation. This period of time, and hence the amount of decay that takes place, is adjusted according to a simple reward function. Note that the two time scales are separated: the droplets are assumed to travel instantaneously as far as the evaporation time scale is concerned.

Each droplet enters the system at a uniformly random position on the top row. Its path to the bottom row is determined by the following algorithm:

1. Look at the amount of moisture in the cell directly below and the cells to either side of it, three cells in total (the edges wrap around).
2. Add the value $\delta = 0.1$ to each of these three numbers (this is to prevent very weak trails from having a strong effect) and normalise them to get a probability distribution.
3. Move to one of the three cells below according to the computed probability distribution
4. Add the value 1.0 to the amount of moisture in the previously occupied cell

After a droplet has reached the bottom there is a period of simulated time in which the moisture in each cell decays exponentially. This amounts to multiplying each cell in the grid by the same number between zero and one. The length of this time, and thus the amount of decay, depends on the reward function, which is described below. In this way the rate of build up of trails can be controlled by modulating the time that elapses between each droplet entering the system.

In addition to this, before each drop enters the system the value $d = 0.01$ is subtracted from the amount of moisture in each cell (we set the value to zero if it becomes negative). This has a proportionally greater effect on weaker trails, and means that they effectively decay more rapidly when the rate at which droplets enter the system is high.

6.2.2 The Reward Function

The task that we set our system is substantially simpler than Pask's. We want the droplets to arrive at the bottom within a specific range of columns, 19 to 31 inclusive (twelve columns out of the 50 in the system). When a droplet hits the target we increase the rate at which droplets enter the system. Since each droplet lays down the same amount of moisture (500 units in total) the rate at which droplets arrive is proportional to the rate at which water is added to the system in the form of moisture trails, and therefore limits the total strength of trails that can be maintained in the system.

The details of the scheme used for the presented results are as follows:

1. Let the score for iteration i be $S_i = 1$ if the droplet arrives at the bottom of the grid within the target interval, 0 if it misses.

2. This value is smoothed out in time slightly using a leaky integrator: Let the reward value $R_i = R_{i-1} + (S_i - R_{i-1})/\lambda$. We give the parameter λ the value 2.0 and let $R_0 = 0$.

3. Droplets are assumed to arrive at a higher rate when R is high than when it is low. This is represented by multiplying each moisture value by $1 - 1/(495R_i + 5)$ after the droplet has arrived at the bottom. This is to represent the constant decay of moisture during the variable time period between droplets arriving.

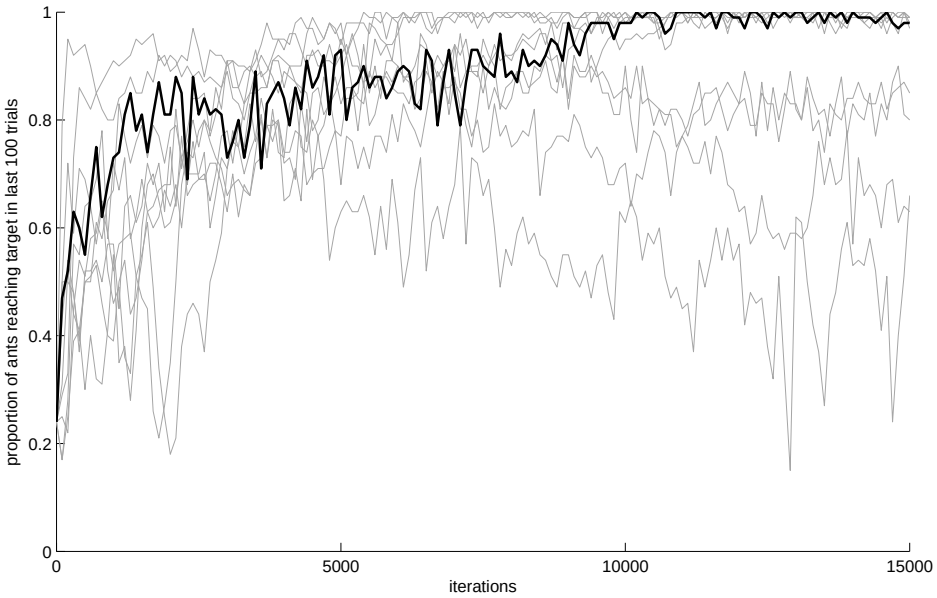


Figure 6.1: Increase in accuracy over time for ten independent runs of the model. The black line corresponds to the run shown in more detail in figure 6.2. Each data point represents the proportion of droplets which hit the target over a period in which 100 droplets are released. The expected probability for a drop to hit the target in the absence of any moisture trails is 0.24, so the first data point is set to this value for each run.

6.2.3 Experimental Results

Figure 6.1 shows the proportion of droplets hitting the target over time for ten independent runs of the experiment. Four of these systems do not converge on a good solution before the end of the run at iteration 15000 but six of them perform well, converging to a stable state in which very few drops miss the target.

Figure 6.2 shows the positions of the trails over time for one such run. One can see that the system converges fairly rapidly to a state in which there is a fairly strong trail leading to the target, which has a catchment area that catches almost all the droplets entering the system. However, all trails are rewarded for this, not just the ones that lead to the target. This allows several "parasite" trails to persist, which benefit from the droplets reaching the target but do not contribute towards it. These are less stable than the one which reaches the target and are eventually out-competed by it.

