4 Maximum Entropy Production and Non-equilibrium Statistical Mechanics

Roderick C. Dewar

Unité de Bioclimatologie, INRA Centre de Bordeaux-Aquitaine, B.P. 81, 33883 Villenave d'Ornon CEDEX, France.

Summary. Over the last 30 years empirical evidence in favour of the Maximum Entropy Production (MEP) principle for non-equilibrium systems has been accumulating from studies of phenomena as diverse as planetary climates, crystal growth morphology, bacterial metabolism and photosynthesis. And yet MEP is still regarded by many as nothing other than a curiosity, largely because a theoretical justification for it has been lacking. This chapter offers a non-mathematical overview of a recent statistical explanation of MEP stemming from the work of Boltzmann, Gibbs, Shannon and Jaynes. The aim here is to highlight the key physical ideas underlying MEP. For non-equilibrium systems that exchange energy and matter with their surroundings and on which various constraints are imposed (*e.g.*, external forcings, conservation laws), it is shown that, among all the possible steady states compatible with the imposed constraints, Nature selects the MEP state because it is the most probable one, *i.e.*, it is the macroscopic state that could be realised by more microscopic pathways than any other. That entropy production is the extremal quantity emerges here from the universal constraints of local energy and mass balance that apply to all systems, which may explain the apparent prevalence of MEP throughout physics and biology. The same physical ideas also explain self-organized criticality and a result concerning the probability of violations of the second law of thermodynamics (the Fluctuation Theorem), recently verified experimentally. In the light of these results, dissipative structures of high entropy production, which include living systems, can be viewed as highly probable phenomena. The prospects for applying these results to other types of non-equilibrium system, such as economies, are briefly outlined.

> *If one grants that [the principle of maximum Shannon entropy] represents a valid method of reasoning at all, one must grant that it gives us also the long-hoped-for general formalism for the treatment of irreversible processes ... [T]he issue is no longer one of mere philosophical preference for one viewpoint or another ; the issue is now one of definite mathematical fact. For the assertion just made can be put to the test by carrying out specific calculations, and will prove to be either right or wrong.*
>
> Jaynes (1979)

4.1 Introduction

Edwin Thompson Jaynes[1] (1922–1998) made many original and fundamental contributions to science in fields as diverse as applied classical electrodynamics, information theory, the foundations of probability theory, the interpretation of quantum mechanics, and radiation theory.

Much of his work is the expression of a single conviction, that probability theory – in which probability is interpreted in the original sense understood by Laplace and Bernoulli, as a measure of our state of knowledge about the real world – provides the uniquely valid rules of logic in science (Jaynes and Bretthorst 2003). In the vast majority of scientific problems actually encountered, we do not have sufficient information to apply deductive reasoning. What we need, said Jaynes, are the logic and tools of statistical inference (*i.e.*, of probability theory) so that we may draw rational conclusions from the limited information we do have.

A key outcome of that conviction was Jaynes' reformulation of statistical mechanics in terms of information theory (Jaynes 1957a,b). This opened the way to the extension of the logic underlying equilibrium statistical mechanics (ESM) – implicit in the work of Boltzmann and Gibbs – to non-equilibrium statistical mechanics (NESM), as well as to many other problems of statistical inference (*e.g.*, image reconstruction, spectral analysis, inverse problems). In all applications of this logic, the basic recipe consists of the maximisation of Shannon information entropy, subject to the constraints imposed by the available information – an algorithm now known as MAXENT (*e.g.*, Jaynes 1985a).

How is it, then, that Jaynes' MAXENT formulation of NESM has for so long failed to be accepted by the majority of scientists when the logic of it is precisely that of Boltzmann and Gibbs?

Part of the answer lies with the relative paucity of published results from the MAXENT school (Dougherty 1994), especially with regard to new testable predictions far from equilibrium. The most extended account of Jaynes' NESM appears as part of a conference paper (Jaynes 1979, Sect. D). While that account makes clear the generality of the approach in principle, it is applied there within a perturbative approximation only to reproduce some known results for near-equilibrium behaviour.

Another reason why the MAXENT formulation of NESM has not caught on as it might have done almost certainly lies with the conceptual gulf between the Bayesian and frequency viewpoints of probability (Jaynes 1979, 1984). The frequency viewpoint – that probability is an inherent property of the real world (the sampling frequency) rather than a property of our state of knowledge about the real world (the Bayesian viewpoint) – dominated

[1] A biographical sketch and bibliography are available at http://bayes.wustl.edu/etj/etj.html

scientific thinking for much of the twentieth century. No wonder, then, that progress has been slow.

