

# Microscopic Dynamics and the Second Law of Thermodynamics

Michael C. Mackey \*  
Departments of Physiology, Physics, & Mathematics  
Centre for Nonlinear Dynamics  
McGill University

September 25, 2001

## Abstract

This paper considers the origin of the thermodynamic behaviour captured by the second law of thermodynamics. It is first shown the Gibbs definition of entropy is the unique definition for both equilibrium and non-equilibrium situations when entropy is an extensive quantity. It is then demonstrated that the invertible microscopically formulated physical laws are incapable of explaining the second law of thermodynamics. A necessary and sufficient condition to derive the second law of thermodynamics is that we have *exact dynamics* at the microscopic level. Since there is no currently known physical law that displays exact dynamics, this result suggests at least two possibilities: (1) Either the microscopic laws of physics are incorrectly formulated; or (2) There is another process operating in concert with the microscopic laws to produce exact dynamics and second law behaviour. The latter possibility is explored by looking at coarse graining, the effects of noise, and the taking of a trace. All three are rejected as likely bases for the second law of thermodynamics.

Published in "Time's Arrows, Quantum Measurement and Superluminal Behavior" (eds. C. Mugnai, A. Ranfagni, L.S. Schulman) (2001), Consiglio Nazionale Delle Ricerche (Roma), pp. 49-65

---

\*Corresponding author [tel: (514) 398-4336; fax: (514) 398-7452; e-mail [mackey@cnd.mcgill.ca](mailto:mackey@cnd.mcgill.ca); 3655 Drummond Street, Room 1124, Montreal, Quebec, CANADA H3G 1Y6]

# 1 Introduction

“The law that entropy always increases, the second law of thermodynamics, holds, I think, the supreme position among the laws of Nature.

...if your theory is found to be against the second law of thermodynamics I can give you no hope; there is nothing for it but to collapse in deepest humiliation.”

Eddington [5]

This quotation—from almost 70 years ago—captures some of the feeling surrounding the literature of the past century that attempts to reconcile the second law of thermodynamics with the nature of our microscopic physical world as it is understood.

A few years ago I wrote a review [15] followed by a small monograph [16] in which I outlined the major issues related to the understanding the second law of thermodynamics from microscopic dynamics. In this paper I repeat the essential points of that argument, correcting presentation points that were confusing to readers of [15] and [16], and emphasizing conclusions more forcefully than previously so that there can be no confusion as to the end point of the argument. My aim is to highlight the situation *vis á vis* deriving macroscopic thermodynamic behaviour from microscopic dynamics.

The outline of this paper is as follows. Section 2 explicitly defines what is meant by the second law of thermodynamics. In Section 3 I give some background material related to thermodynamics, densities, and dynamics. Section 4 defines the Gibbs entropy and shows that it is the unique definition of entropy for both equilibrium and non-equilibrium situations if the entropy is an extensive quantity. Section 5 briefly deals with maximal entropy principles and shows how they can illuminate the nature of equilibrium state when conserved quantities are known—but fail to make any commentary about the way in which those equilibrium states are reached. Section 6 considers the relation between invertible and non-invertible dynamics and entropy evolution. There it is shown that non-invertibility of dynamics is a necessary condition for second law-like behaviour. Since no current formulation of microscopic dynamics is non-invertible, this is the first clue of the problems faced by theoretical physics in trying to offer microscopic insight into the second law.

Section 7 is, in a very real sense, the *pièce de résistance* of this paper. There I offer a necessary and sufficient condition for the second law of thermodynamics to follow from microscopic dynamics. The only problem is that the condition that allows this to take place is not contained in any microscopically formulated physical law in the year 2000. Section 8 treats potential resolutions of this apparent paradox. In Section 8.1 we look at coarse graining, at the effects of noise in Section 8.2 and the effect of taking a trace in Section 8.3. Section 8.4 is very short and mentions another alternative way out of the paradox. I have relegated my *opinions* to Section 9.

I am not going to explicitly deal with the issue of the apparent unidirec-

tionality of time since other people have gone through the arguments for one or another source of this phenomena (cf. [4, 19, 22] for representative contributions). However, in providing a necessary and sufficient condition (exactness) on microscopic dynamics for the validity of the second law of thermodynamics this issue of temporal unidirectionality is solved automatically. This follows since the exact dynamics are, by necessity, non-invertible.

If, in the course of reading this paper, you come across results or statements that are not proven or inadequately explained consult [15, 16] first. If that doesn't work, then go to [13, 14].