The four systems which do not converge to a good solution have stronger, more established parasite trails. Since more established trails fluctuate more slowly it can take a very long time for

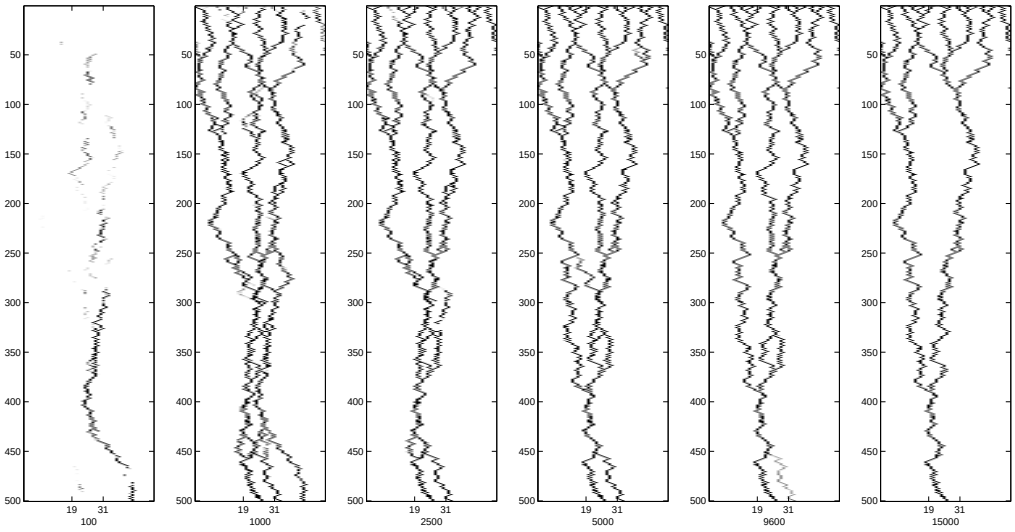


Figure 6.2: Snapshots of the moisture trails after 100, 1000, 2500, 5000, 9600 and 15000 droplets have passed through the system. The target is the marked interval between columns 19 and 31 at the bottom of the grid. After 100 iterations there is only a weak trail which does not lead to the target. By iteration 1000 a stable trail to the target has been formed which fans out at the top into a large catchment area, but it also supports a number of 'parasite' trails which do not hit the target. These gradually disappear, with the last fading out at around iteration 9600. After this almost all the droplets hit the target (see figure 6.1). The system then changes very little until the end of the run at iteration 15000.

these to decay. Note that although these replicates have not converged on a perfect solution, all but one of them consistently do better than the 25% that would be expected in the absence of any reinforcement effect: they have attained solutions which work but are not perfect. The possibility of getting stuck on 'local' optima is perhaps another thing that adaptive processes have in common with evolution by natural selection.

6.3 Discussion

It is important to be clear about the relationship of our model to the subjects under discussion, namely Pask's experiment and adaptive growth processes. Our claim is that our model and Pask's device operate according to a common principle, the adaptive growth mechanism that we have described. Our model is intended as a simple instantiation of an adaptive growth process rather than as a direct model of Pask's ear, although loosely speaking the moisture trails can be seen as a metaphor for the metallic threads in Pask's ear, with the rate of input of water droplets taking the place of electric current.

6.3.1 Comparison to Ant-Colony Methods

Our model is not intended to compete with ant colony optimisation methods as it does not have the same purpose, but since it bares some similarity to them it is worth discussing the how our system differs and the reasons for taking our approach.

It might appear that our system has several disadvantages compared to a more traditional ant-based system in which a successful trail is rewarded. We can see from figure 6.1 that convergence to a solution is slow and not guaranteed to occur. However, the purpose of our model is explanatory. The threads in Pask's experiments were not selectively reinforced. Instead he rewarded the whole system with an increase in current flow, and our claim is that our model captures the way in which this directed the system's growth towards a solution. Moreover, the solution found by Pask's electrochemical device does not consist of a single thread; it is a complex solution which requires the co-operation of multiple threads performing a variety of tasks. Rewarding "successful" threads in this context would not make sense, since it is not in general possible to determine which threads are contributing to the solution and which detract from it (see the discussion of the credit assignment problem above).

In our system the task is not to create a single path leading from the top of the grid to the target area, but to create a network of threads which funnels droplets from all positions on the top of the grid towards the target area. There is some similarity between this and the system-level nature of the solution found by Pask's device. Our hope is that with a better understanding of adaptive growth processes it will be possible to design systems, either *in silico* like our model or in physical substrates, which can solve more complex tasks, forming solutions which equal or surpass the sophistication that Pask was able to achieve.

6.3.2 Implications for Adaptive Growth Processes

It seems reasonable to call the growth process in our model adaptive because it shares with Pask's ear the property that structures which contribute towards performing the task (and thus increasing resource availability) are more stable and out-compete structures which do not help perform the task. Our computational experiment thus demonstrates that adaptive growth is a general phenomenon, rather than something which only occurs in the specific electrochemical environment of Pask's experiment.

However, adaptive growth does not occur in all possible substrates. The parameters and reward function of our model had to be chosen from the right ranges in order for the phenomenon to occur. For instance, the subtraction of d from the moisture level in each cell for each drop that enters the system seems to be important for adaptive growth to occur. Without it stable structures that achieve the funnelling task do form, but they spontaneously collapse much more readily, reforming again soon after. Perhaps this is because parasitic structures can grow too easily, diminishing the resource supply and destabilising the whole system. It appears in this case as if the system performs a random walk between more and less stable structures, spending a high proportion of its time in a state containing a stable structure simply because those states change more slowly. But the subtraction of d on each iteration seems to provide a ratchet effect, making transitions from more to less stable structures very unlikely.

Another factor which seems important is that the growth process is fast but that the decay process is slower for more established structures. If the two processes took place on the same time scale then it would be harder for structures to persist on long time scales. The difference in

time scales between new and old structures means that it's very unlikely for the whole structure to unravel by chance. If the resource availability drops it is more likely that a new structure will be dissolved than an old one, giving a trial-and-error quality to the growth process.

It also seems important that the reward function is modulated on a similar time scale to the fluctuations in the structure. This is the reasoning behind applying a leaky integrator to our reward function. A good substrate for an adaptive growth process might therefore be one which exhibits fluctuations over a wide range of time scales.

There is clearly a long way to go before we have a full understanding of why Pask's ear works, and of adaptive growth processes in general. Our example is simple and illustrative but we anticipate that more elegant and successful examples will be found which are capable of adapting to more involved tasks, leading to the possibility of practical application.