The hypothesis of maximum entropy production (MEP), which is explored by several authors in this volume, has likewise made slow progress. It has still to be widely accepted as a generic property of non-equilibrium systems despite a growing body of empirical evidence pointing in that direction (Lorenz 2003; Ozawa et al. 2003). The sticking point has been the perceived lack of a rigorous theoretical explanation for MEP.

So here we have, on the one hand, a theory in search of evidence (the MAXENT formulation of NESM) and, on the other hand, evidence in search of a theory (MEP). This Chapter gives an overview of some recent work proposing a mutually beneficial marriage between the two (Dewar 2003). As is often the case with such proposals, while the purpose might be well intentioned the result may be to have rocks thrown from both sides.

The main difficulties encountered at this stage are not so much technical as conceptual in nature, and so here I will try to give a non-mathematical account that emphasises the key physical ideas leading to MEP. I begin by briefly retracing the historical path of ideas from Boltzmann to Gibbs and Shannon which eventually led to Jaynes' MAXENT formulation of NESM. But if Jaynes' formulation is essentially an algorithm for statistical inference, what is the guarantee that it should work as a description of Nature? I discuss two key ideas of Jaynes – *macroscopic reproducibility* and *caliber* – that make the physical relevance of the algorithm intuitively clear (Jaynes 1980, 1985b).

Building on these ideas, I then present the path information formalism of NESM and discuss some new far-from-equilibrium predictions that have recently been obtained from it (Dewar 2003) – specifically, the emergence of MEP and self-organized criticality, and a result (known as the Fluctuation Theorem) concerning the probability of violations of the second law of thermodynamics.

I conclude that the MAXENT derivation of MEP explains its apparent prevalence throughout physics and biology, and suggests how MEP might be applicable to non-equilibrium systems more generally (*e.g.*, economies, also see Ruth, this volume). In the light of this derivation, dissipative structures of high entropy production, which include living systems, may now be understood as phenomena of high probability.

4.2 Boltzmann, Gibbs, Shannon, Jaynes

Boltzmann interpreted Clausius' empirical entropy (S) as the logarithm of the number of ways (W), or microstates, by which a given macroscopic state can be realized ($S = k\log W$, where k is Boltzmann's constant). The second law of thermodynamics (maximum entropy) then simply means that the observed macrostate is the most probable one, *i.e.*, it is the one that could be realized by Nature in more ways than any other. Microstate counting worked fine for isolated systems with fixed total energy and particle number.

Gibbs noted that Boltzmann's results could also be obtained by minimising the somewhat obscure quantity $\Sigma_i p_i \log p_i$ with respect to the microstate probabilities p_i, subject to the appropriate constraints on energy and particle number. Gibbs (1902) called the quantity $\Sigma_i p_i \log p_i$, or $<\log p_i>$, the 'average index of probability of phase'. He was then able to generalise ESM to open systems, by extending the imposed constraints to include system interactions with external heat and particle reservoirs. However, just what the Gibbs algorithm meant, and its relation to Boltzmann's insight, remained obscure.

Much later, Shannon (1948) introduced the information entropy, $-\Sigma_i p_i \log p_i$, as a measure of the amount of missing information (*i.e.*, uncertainty) associated with a probability distribution p_i. In the context of ESM, Shannon's information entropy measures our state of ignorance about the actual microstate the system is in. More quantitatively, a result called the Asymptotic Equipartition Theorem tells us that for systems with many degrees of freedom, the information entropy is equal to the logarithm of the number of microstates having non-zero probability (Jaynes 1979). It is a direct measure of the extent (or spread) of the microstate distribution p_i in phase space.

Jaynes' first insight was to see, in the light of Shannon's work, what the Gibbs algorithm meant and how it related to Boltzmann's insight. By maximising the information entropy with respect to p_i, Gibbs was constructing the microstate distribution with the largest extent compatible with the imposed constraints, thus generalising Boltzmann's logic of microstate counting to the microstate distribution (Jaynes 1957a,b).

But as Jaynes went on to realise, the Gibbs algorithm is much more than that. For the quantity $-\Sigma_i p_i \log p_i$ can be missing information about *anything*, not just about the microstates of equilibrium systems. Thus, during the late 1950s and early 1960s Jaynes developed his information theory formulation of NESM based on applying the Gibbs algorithm to non-equilibrium systems (Jaynes 1979).

But it did not stop there. Jaynes saw the Gibbs algorithm as a completely general recipe for statistical inference in the face of insufficient information (MAXENT), with useful applications throughout science, not just in statistical mechanics. Viewed as such, it is a recipe of the greatest rationality because it makes the least-biased assignment of probabilities, *i.e.*, the one that incorporates only the available information (imposed constraints). To make any other assignment than the MAXENT distribution would be unwarranted because that would presume extra information one simply does not have, leading to biased conclusions.

4.3 Macroscopic Reproducibility

But if MAXENT is essentially an algorithm of statistical inference (albeit the most honest one), what guarantee is there that it should actually work as a description of Nature? The answer lies in the fact that we are only concerned with describing the reproducible phenomena of Nature.

Suppose certain external constraints act on a system. Examples include the solar radiation input at the top of Earth's atmosphere, the temperature gradient imposed across a Bénard convection cell, the velocity gradient imposed across a sheared fluid layer, or the flux of snow onto a mountain slope. If, every time these constraints are imposed, the same macroscopic behaviour is reproduced (atmospheric circulation, heat flow, shear turbulence, avalanche dynamics), then it must be the case that knowledge of those constraints (together with other relevant information such as conservation laws and the spectrum of possible microstates) is sufficient for theoretical prediction of the macroscopic result. All other information must be irrelevant for that purpose. It cannot be necessary to know the myriad of microscopic details that were not under experimental control and would not be the same under successive repetitions of the experiment (Jaynes 1985b). We can only imagine with horror the length of scientific papers that would be required for others to reproduce our results if this were not the case.

MAXENT acknowledges this fact by discarding the irrelevant information at the outset. By maximising the Shannon information entropy (*i.e.*, missing information) with respect to $p_{\rm i}$ subject only to the imposed constraints, MAXENT ensures that only the information relevant to macroscopic prediction is encoded in the distribution $p_{\rm i}$. Therefore, *if* we have correctly identified all the relevant constraints (and other prior information), then macroscopic predictions calculated as expectation values over the MAXENT distribution will match the experimental results reproduced under those constraints.

But of course that last *if* is crucial. In any given application of MAXENT there is no *a priori* guarantee that we have incorporated all the relevant constraints. But if we have not done so, then MAXENT will signal the fact *a posteriori* through a disagreement between predicted and observed behaviours, the nature of the disagreement indicating the nature of the missing constraints (*e.g.*, new physics). MAXENT's failures are more informative than its successes. This is the logic of science.

Jaynes considered reproducibility – rather than disorder – to be the key idea behind the second law of thermodynamics (Jaynes 1963, 1965, 1988, 1989). Suppose that under given experimental conditions a system evolves reproducibly from initial macrostate A to final macrostate B (Fig. 4.1). The initial microstate lies somewhere in the phase volume $W_{\rm A}$ compatible with A, although we do not know where exactly because we cannot set up the system with microscopic precision. By Liouville's theorem, the system ends up somewhere in a new region of phase space $W'_{\rm A}$ having the same volume $W_{\rm A}$. If the macroscopic transition A $\rightarrow$ B is reproducible for all initial microstates, then W'_A cannot be greater than the phase volume W_B compatible with B. Hence $S_A = k\log W_A = k\log W'_A \leq k\log W_B = S_B$. This is the second law.

By the same token, while the reverse macroscopic process B → A (squeezing the toothpaste back into the tube) is possible because the microscopic equations of motion are reversible, it is not achievable reproducibly (*i.e.*, it is highly improbable) because we cannot ensure by macroscopic means that the initial state lies in the appropriate subset of W_B (*i.e.*, W'_A with all molecular velocities reversed) to get us back to A. Jaynes (1988) put some numbers to this: the probability of B → A, he conjectured[2], is something like $p = W_A/W_B = \exp(-(S_B - S_A)/k)$. If the entropy difference corresponds to just one microcalorie at room temperature, then we have $p < \exp(-10^{15})$. Although the macroscopic process B → A is possible, it is not macroscopically reproducible. The second law is the price paid for macroscopic reproducibility.

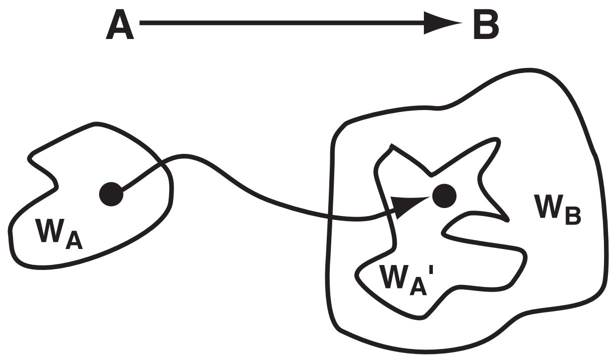

Fig. 4.1. The second law explained by macroscopic reproducibility (see text)

Already in 1867 James Clerk Maxwell understood that the second law was 'of the nature of a strong probability ... not an absolute certainty' like dynamical laws (Harman 1998). He introduced his 'finite being' (the term 'demon' was later coined by William Thomson) to underline this very point and to reject current attempts (notably by Clausius and Boltzmann) to reduce the law to a theorem in dynamics. Perhaps Maxwell had something like Fig. 4.1 in mind when he wrote that 'the 2nd law of thermodynamics has the same degree of truth as the statement that if you throw a tumblerful of water into the sea you cannot get the same tumblerful out again' (Maxwell 1870). Without the means to identify and pick out the individual molecules involved, the process is effectively irreversible. Water flow is indeed a good analogue of Fig. 4.1; according to Liouville's theorem, probability in phase space behaves like an incompressible fluid.