6.3.3 Relationship to the Maximum Entropy Production Principle

The boundary conditions applied to the model in this chapter were that the flow of water droplets through the system was a function of the system's behaviour. This can be seen as a generalisation of the negative feedback boundary conditions discussed in earlier chapters. Under those conditions, the driving gradient was constrained to be a decreasing function of its associated flow; in this example the flow is an increasing function of some more complicated aspect of the system's behaviour.

In the model presented here the driving gradient was not explicitly modelled, but it seems reasonable to think of it as remaining fixed while the flow varied. (A more physically realistic model would be required to make this explicit.) In this case an increase in the flow corresponds to an increase in entropy production. Thus, if the maximum entropy production principle generalises successfully beyond negative feedback boundary conditions it should predict a maximisation of the flow under these positive reinforcement boundary conditions. However, it is clear that not all systems act so as to increase flow in this way, and so the relationship between the ideas in this chapter and the MEPP must be investigated in future work.

6.4 Conclusion

I have introduced the notion of an adaptive growth process in order to explain the results of Pask's electrochemical experiment. This allows us to see the enormous potential of his result: in terms of practical applications, adaptive growth processes could be used to produce control systems that can adapt to new tasks and even adjust their level of complexity when necessary. The idea may also be important for our understanding of biology since adaptive growth processes may play a role in the development of organisms and of ecosystems.

I have presented an illustrative computational model in which an adaptive growth process occurs, demonstrating that adaptive growth is a general phenomenon and paving the way for a better understanding of the circumstances under which adaptive growth can occur.

From a broader perspective, adaptive growth processes are a possible behaviour of physical systems under boundary conditions that are modulated by the system's behaviour. This can be

seen as a generalisation of the negative feedback boundary conditions that were the focus of the previous chapters of this thesis.

Like the processes investigated in the previous chapter, the phenomenon of adaptive growth can occur in purely physical structures but also has relevance for biological systems. Adaptive growth thus represents another aspect of the continuity between life and physics that has been the focus of this thesis.

Chapter 7

Conclusion

This thesis has been concerned with the continuity between life and physics, and the study of living systems from a physical point of view. In Chapter 5 I demonstrated that some simple dissipative structures have important features in common with living organisms; chief among them is the phenomenon of individuation, the on-going maintenance of spatially distinct individuals. In these systems we also saw adaptive behaviour, reproduction with limited heredity and the formation of individuals with a complex structure composed of functionally differentiated components.

Chapter 4 was concerned with the evolution of organisms, seen from the most general physical point of view as thermodynamic machines performing chemical work in order to maintain their structures. Chapter 3 was about the application of statistical mechanics to a very general class of physical systems that includes ecosystems; and in Chapter 6 we saw that the process of growing in an adaptive manner in response to an externally imposed "reward" signal, which we normally think of as an exclusive property of living systems, can plausibly happen in some purely physical systems as well.

These results are tied together by a number of themes. In this conclusion chapter I will revisit a number of these themes, showing how the results of this thesis relate to them.

7.1 Functional Boundary Conditions

The main theme connecting all the results of this thesis is the notion of functional boundary conditions, where the thermodynamic gradient (temper difference) that powers a system depends in some way upon the system's overall behaviour. In particular, most of the thesis has concerned the response of ecosystems and complex physical systems to a type of non-equilibrium situation that I call negative feedback boundary conditions, where a temper difference powering the system is a decreasing function of its associated flow.

Chapter 3 concerned a general statistical-mechanical framework for making predictions in such situations; Chapter 4 examined the population dynamics and evolutionary dynamics of simple model organisms under these conditions; in Chapter 5 they were applied to simple physical

systems, resulting in individuation. Chapter 6 dealt with a more general form of functional boundary conditions, where the driving gradient increases when the system behaves in some externally specified way.

7.2 What is an Organism?

I have deliberately steered clear of defining life in this thesis. Differentiating life from non-life on present-day Earth is easy, since all known living organisms share many features in common. These range from general physical properties, such as the maintenance of a structure by increasing the entropy of the environment as described by Schrödinger (1944), to the use of specific molecules, such as ATP as an energy carrier and nucleic acids as genetic material. Presumably these features did not all arise simultaneously, however, and so there must have been a time in the distant past when the line between living and non-living was not so easy to draw. The origins of each of modern life's features must be studied in order for the origins of life to be fully understood, and by prematurely stating a definition of life we would risk biasing our direction of enquiry.

Another reason for not defining life is that, as I have argued with McGregor (McGregor & Virgo, 2009) and in Chapter 5 of this thesis, studying physical systems that share important properties with living systems can be a useful way to understand the structure of life, because it allows us to understand these features of life in isolation from the full complexity of modern organisms. Thus, rather than trying to define life I have intentionally considered broader classes of physical systems that include life as a subset. This allows simpler systems to be used as models.

I have concentrated more on capturing the metabolic than the genetic aspects of modern organisms. The approach I have taken owes much to Maturana and Varela's (1980) theory of autopoiesis, as I will discuss below, although it differs in important respects.

In Chapter 4 I characterised organisms as chemical engines, whose output of chemical work was put to use in maintaining the engines' own structures. This circularity was first pointed out by Schrödinger (1944) and was described by Kauffman (2000) as a "work-constraint cycle". This thermodynamic circularity is also closely related to the organisational circularity of autopoiesis (Moreno & Ruiz-Mirazo, 1999; Ruiz-Mirazo & Moreno, 2000).

In Chapter 5 I characterised organisms as individuated, precarious dissipative structures. These can also be seen as self-maintaining engines, since their structure is maintained by a dynamical balance between processes that are ultimately powered by reducing the thermodynamic disequilibrium of their environment. However, the examples of Chapter 5 show that the construction of such self-maintaining engines can be much simpler than one might at first expect. Indeed the spontaneous formation of structures fitting this description is not uncommon, hurricanes being a natural example.

The concept of individuation as I describe it is closely related to the autopoietic concept of *homeostasis of organisation*: each individual has a physical structure which can recover when perturbed by a small amount, due to a dynamical balance between on-going processes. Larger perturbations can disrupt the structure beyond its ability to recover in this way, leading to the cessation of its maintenance in a processes analogous to death. The possibility of this kind of

disintegration is called precariousness.