[2] Jaynes's conjecture that 2nd law violating processes are exponentially improbable anticipates the Fluctuation Theorem (see below).

4.4 The Concept of Caliber

Mathematically, Jaynes expressed his MAXENT approach to NESM in a rather formal way, as the solution to a statistical inference problem of the most general kind (Jaynes 1979, Sect. D). From some known but otherwise arbitrary macroscopic history A, say, between times $t = -\tau$ and $t = 0$, we are asked to predict the macroscopic trajectory B for later times $t > 0$ (or to retrodict the previous history for $t < -\tau$). For non-equilibrium systems, A and B will generally involve both space- and time-dependent macroscopic quantities.

In the MAXENT approach one maximises the Shannon information entropy S subject to the known history A, thus obtaining the initial microstate probability distribution $\rho(t = 0)$. In principle, one then integrates the (known) microscopic equations of motion to obtain $\rho(t)$ for all other times t, and constructs the macroscopic trajectory B by calculating the appropriate variables as expectation values over $\rho(t)$.

The maximised value of S depends on the known macroscopic history A, *i.e.*, $S = S(A)$. Jaynes (1980) called that value the *caliber* of history A. It is a measure of the number of initial microstates compatible with that history. Equivalently, if you think of the previous history of each microstate for $-\tau < t < 0$ as a path in phase space, then the caliber is also a measure of the cross-sectional area of a tube formed by bundling together (like a stack of spaghetti) all those microscopic phase space paths compatible with the known macroscopic history A.

Jaynes (1980) suggested that the unknown macroscopic trajectory B could be inferred by an extension of the Gibbs algorithm that had given the caliber $S(A)$. Out of all possible macroscopic trajectories B, choose that one for which the combined caliber $S(A,B)$ is the greatest, because that is the one that could be realized by Nature in more ways than any other consistent with A. He later referred to this as the *maximum caliber principle*, and noted a tantalising analogy between the caliber in NESM and the Lagrangian in mechanics (Jaynes 1985b).

4.5 Path Information Formalism of NESM

Recently I reformulated the MAXENT approach to NESM directly in terms of phase space paths (Dewar 2003). That is, one maximises the *path information entropy* $S = -\Sigma_\Gamma p_\Gamma \log p_\Gamma$ with respect to p_Γ, subject to the imposed constraints and other relevant information (*e.g.*, conservation laws). Here p_Γ is the (Bayesian) probability of microscopic phase space path Γ, and the sum is over all paths permitted by the microscopic equations of motion. S is a measure of our state of ignorance about which microscopic path the system actually follows over time.

As an initial test, I applied this formalism to non-equilibrium stationary states in which all macroscopic variables are independent of time, although spatial variations (*e.g.*, temperature gradients) are of course present. My main interest was to try to put maximum entropy production and a result known as the Fluctuation Theorem (Evans and Searles 2002) – which also concerns entropy production – on a common theoretical footing.

Why a path formalism? Non-equilibrium systems are characterised by the presence of macroscopic fluxes, *i.e.*, macroscopic exchanges of energy and matter over time. It seemed desirable then also to represent the microscopic behaviour explicitly over time, so phase space paths rather than microstates became the natural objects of interest. This choice was also influenced by my reading of the literature on the Fluctuation Theorem, for which the standard proofs explicitly considered pairs of phase space trajectories related by path reversal (Evans and Searles 2002).

I also sensed that a path formalism was somehow truer to the spirit of Jaynes' maximum caliber principle, although at the time I did not see explicitly how his presentation of it (predicting unknown trajectory B from known history A) related to my steady-state problem[3].

Before discussing some specific predictions of this path information formalism of NESM, let us pause to state the problem more precisely and to recap the rationale for its solution by MAXENT, now in the twin contexts of reproducibility and maximum caliber.

Problem: Given the external constraints (and any other relevant constraints) acting on our system, which macroscopic steady state – among all possible macroscopic steady states compatible with those constraints – is the one selected by Nature? For example, given the input of solar radiation at the top of the Earth's atmosphere (and local energy and mass conservation), which climate state is selected?