The possibility of death, or its analogue, arises because the processes responsible for the maintenance of the structure are all dependent on one another. In a reaction-diffusion spot, for instance, the process of diffusion of food into the spot cannot continue unless the autocatalytic reaction is using up food in the centre of the spot, and likewise the autocatalytic reaction cannot continue without the diffusion transport of food into the spot. In a living vertebrate, the heart cannot pump blood without a supply of oxygen from the lungs while the lungs cannot supply oxygen without a constant flow of blood. In both cases, all the other processes responsible for the structure's maintenance are interlinked in a similar way. This interdependence is at the heart of the autopoietic concept of operational closure, at least as interpreted by Virgo et al. (2009).

All known living organisms share many features beyond being individuated, precarious dissipative structures. Perhaps the most important of these is the capacity for heredity, allowing evolution by natural selection. Exploring true evolution in dissipative structures would require different modelling techniques than the ones used in this thesis, but I have made a start on such a project by showing that individuated dissipative structures with limited heredity can arise even in very simple physical systems.

Maturana and Varela saw autopoiesis as an attempted definition of life. My approach differs in that I have explicitly studied inanimate dissipative structures from this point of view. In addition, Froese and Stewart (in press) argue that the theory of autopoiesis could never have taken physical/thermodynamic constraints into account because of its basis in the cybernetic framework of Ashby (1960). Ashby's work was instrumental in the development of modern dynamical systems theory, but it was a framework in which physics was abstracted away and systems characterised purely in terms of their dynamics. The approach developed in this thesis has the physical aspects of biological organisation at its heart. Finally, according to some interpretations, the definition of autopoiesis hinges on the presence of a bounding membrane, whereas my approach is centred on the construction of a spatially distinct individual, with a membrane being part of one mechanism by which this can be achieved.

7.3 Adaptive Behaviour and Cognition

An important part of Maturana and Varela's work was the idea that the notions of life (in the sense of autopoiesis) and cognition are intimately related. Indeed, one interpretation of their ideas is that "life = cognition" (Stewart, 1992). Perhaps some evidence for this deep relationship can be seen in the behaviour of reaction diffusion spots, which move away from areas of low food concentration, as seen in Chapter 5. This happens simply because they grow faster on the side where the food concentration is highest. Whether this behaviour is best described as "cognition" or simply "adaptive behaviour" is a moot point, but its relationship to the self-maintaining aspect of the spots' organisation is interesting to discuss.

Bourgine and Stewart (2004) and Di Paolo (2005) have proposed hypotheses along the lines that cognition (or adaptive behaviour) is not an automatic consequence of autopoiesis after all, and that life can be defined as something along the lines of "autopoiesis + cognition" (or autopoiesis +

adaptivity in Di Paolo's case). Logically it may be true that autopoiesis does not imply adaptive behaviour. This of course depends on the precise definitions of autopoiesis and of cognition or adaptive behaviour, but it may be possible for a system to exist that is individuated and self-maintaining, but which does not behave in an adaptive manner. However, the results of Chapter 5 show that at least some form of adaptive behaviour is present even in extremely simple non-living individuated dissipative structures. Indeed, in the case of reaction-diffusion spots it seems that they could not remain individuated if they did not behave in this way, since it is their tendency to move away from one another that prevents them from merging together. The adaptiveness of their behaviour seems closely related to the homeostasis of their organisation.

This suggests to me that the boundary between life and non-life cannot be drawn in this way. I hypothesise that cognition, in the sense of behaving in such a way as to remain a viable self-maintaining structure, is common in non-living physical structures. This in turn suggests that the difference between the cognitive capacity of such non-living structures and living organisms is simply one of degree. (It may be a very large difference in degree however: even some single-celled amoebae are capable of gathering grains of material from their environment and building them into a spherical shell known as a test; see, e.g., Ford, 2008.)

Another demonstration of the possibility of adaptive behaviour in non-living systems can be found in Chapter 6. The structures considered in that chapter are not particularly individuated, but grow in an adaptive manner because their structure fluctuates and those fluctuations that result in an increase in energy flow tend to become more permanent structures.

This is quite a distinct mechanism from the process by which reaction-diffusion spots move along a food gradient. Reaction-diffusion spots are probably not capable of exhibiting adaptive growth according to the mechanism presented in Chapter 6, because their structures do not fluctuate, and because they lack the separation of timescales between the creation of new parts of their structure and the decay of old ones. However, the possibility of individuated dissipative structures that do have the capacity for adaptive growth by this mechanism is an interesting one to consider. Presumably such structures would be able to exhibit a much wider range of adaptive behaviours than simple reaction-diffusion structures. In principle it should be possible to demonstrate such structures in a simulation model; this is another task for future work.

7.4 An Hypothesis About The Origins of Life

In Chapters 4 and 5 I proposed and argued for a hypothesis that the free energy density of the environments in which life first arose could have been higher than in any modern environment. The hypothesis also states that this energy density would have reduced as life evolved, because of negative feedback effects between the energy density and life's overall rate of energy use.

In Chapter 4 I argued that such a scenario would have allowed life's early ancestors to survive with much simpler structures and metabolisms than is possible in modern environments. In Chapter 5 I showed that this scenario could potentially explain how individuated structures came into existence in the first place, and made some progress towards showing that structures formed in this way could have been capable of reproduction with heredity.

Putting these together, we end up with a scenario where life arose from a complex autocatalytic network of chemicals that was originally spread throughout space. As these collectively autocatalytic molecules grew in concentration the concentration of the food chemicals which powered their production reduced until eventually a homogeneous concentration of these autocatalytic compounds could no longer be supported. However, small, concentrated regions could still persist, and it was these that went on to become life's ancestors. These could plausibly have had more complicated structures than reaction-diffusion spots. Indeed, if these early individuals existed under anything but the calmest of conditions, something akin to a bounding membrane would probably be required, since reaction-diffusion structures would easily be destroyed by any mixing of the surrounding milleu. Demonstrating the formation of more complex individuated structures in this way is a task for future work. However, early individuals that formed according to such a process would probably have been much simpler than today's organisms, in the sense of being less specifically constructed.