Solution: The selected steady state is completely described by that path distribution which maximises the path information entropy S subject to those (and only those) constraints. Macroscopic quantities are calculated as expectation values over the path distribution.

Rationale in terms of reproducibility: If Nature selects the state reproducibly, *i.e.*, every time the constraints are imposed, then knowledge of those constraints alone must be sufficient to predict the result. Provided we have incorporated all the relevant constraints, macroscopic predictions inferred as expectation values over the MAXENT path distribution will agree with Nature's selected state.

Rationale in terms of maximum caliber: For systems with many degrees of freedom, the Asymptotic Equipartition Theorem implies that the value of S for a given path distribution p_Γ is the logarithm of the number of paths with non-zero probability. Let us call S the caliber of p_Γ (*i.e.*, the path analogue of a microstate distribution's extent in phase space). The MAXENT path distribution therefore encompasses the largest number of paths compatible with the imposed constraints, *i.e.*, it describes the macroscopic state that can be realized by more paths than any other.

[3] This only became clear to me during the preparation of the present article.

Reproducibility = maximum caliber: These two rationales are physically equivalent. The selected steady state is reproducible precisely because its macroscopic properties are characteristic of *each* of the overwhelming majority of possible microscopic paths compatible with the constraints. That is why it does not matter which particular path the system follows in any given repetition of the experiment.

In this formalism the caliber $S = -\Sigma_\Gamma p_\Gamma \log p_\Gamma$ is defined for any path distribution p_Γ, not just the MAXENT distribution. Using this more general definition of caliber (logarithm of the number of paths with $p_\Gamma > 0$), the maximum caliber principle is then identical to the Gibbs algorithm, rather than an extension of it (cf. Jaynes 1985b).

Finally it is worth emphasising that expectation values calculated from the path distribution are statistical inferences (Bayesian viewpoint). They are not the result of the system somehow sampling different paths in the real world (frequency viewpoint). While microstate distributions have always been open to an ergodic interpretation (expectation value = time average), clearly this interpretation makes no sense at all for the path distribution. In any given experiment the system only ever follows one path, S measures our state of ignorance about which one, and MAXENT predicts the macroscopic behaviour reproduced each time.

4.6 New Results Far from Equilibrium

The above formalism was used to predict the stationary steady-state properties of a general open, non-equilibrium system exchanging energy and matter with its external environment (Dewar 2003). The path information entropy $S = -\Sigma_\Gamma p_\Gamma \log p_\Gamma$ was maximised with respect to p_Γ subject to the relevant constraints, denoted collectively by A. Typically, these constraints consist of: ($A1$) Local energy and mass balance (conservation laws); ($A2$) Global steady-state energy and mass balance (stationarity); ($A3$) External forcings (*e.g.*, solar radiation input, temperature gradient, velocity gradient, snow flux) by which the system is maintained out of equilibrium. Note that $A1$ and $A2$ are universal constraints common to all such systems, whereas $A3$ is specific to each system. In any practical application, we must also specify the spectrum of possible microscopic paths Γ, although the results discussed here only depend on very general properties of that spectrum.

All macroscopic quantities can then be calculated as expectation values over the MAXENT path distribution. Which macroscopic quantities are we interested in? Typically these are the local distributions of heat and mass density within the system and the fluxes of heat and mass across the system boundary. These describe both the internal state of the system and its interaction with the external environment. Let us denote this macroscopic information collectively by B.

It then proves useful to apply the MAXENT algorithm in two steps, with the unknown macroscopic state B acting as a temporary constraint which is subsequently relaxed. S is first maximised with respect to the path distributionp_Γ, subject to $A1$ and B. The maximised value of S after this first step (denoted S_1) depends on B, *i.e.*, $S_1 = S_1(B)$. In the second step, $S_1(B)$ is maximised with respect to B, subject to the remaining constraints $A2$ and $A3$. This completes the MAXENT algorithm. The choice of B that maximises $S_1(B)$ represents the macroscopic steady state that is reproduced under the imposed constraints A. The final result for B is the same as if we had constructed the MAXENT path distribution imposed by A alone, and then calculated the properties of B directly as expectation values over that distribution.

We now note the close formal analogy between this procedure and Jaynes' procedure for inferring an unknown macroscopic trajectory B from a known macroscopic history A. Here we are inferring an unknown macroscopic steady state B from known constraints A. In each case the joint caliber $S(A,B)$ is maximised with respect to B, subject to A. In each case, if A is sufficient to reproduce B, then MAXENT will correctly predict the observed B.