The ability to reproduce is a natural feature of such individuated entities, as demonstrated by the reaction-diffusion spot examples, but in order for them to have been the ancestors of life they must through some mechanism or other have had the capacity for heredity. This opens up the possibility for evolution by natural selection, producing faster and more efficient metabolic structures, which would further reduce the environmental energy availability, according to the results of Chapter 4. This means that, over very long time periods, only individuals with the fastest and most efficient metabolisms, and the structures most resistant to decay (or rather, the organisms with the best trade-offs between all three of these factors) would have survived. I hypothesise that this process led eventually to the very specific metabolic and genetic mechanisms that are found in all modern cells.

Like all hypotheses about the origin of life, this involves a fair amount of conjecture about the state of the Earth at a time for which very little evidence is available. However, this hypothesis differs from many theories about the origins of life in that it is not tied to any particular set of chemical processes. It is a theory about the origins of metabolism and cellular individuality, and is somewhat independent of theories about the origins of genetic material.

If this hypothesis holds up to further scrutiny, its general nature may have interesting implications for astrobiology, since it suggests that the formation of evolving individuated entities may be a natural result of common planetary scenarios, even on bodies where the specific physical and chemical conditions are quite different from those found on Earth. The hypothesis' general nature also makes it suited to investigation by modelling, as we have seen in this thesis. There is plenty of scope for more detailed models to be constructed in the future, with which the plausibility of these ideas could be further tested.

7.5 Ecosystems and the Maximum Entropy Production Principle

In Chapter 3 I presented a statistical mechanical argument that the maximum entropy production principle should apply quite generally to non-equilibrium physical systems. I do not expect this argument to be the last word on the maximum entropy production principle, which is still far

from conclusively proven, but the argument I have presented seems to lend at least some plausibility to the idea that the MEPP might be a universal physical principle akin to the second law of thermodynamics.

According to the argument developed in Chapter 3, the MEPP is a principle for predicting the future behaviour of a system based on knowledge of its boundary conditions and the factors that constrain its behaviour. It states that the best predictions of a system's future state can be made by assuming it produces entropy at the maximum possible rate, subject to constraints formed by one's knowledge of the system's kinetics. This is very similar to the way one calculates an isolated system's thermal equilibrium state by assuming a maximum in its thermodynamic entropy, subject to constraints formed by conservation laws and the presence of adiabatic barriers between parts of the system, etc. The difference is that kinetic constraints are potentially much more complex and hard to take account of than the kinds of constraints that are relevant to equilibrium thermodynamics.

However, it seems that for at least some systems, such as the Earth's atmosphere, any kinetic constraints are effectively irrelevant for predicting certain aspects of their behaviour. Thus the rate of heat transport in the atmosphere can be predicted using the MEPP from the boundary conditions alone. Intuitively, this under-constrainedness of the atmospheric heat transport seems to be related to its complexity: the atmosphere transports heat by a complex network of convective flows which could conceivably be arranged in a huge variety of different ways, but which are in fact configured in quite a specific way. Every possible configuration of the heat flows in the atmosphere would transport heat at a different rate, and the actual observed configuration transports heat at close to the rate that maximises the entropy production. The MaxEnt interpretation of the MEPP effectively says that this is because there are a great many more microstates compatible with this rate of flow than with any other rate.

However, this thesis has left open the question of whether ecosystems are under-constrained in a similar way, which would allow the rate of energy flow through them to be predicted using the MEPP in a similar manner. The evolutionary model in Chapter 4 does not exhibit maximisation of the entropy production, and neither does the application of negative feedback boundary conditions to reaction-diffusion systems in Chapter 5.

One possible explanation for this is that the MEPP is not a genuine physical principle. Although the arguments in Chapter 3 lend some plausibility to the principle its theoretical justification is still an active area of research.

However, if we assume that the MaxEnt interpretation of the MEPP is correct then there are two possibilities. The first is that ecosystems could, at least in some cases, be under-constrained enough for the MEPP to be applied based upon the boundary conditions alone, but that the models in this thesis are not. If this is case one would expect it to be possible to develop models in the future which do maximise the production of entropy.

However, the second possibility is that ecosystems are not, in general, under-constrained enough for the MEPP to be applied to them in the same way that it can be applied in climate science. Perhaps in order to predict the rate of energy flow through an ecosystem one must always

take into account a great many factors, including details of every species present. If this turns out to be the case then the MEPP principle will not be invalidated and may still turn out to be useful in some contexts in ecology, but it will not be possible to use it in the same way as it is used in climate science. These two possibilities can only be distinguished empirically, either using microcosms with carefully controlled boundary conditions, or using appropriate field measurements.

7.6 Future Work

This thesis has served to advance the study of non-equilibrium thermodynamics as applied to living systems and complex systems. In doing so it has raised a number of interesting questions and possibilities for further research. These are discussed in detail in the individual chapters, but it is worth summarising them here.

Chapter 3 concerned the theory behind the maximum entropy production principle. Further advancement of the theory would be greatly enhanced by experimental study. In particular, if maximisation of entropy production was observed in an experiment in which negative feedback boundary conditions are applied to some complex system under a range of different conditions this would provide a tremendous boost to the theory. It would also open up the possibility of investigating experimentally the conditions under which systems are under-constrained enough for the principle to be applied in this way. This type of experiment could also in principle be conducted using biological microcosms, which would provide a way to experimentally test the extent to which the principle can be applied to ecosystems.

In Chapter 4 I developed a simple evolutionary model which included thermodynamic constraints on the organisms' metabolism. As discussed in that chapter, there are many possibilities for extending this model to include additional constraints on the organisms' metabolism, as well as additional processes such as predation.

The results of Chapter 5 included demonstrations of individuation occurring in response to negative feedback boundary conditions, in simple reaction-diffusion systems. I argued that this phenomenon is likely to occur in other types of system where processes other than diffusion and chemical reactions can occur. This claim could be investigated by further simulation work, or by the implementation of system-wide negative feedback in real physical/chemical systems. In particular, the relationship between individuation and the origins of the cell membrane could be investigated by the use of more advanced simulations, in which the production of lipid bilayers is possible. It would also be interesting to continue investigating the degree to which more complex individuated structures can be formed using this technique. I demonstrated that a very limited form of heredity is possible in reaction-diffusion systems, and I would expect that more advanced forms of heredity could be exhibited by future models.

Including complex molecules and processes such as membrane formation and fluid dynamics in thermodynamically realistic models on a large enough scale to exhibit complex individuated entities would be a technical challenge. Perhaps this challenge could be sidestepped by performing physical experiments instead. Such an approach would complement the "wet" A-Life research programme (reviewed by Rasmussen et al., 2004), potentially contributing a technique by which

the spontaneous formation of protocells could be encouraged more easily.

Finally, in Chapter 6 I presented a very simple model of an adaptive growth process. The theory of adaptive growth processes could be further advanced by the development of more physically realistic models and by further empirical investigation. In particular, Pask's experiment has never been replicated, and a repeat of his result would highlight its continuing relevance today.

7.7 Conclusion

This thesis has been concerned with the study of living systems from a physical point of view, and with the search for general principles that govern the dynamics of both living and non-living systems. In particular, the response of living and physical systems to functional boundary conditions has been a central theme.

In particular, this thesis has included a statistical argument in favour of the maximum entropy production principle, which shows some promise as a general way to reason about far-from-equilibrium systems. I have presented the hypothesis that the availability of energy in the environments of early life was higher than in today's environments, and that this played a crucial role in the origins of life. This hypothesis was backed up with two models: a model of the population and evolutionary dynamics of thermodynamically self-maintaining organisms, and a reaction-diffusion model demonstrating the formation of spatially distinct individuals in response to negative feedback.

I argued that the individuation of these reaction-diffusion spots has much in common with the individuation of living organisms, especially when seen from the point of view of the theory of autopoiesis, and demonstrated that simple dissipative structures of this kind are capable of limited heredity. Finally, I proposed a mechanism by which some non-living systems can grow in such a way as to maximise an externally applied energetic reward, and presented a simple model of this process.

This thesis has contributed to our understanding of living systems and their origins form a thermodynamic point of view, and in so doing has opened up several exciting new avenues of research.

References

Aoki, I. (1989). Entropy flow and entropy production in the human body in basal conditions. *Journal of Theoretical Biology*, *141*(1), 11–21.

Aoki, I. (1990). Effects of exercise and chills on entropy production in human body. *Journal of Theoretical Biology*, *145*(3), 421–428.

Aoki, I. (1991). Entropy principle for human development, growth and aging. *Journal of Theoretical Biology*, *150*(2), 215–223.

Ashby, W. R. (1960). *Design for a brain*. New York: Wiley.

Attard, P. (2006). Theory for non-equilibrium statistical mechanics. *Physical Chemistry Chemical Physics*, *8*, 3585–3611.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, *53*, 370–418.

Beer, R. D. (2004). Autopoiesis and cognition in the game of life. *Artificial Life*, *10*(3), 309-326.

Bird, J., & Di Paolo, E. A. (2008). Gordon Pask and his maverick machines. In P. Husbands, M. Wheeler, & O. Holland (Eds.), *The mechanization of mind in history.* MIT Press.

Boltzmann, L. (1886). *The second law of thermodynamics*. (reprinted in Boltzmann, L. Theoretical physics and philosophical problems, S. G. Brush (Trans.), 1974)

Bourgine, P., & Stewart, J. (2004). Autopoiesis and cognition. *Artificial Life*, *10*, 327–345.

Cairns-Smith, A. G. (1985). *Seven clues to the origin of life: A scientific detective story*. Cambridge University Press.

Cantrell, R. S., & Cosner, C. (2003). *Spatial ecology via reaction-diffusion equations*. Wiley InterScience.

Cariani, P. (1993). To evolve an ear: Epistemological implications of Gordon Pask's electrochemical devices. *Systems Research*, *10*(3), 19–33.

Cox, R. T. (1946). Probability, frequency and reasonable expectation. *American Journal of Physics*, *14*(1). (Reprinted in Shafer, G. and Pearl, J. (eds) Readings in Uncertain Reasoning. Morgan Kaufman Publishers, 1990)

Cox, R. T. (1961). *The algebra of probable inference*. Baltimore, MD: John Hopkins University Press.

Danovaro, R., Dell'Anno, A., Pusceddu, A., Gambi, C., Heiner, I., & Kristensen, R. M. (2010). The first metazoa living in permanently anoxic conditions. *BMC Biology*, *8*(30).

Dewar, R. C. (2003). Information theory explanation of the fluctuation theorem, maximum entropy production and self-organized criticality in non-equilibrium stationary states. *Journal of Physics A: Mathematical and General*, *36*(3), 631–641.

Dewar, R. C. (2005a). Maximum entropy production and the fluctuation theorem. *Journal of Physics A: Mathematical and General*, *38*, 371–381.

Dewar, R. C. (2005b). Maximum entropy production and non-equilibrium statistical mechanics. In A. Kleidon & R. D. Lorenz (Eds.), *Non-equilibrium thermodynamics and the produciton of entropy* (pp. 41–53). Berlin: Springer. (at some point the year was written as 2005)

Di Paolo, E. A. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences*, *4*(4), 429–452.

Di Paolo, E. A. (2009). Extended life. *Topoi*, *28*(1), 9–12.

Downing, K., & Zvirinsky, P. (1999). The simulated evolution of biochemical guilds: Reconciling gaia theory and natural selection. *Artificial Life*, *5*(4), 291–318.

Eigen, M., & Schuster, P. (1979). *The hypercycle: A principle of natural self-organisation*. Springer.

Essex, C. (1984). Radiation and the irreversible thermodynamics of climate. *Journal of Atmospheric Sciences*, *41*(12), 1985–1991.