4.6.1 Maximum Entropy Production (MEP)

After step 1 we find that $S_1(B) = \log W(EP{=}EP_\mathrm{B})$, where $W(EP{=}EP_\mathrm{B})$ is the number of paths whose thermodynamic entropy production rate (EP) equals that of macrostate B. The 'density of paths', W, is analogous to the density of states in equilibrium statistical mechanics. It depends on the spectrum of paths permitted by the microscopic equations of motion. The entropy production rate that emerges here (EP) is just the familiar near-equilibrium expression involving products of fluxes and thermodynamic forces, but here it is also valid far from equilibrium. We have not assumed a local equilibrium hypothesis. The various contributions to EP derive directly from the contributions to local energy and mass balance (constraint $A1$), *e.g.*, heat flow and frictional heating (from local heat balance), mass flow and chemical reactions (from local mass balance).

In step 2 of the MAXENT algorithm we choose the value of B for which $\log W(EP{=}EP_\mathrm{B})$ is maximal, subject to $A2$ and $A3$. This occurs when EP_B is maximal. Therefore step 2 is equivalent to MEP subject to $A2$ and $A3$. The only requirement here is that the density of paths W is an increasing function of EP in the neighbourhood of the maximum EP_B, so that a maximum in EP_B corresponds to a maximum in $\log W(EP{=}EP_\mathrm{B})$. That EP_B has a maximum reflects the trade-off between the component thermodynamic fluxes and forces, in which increased fluxes tend to dissipate the thermodynamic forces.

The conclusion here is that the MEP state is selected because it is the non-equilibrium steady state with the highest caliber, *i.e.*, the one that can be realised by more microscopic paths than any other steady state compatible with the constraints. Because entropy production emerges here directly from local energy and mass balance – the universal constraint $A1$ valid for all systems – it becomes clear why MEP is so prevalent across physics and biology. The predictions of MEP under constraints $A2$ and $A3$ will vary from one system to another, reflecting the specific nature of $A3$, but the principle of MEP itself would appear to have the same validity as energy and mass conservation (constraint $A1$).

Provided we have identified all the relevant constraints, MEP will predict the experimentally reproduced result. Failure to do so will signal the presence of unaccounted constraints; but it could also indicate that we have ignored some contributions to the entropy production itself, signalling missing terms in our equations for local energy and mass balance.

4.6.2 The Fluctuation Theorem (FT)

Another general result that emerges from step 1 concerns the probability of violations of the second law (which, as Maxwell appreciated, is statistical in character).

Specifically, we find that the MAXENT probability of path Γ is proportional to $\exp(\tau EP_\Gamma/2k)$, where τ is the time duration of path Γ, EP_Γ is its entropy production rate, and k is Boltzmann's constant. Now consider the reverse path Γ_R obtained by starting from the end of path Γ and reversing all the molecular velocities so that we end up at the start of path Γ (*i.e.*, reversing the curved path in Fig. 4.1). The entropy production rate of Γ_R is equal to $-EP_\Gamma$ by time-reversal symmetry of the microscopic equations of motion (*i.e.*, all fluxes are reversed). This immediately implies that the ratio of the probability of Γ_R to that of Γ is equal to $\exp(-\tau EP_\Gamma/k)$.

As is easily shown, this result implies that the second law holds on the average, *i.e.*, $<EP> \geq 0$. It also says that second law violating paths with negative entropy production are possible, although exponentially improbable. This result is known as the Fluctuation Theorem (FT) (Evans and Searles 2002). The FT was first[4] derived heuristically in 1993. Subsequent derivations of the FT have been based on ergodicity and causality assumptions. Computer simulations of various models of microscopic dynamics have confirmed its validity. The first truly experimental verification of the FT was obtained in a delicate experiment which followed the Brownian motion of colloidal particles in an optical trap (Wang et al. 2002).

The path information formalism of NESM puts the FT and MEP on a common theoretical footing, and predicts that both are valid on very general grounds. The Bayesian rationale of MAXENT implies that ergodicity is not required to explain the FT. Rather, MAXENT suggests that the exponentially small probability of violations of the second law is, like MEP, characteristic of the reproducible behaviour of all systems obeying local energy and mass conservation.

[4] Jaynes (1988) was essentially there when he conjectured on grounds of macroscopic reproducibility (Jaynes 1963, 1965) that $p = W_\mathrm{A}/W_\mathrm{B} = \exp((-S_\mathrm{B} - S_\mathrm{A})/k)$ for the entropy-consuming transition B $\rightarrow$ A (Fig. 4.1).

4.6.3 Self-Organized Criticality (SOC)

Now we come to a result that emerged as an unexpected bonus (Dewar 2003). Some non-equilibrium systems such as earthquakes, snow avalanches, sandpiles, and forest fires tend to organize themselves into steady states which are characterised by large-scale fluctuations (Jensen 1998). This behaviour is reminiscent of equilibrium systems at phase transitions, obtained when variables such as temperature and pressure are tuned to critical values. Only, many non-equilibrium systems appear to organize themselves into a critical state (SOC), apparently without tuning.