Fath, B. D., Jœgensen, S. E., Patten, B. C., & Straškraba. (2004). Ecosystem growth and development. *BioSystems*, *77*, 213–228.

Fernando, C., & Rowe, J. (2008). The origin of autonomous agents by natural selection. *Biosystems*, *91*(2), 355–373.

Fienberg, S. E. (2006). When did Bayesian inference become "Bayesian"? *Bayesian Analysis*, *1*(1), 1–40.

Ford, B. (2008). *Microscopical substantiation of intelligence in living cells.* in Royal Microscopical Society's inFocus magazine.

Froese, T., & Stewart, J. (in press). Life after Ashby: Ultrastability and the autopoietic foundations of enactive cognitive science. *Cybernetics and Human Knowing*.

Gánti, T. (2003). *The principles of life*. Oxfod University Press.

Gray, P., & Scott, S. K. (1983). Autocatalytic reactions in the isothermal, continuous stirred tank reactor: isolas and other forms of multistability. *Chemical Engineering Science*, *38*(1), 29–43.

Hamilton, W. D. (1964). The genetical evolution of social behaviour. *Journal of Theoretical*

Biology, *7*(1), 1–52.

Jaynes, E. (1957a). Information theory and statistical mechanics. *Phys. Rev.*, *106*(4), 620–630.

Jaynes, E. (1957b). Information theory and statistical mechanics ii. *Phys. Rev.*, *108*(2), 171–190.

Jaynes, E. (1965). Gibbs vs boltzmann entropies. *Am. J. Phys*, *33*(5), 391–398.

Jaynes, E. (1979). Where do we stand on maximum entropy? In R. D. Levine & M. Tribus (Eds.), *The maximum entropy formalism* (p. 15). Cambridge: MIT Press.

Jaynes, E. (1980). The minimum entropy production principle. *Ann. Rev. Phys. Chem.*, *31*, 579–601.

Jaynes, E. (1985). Macroscopic prediction. In H. Haken (Ed.), *Complex systems – operational approaches* (p. 254). Berlin: Springer-Verlag.

Jaynes, E. (1988). The evolution of Carnot's principle. In G. J. Erickson & C. R. Smith (Eds.), *Maximum-entropy and bayesian methods in science and engineering* (p. 267). Dordrecht: Kluwer.

Jaynes, E. (1989). Clearing up mysteries - the original goal. In J. Skilling (Ed.), *Maximum entropy and bayesian methods.* Dordrecht: Kluwer.

Jaynes, E. (1992). The Gibbs paradox. In G. J. Erickson, P. Neudorfer, & C. R. Smith (Eds.), *Maximum entropy and bayesian methods.* Kluwer.

Jaynes, E. (2003). *Probability theory: the logic of science*. Cambridge: Cambridge University Press.

Jonas, H. (1968). Biological foundations of individuality. *International Philosophical Quarterly*, *8*, 231–251.

Kauffman, S. (2000). *Investigations*. New York: Oxford University Press US.

Kerner, B. S., & Osipov, V. V. (1994). *Autosolitons: a new approach to problems of self-organization and turbulence*. Kluwer.

Kleidon, A., & Lorenz, R. D. (2005). Entropy production by earth system processes. In A. Kleidon & R. D. Lorenz (Eds.), *Non-equilibrium thermodynamics and the production of entropy: life, earth, and beyond.* Springer Verlag.

Langton, C. (1989). Introduction: Artificial life. In C. Langton (Ed.), *Artificial life.* Los Alamos, New Mexico: Santa Fe Institute.

Lorenz, E. N. (1960). Generation of available potential energy and the intensity of the general circulation. In R. C. Pfeffer (Ed.), *Dynamics of climate* (pp. 86–92). Pergamon.

Lorenz, R. D., Lunine, J. I., & Withers, P. G. (2001). Titan, Mars and Earth: Entropy production by latitudinal heat transport. *Geophysical Research Letters*, *28*(3), 415–151.

Lotka, A. J. (1922). Contribution to the energetics of evolution. *PNAS*, *8*(6), 147–418.

Lotka, A. J. (1925). *Elements of physical biology*. Williams and Wilkins Company.

Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S. J., et al. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *PNAS*, *97*(8), 4298–4403.

Martyushev, L. M., & Seleznev, V. D. (2006). Maximum entropy production principle in physics, chemistry and biology. *Physics Reports*, *426*(1), 1–45.

Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and cognition: The realization of the living*. Dordrecht, Holland: Kluwer Academic Publishers.

Maturana, H. R., & Varela, F. J. (1987). *The tree of knowledge: The biological roots of human understanding*. Boston, MA: Shambhala Publications.

Maynard-Smith, J., & Szathmáry, E. (1997). *The major transitions in evolution*. Oxford University Press.

McGregor, S., & Virgo, N. (2009). *Life and its close relatives.* To appear in Proceedings of the Tenth European Conference on Artificial Life. Springer.

McMullin, B. (2000). Some remarks on autocatalysis and autopoiesis. *Annals of the New York Academy of Sciences*, *901*, 163–174.

McMullin, B. (2004). Thirty years of computational autopoiesis: a review. *Artificial Life*, *10*(3), 277–294.

Miller, S. L. (1953). Production of amino acids under possible primitive earth conditions. *Science*, *117*, 528–529.

Moreno, A., & Ruiz-Mirazo, K. (1999). Metabolism and the problem of its universalization. *BioSystems*, *49*(1), 45–61.

Morowitz, H. (1968). *Energy flow in biology*. New York and London: Academic Press.

Morowitz, H. (1978). *Foundations of bioenergetics*. Academic Press.

Niven, R. K. (2009). Steady state of a dissipative flow-controlled system and the maximum entropy production principle. *Physical Review E*.

Odum, E. P. (1969). The strategy of ecosystem development. *Science*, *164*(3877), 262–270.

Odum, H. T., & Pinkerton, R. C. (1955). Time's speed regulator: The optimum efficiency for maximum output in physical and biological systems. *American Scientist*, *43*, 331–343.