Nevertheless, all these systems are tuned to some extent. Typically they are forced out of equilibrium by a fixed but very slow input flux, $F_{\rm in}$ (of momentum, snow, sand, and lightning strikes, in the above examples – cf. constraint $A3$). In the archetypal example where a sprinkling of grains falls onto a sandpile (Bak et al. 1987), the slope of the sandpile tends to its largest possible value (the critical angle of repose), while critical fluctuations in the output grain flux about its average value (equal to $F_{\rm in}$ in the steady state) are induced in the form of sand avalanches of all sizes.

We can begin to understand SOC from an MEP perspective simply by noting that in the steady state, the sandpile entropy production is the product of the grain flux ($F_{\rm in}$) and the slope. Because $F_{\rm in}$ is fixed, MEP predicts that the slope adopts its largest possible value. In other words, SOC is a special case of MEP applied to flux-driven systems. But what about the fluctuations?

In the path information formalism of NESM, fluctuations are described by the path distribution p_Γ. It can be shown from the path-reversal symmetry properties of p_Γ that, in the limit of slow input flux $F_{\rm in} \to 0$, the variance of the magnitude of the output grain flux (the avalanches) is proportional to $1/{F_{\rm in}}^2$ and therefore diverges to infinity as $F_{\rm in} \to 0$, the characteristic signature of SOC. This result involves exactly the same mathematics as in classical theories of equilibrium phase transitions, with ${F_{\rm in}}^2$ playing the role of the control parameter (cf. $|T - T_{\rm c}|$, the amount by which a ferromagnet is cooled below its Curie temperature $T_{\rm c}$) and the avalanche flux playing the role of the order parameter (cf. spontaneous magnetisation). As the control parameter is tuned to zero, the order parameter goes to zero but fluctuations in the order parameter emerge on all length scales (cf. divergence of magnetic susceptibility).

We can understand SOC intuitively from an information perspective. As $F_{\rm in} \to 0$, the external constraint becomes scale-free. Provided there is no scale set by other internal constraints (*e.g.*, friction), then the steady state reproduced under these constraints must also be scale-free. Consequently, the system is dominated by fluctuations on all scales.

4.7 Thermodynamics of Life

What conclusions may we now draw regarding the thermodynamics of life? Non-equilibrium dissipative structures, which include living systems, appear to be consistent with MEP. They couple extended regions of high order (*e.g.*, convention cells, mass transport pathways) with localised regions of high dissipation (*e.g.*, boundary layers, chemical reaction sites). The localised regions are responsible for most of the system entropy production, while the ordered regions act as transport structures which permit this entropy to be produced and exported at the greatest rate possible under the combined constraints of stationarity and local energy and mass balance. Far from equilibrium, the coexistence of ordered and dissipative regions produces and exports more entropy to the environment than a purely dissipative 'soup'.

Since Schrödinger's influential book *What is Life?* (Schrödinger 1944), discussion of the thermodynamics of life has taken a rather biocentric viewpoint along the lines that in order to maintain their internal order living systems must export entropy to their surroundings. This viewpoint sees life as constantly competing against the second law. But if we are to understand the emergence of living systems and other dissipative structures then it is the coexistence of ordered and dissipative regions that we need to focus on, and whose natural selection we need to explain. MAXENT provides the proper viewpoint – we take the imposed constraints as our starting point and we ask: which pattern of energy and mass flows is reproducibly selected under those constraints? In the light of the MAXENT derivation of MEP we can now view living systems (and dissipative structures more generally) as highly probable phenomena. They are selected because they are characteristic of each of the overwhelming majority of ways in which energy and matter could flow under the imposed constraints.

4.8 Further Prospects

MAXENT is a general algorithm for predicting reproducible macroscopic phenomena under given constraints. The derivation of MEP from MAXENT suggests that MEP itself may apply beyond purely physical and biological systems involving energy and mass transfer. Specifically, if a system's macroscopic state B is described by a local variable ρ (*e.g.*, the analogue of heat density) which obeys a balance equation $\partial\rho/\partial t = -\nabla \cdot F + Q$ (local rate of change = flux convergence + net local source), then the two-step MAXENT procedure will lead (after step 1) to the emergence of a generalised entropy production involving contributions from both flux F and source Q (*e.g.*, analogues of thermal and frictional dissipation), and then (after step 2) to analogues of MEP, FT and SOC (Dewar 2003). Explicitly, the generalised entropy production takes the form $EP = \int_V \left(\bar{F} \cdot \nabla\theta + \theta\bar{Q}\right)$ where θ is the analogue of inverse temperature ($1/T$), the overbar indicates a time-average over interval τ, and the space integral extends over the system volume V.