Oparin, A. I. (1952). *The origin of life*. New York: Dover.

Paltridge, G. W. (1979). Climate and thermodynamic systems of maximum dissipation. *Nature*,

279, 630–631.

Paltridge, G. W. (2005). Stumbling into the MEP racket: An historical perspective. In A. Kleidon & R. D. Lorenz (Eds.), *Non-equilibrium thermodynamics and the produciton of entropy* (chap. 3). Springer Verlag.

Pask, G. (1958). Physical analogues to the growth of a concept. In *Mechanisation of thought processes: Proceedings of a symposium held at the national physical laboratory* (pp. 879–922). London: H.M.S.O.

Pask, G. (1960). The natural history of networks. In M. C. Yovits & S. Cameron (Eds.), *Self-organising systems* (pp. 232–263). Pergamon.

Pask, G. (1961). A proposed evolutionary model. In H. Van Foerster & G. W. Zopf (Eds.), *Principles of self-organisation* (pp. 229–254). Pergamon.

Pearson, J. E. (1993). Complex patterns in a simple system. *Science*, *261*(5118), 189–192.

Penn, A. S. (2005). *Ecosystem selection: Simulation, experiment and theory*. Unpublished doctoral dissertation, University of Sussex.

Price, H. (1996). *Time's arrow and archimedes' point: New directions for the physics of time*. New York ; Oxford: Oxford University Press.

Prigogine, I. (1955). *Introduction to thermodynamics of irreversible processes* (3rd ed.). Wiley InterScience.

Prigogine, I. (1978). Time, structure and fluctuations. *Science*, *201*(4358), 777–785.

Rasmussen, S., Chen, L., Deamer, D., Krakauer, N., Packard, P., Stadler, P., et al. (2004). Transitions from nonliving to living matter. *Science*, *303*, 963–965.

Richardson. (1969). On the principle of minimum entropy production. *Biophysics*, *9*(2), 265–267.

Ruiz-Mirazo, K., & Moreno, A. (2000). Searching for the roots of autonomy: the natural and artificial paradigms revisited. *Communication and Cognition–Artificial Intelligence*, *17*, 209–228.

Schneider, E. D., & Kay, J. J. (1994). Life as a manifestation of the second law of thermodynamics. *Mathematical and Computer Modelling*, *19*(6–8), 25–48.

Schrödinger, E. (1944). *What is life?* Cambridge University Press.

Schulze-Makuch, D., & Grinspoon, D. H. (2005). Biologically enhanced energy and carbon cycling on Titan. *Astrobiology*, *5*(4), 560–567.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*, 379–423, 623–656.

Shimokawa, S., & Ozawa, H. (2005). Thermodynamics of the ocean circulation: A global per-

spective on the ocean system and living systems. In A. Kleidon & R. D. Lorenz (Eds.), *Non-equilibrium thermodynamics and the production of entropy* (chap. 10). Springer Verlag.

Simondon, G. (1964/1992). Genesis of the individual. In J. Crary & S. Kwinter (Eds.), *Incorporations.* New York: Zone Books. (Translated by Crary, J. and Kwinter, S.)

Smith, M. L., Bruhn, J. N., & Anderson, J. B. (1992). The fungus *Armillaria bulbosa* is among the largest and oldest living organisms. *Nature*, *356*, 428–431.

Stewart, J. (1992). Life = cogntion: The epistemological and ontological significance of artificial life. In F. J. Varela & P. Bourgine (Eds.), *Towards a practice of autonomous systems: Proceedings of the first european conference on artificial life* (pp. 475–483). MIT Press.

Tribus, M. (1961). *Thermostatics and thermodynamics*. Van Norstrand.

Tykodi, R. J. (1967). *Thermodynamics of steady states*. New York: Macmillan.

Ulanowicz, R. E. (1980). An hypothesis on the development of natural communities. *Journal of Theoretical Biology*, *85*, 223–245.

Varela, F. J., Maturana, H. R., & Uribe, R. (1974). Autopoiesis: The organization of living systems, its characterization and a model. *BioSystems*, *5*, 187–196.

Virgo, N. (2010). From maximum entropy to maximum entropy production: A new approach. *Entropy*, *21*(1), 107–126.

Virgo, N., Egbert, M., & Froese, T. (2009). *The role of the spatial boundary in autopoiesis.* To appear in Proceedings of the Tenth European Conference on Artificial Life. Springer.

Virgo, N., & Harvey, I. (2007). Entropy production in ecosystems. In F. Almeida e Costa et al. (Eds.), *Proceedings of the ninth european conference on artificial life* (pp. 123–132). Springer Verlag.

Virgo, N., & Harvey, I. (2008a). Adaptive growth processes: A model inspired by Pask's ear. In S. Bullock, J. Noble, R. A. Watson, & M. A. Bedau (Eds.), *Proceedings of the eleventh international conference on artificial life* (pp. 656–651). MIT Press.

Virgo, N., & Harvey, I. (2008b). Reaction-diffusion spots as a model for autopoiesis (abstract). In S. Bullock, J. Noble, R. A. Watson, & M. A. Bedau (Eds.), *Proceedings of the eleventh international conference on artificial life* (p. 816). MIT Press.

Virgo, N., Law, R., & Emmerson, M. (2006). Sequentially assembled food webs and extremum principles in ecosystem ecology. *Journal of Animal Ecology*, *75*(2), 377–386.

Wächtershäuser, G. (1988). Before enzymes and templates: Theory of surface metabolism. *Microbiology and Molecular Biology Reviews*, *52*(4), 452–484.

Weber, A., & Varela, F. J. (2002). Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the Cognitive Sciences*, *1*(2), 97–125.

Williams, H., & Lenton, T. (2007). Artificial selection of simulated microbial ecosystems. *Proceedings of the National Academy of Sciences*, *104*(21), 8198–8923.

Zupanovic, P., Botric, S., & Juretic, D. (2006). Relaxation processes, MaxEnt formalism and Einstein's formula for the probability of fluctuations. *Croatia Chemica Acta*, *79*(3), 335-338.