For example, Jaynes (1991) anticipated the application of MAXENT to the prediction of macro-economic behaviour. Does MEP apply there? Are financial crashes SOC? Is there a 2^{nd} law analogue for economies? A starting point would be to identify ρ, F and Q for economies, and to specify the micro- and macro-economic constraints that apply (cf. constraints *A1, A2, A3*).

On the theoretical side, we can see the prospect of generalising the path information formalism of NESM to non-stationary macroscopic phenomena, in the spirit of Jaynes (1979). Can MEP be extended to time-dependent macroscopic trajectories such as cyclic steady states?

But above all, let us go ahead and apply the path information formalism of NESM and its predictions to as wide a range of real-world problems as possible. Then, as Edwin Jaynes would have put it, let the results speak for themselves.

References

Bak P, Tang C, Wiesenfeld K (1987) Self-organized criticality : an explanation of $1/f$ noise. Phys Rev Lett 59:381–384.

Dewar RC (2003) Information theory explanation of the fluctuation theorem, maximum entropy production and self-organized criticality in non-equilibrium stationary states. J Phys A 36:631–641.

Dougherty JP (1994) Foundations of non-equilibrium statistical mechanics. Phil Trans R Soc Lond A346:259–305.

Evans DJ, Searles DJ (2002) The fluctuation theorem. Adv Phys 51:1529–1585.

Gibbs JW (1902) Elementary principles of statistical mechanics. Reprinted by Ox Bow Press, Woodridge CT (1981).

Harman PM (1998) The natural philosophy of James Clerk Maxwell. Cambridge University Press, Cambridge UK.

Jaynes ET (1957a) Information theory and statistical mechanics. Phys Rev 106:620–630.

Jaynes ET (1957b) Information theory and statistical mechanics II. Phys Rev 108:171–190.

Jaynes ET (1963) Information theory and statistical mechanics. In: Ford KW (ed) Brandeis Summer Institute 1962, Statistical Physics. Benjamin, New York, pp 181–218.

Jaynes ET (1965) Gibbs vs Boltzmann entropies. Am J Phys 33:391–398.

Jaynes ET (1979) Where do we stand on maximum entropy? In: Levine RD, Tribus M (eds) The maximum entropy principle. MIT, Cambridge MA, pp 15–118.

Jaynes ET (1980) The minimum entropy production principle. Ann Rev Phys Chem 31:579–601.

Jaynes ET (1984) The intuitive inadequacy of classical statistics. Epistemologia 7(special issue):43–74.

Jaynes ET (1985a) Where do we go from here? In: CR Smith, WT Grandy (eds) Maximum entropy and Bayesian methods in inverse problems. D Reidel, Dordrecht, pp 21–58.

Jaynes ET (1985b) Macroscopic prediction. In: H Haken (ed) Complex systems – operational approaches in neurobiology. Springer-Verlag, Berlin, pp 254–269.

Jaynes ET (1988) The evolution of Carnot's principle. In: Erickson GJ, Smith CR (eds) Maximum-entropy and Bayesian methods in science. Kluwer, Dordrecht, vol. 1, pp 267–282.

Jaynes ET (1989) Clearing up mysteries – the original goal. In: Skilling J (ed) Maximum entropy and Bayesian methods. Kluwer, Dordrecht, pp 1–27.

Jaynes ET (1991) How should we use entropy in economics? Unpublished manuscript available at http://bayes.wustl.edu/etj/etj.html.

Jaynes ET, Bretthorst GL (ed) (2003) Probability theory : the logic of science. Cambridge University Press, Cambridge UK.

Jensen HJ (1998) Self-organized criticality. Cambridge University Press, Cambridge UK.

Lorenz RD (2003) Full steam ahead – probably. Science 299:837–838.

Maxwell JC (1870). Letter to John William Strutt. In: PM Harman (ed) The scientific letters and papers of James Clerk Maxwell. Cambridge University Press, Cambridge UK 2:582–583.

Ozawa H, Ohmura A, Lorenz RD, Pujol T (2003) The second law of thermodynamics and the global climate system – A review of the maximum entropy production principle, Rev Geophys 41: 1018.

Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27:379–423, 623–656.

Schrödinger E (1944) What is life? Cambridge University Press, Cambridge UK.

Wang GM, Sevick EM, Mittag E, Searles DJ, Evans DJ (2002) Experimental demonstration of violations of the second law of thermodynamics for small systems and short time scales. Phys Rev Lett 89, 050601.