## 2 The Second Law of Thermodynamics

Before starting our examination of the dynamics of entropy behaviour, we must first define precisely what is meant by the second law of thermodynamics. This is given in the following.

**Definition.** [Second Law] Let  $S_{TD}(t)$  be defined as the time dependent thermodynamic entropy. Then for an isolated system

$$S_{TD}(t_2) \geq S_{TD}(t_1) \quad \text{for all } t_2 > t_1, \quad (1)$$

and there is a unique steady state

$$S_{TD}^* = \lim_{t \rightarrow +\infty} S_{TD}(t) \quad (2)$$

for all initial system preparations. The entropy difference satisfies

$$\Delta S(t) \equiv S_{TD}(t) - S_{TD}^* \leq 0 \quad (3)$$

and

$$\lim_{t \rightarrow +\infty} \Delta S(t) = 0. \quad (4)$$

In other words, the system entropy evolves to a unique maximum for all system preparations. In what follows, analogs of  $S_{TD}$ ,  $S_{TD}^*$ , and  $\Delta S$  will be defined.

## 3 Starters

There are a few concepts that are well to have stated at the outset, and this section collects those together.

### 3.1 Densities and Thermodynamics

We assume that a *thermodynamic system* has states distributed in a phase space  $\mathcal{X}$ . The distribution of these states is *characterized by a (time dependent) density*  $f(t, x)$ . A *thermodynamic equilibrium* is characterized by a *stationary* (time invariant) *density*  $f_*(x)$ . Remember that  $f$  is a density if

$$f(x) \geq 0 \quad \int_{\mathcal{X}} f(x) dx = 1. \quad (5)$$

Given a phase space  $\mathcal{X}$  we will denote the space of all densities on  $\mathcal{X}$  by  $\mathcal{D}(\mathcal{X})$  or by  $\mathcal{D}$  if  $\mathcal{X}$  is understood.

### 3.2 Dynamics and Densities

We also consider a dynamics  $S_t$  operating on the phase space  $\mathcal{X}$

$$S_t : \mathcal{X} \rightarrow \mathcal{X}. \quad (6)$$

The usual way of thinking about the evolution of the dynamics (6) is through the evolution of a trajectory in the phase space  $\mathcal{X}$ . However, because we wish to examine the approach of systems to thermodynamic equilibrium, we will study how system dynamics alter densities of initial conditions rather than considering trajectories emanating from a single initial condition.

With a dynamics  $S$  and initial density  $f_0(x) = f(0, x)$  of states, the evolution of the density  $f(t, x)$  is given by

$$f(t, x) = P_S^t f_0(x), \quad (7)$$

wherein  $P_S$  is the *transfer operator* corresponding to  $S$ .

To illustrate the notion of the transfer operator governing the density evolution, consider the following examples.

1. If the dynamics are described by a system of ordinary differential equations

$$\frac{dx_i}{dt} = F_i(x) \quad i = 1, \dots, d, \quad (8)$$

then the evolution of  $f(t, x) \equiv P^t f_0(x)$  is governed by the generalized Liouville equation

$$\frac{\partial f}{\partial t} = - \sum_i \frac{\partial(f F_i)}{\partial x_i}. \quad (9)$$

2. For dynamics described by a system of stochastic differential equations

$$\frac{dx_i}{dt} = F_i(x) + \sigma(x) \xi_i, \quad i = 1, \dots, d, \quad (10)$$

$f(t, x) \equiv P^t f_0(x)$  satisfies the Fokker-Planck equation

$$\frac{\partial f}{\partial t} = - \sum_i \frac{\partial(f F_i)}{\partial x_i} + \frac{1}{2} \sum_{i,j} \frac{\partial^2(\sigma^2 f)}{\partial x_i \partial x_j}. \quad (11)$$

3. If the dynamics are described by a discrete time map  $x_{t+1} = S(x_t)$  then the evolution operator is given by the Frobenius-Perron operator

$$P_S f(x) = \frac{d}{dx} \int_{S^{-1}([0,x])} f(u) du. \quad (12)$$

4. Finally, for dynamic evolutions governed by a stochastically perturbed discrete time map  $x_{t+1} = S(x_t) + \sigma\xi_t$  then the evolution operator is a Markov operator <sup>1</sup>

$$P_S f(x) = \int_{\mathcal{X}} f(u)g(x - \sigma S(u)) du. \quad (13)$$

$g$  is the density of the distribution of the stochastic perturbation  $\xi$ .

All of these evolution operators are specific examples of Markov operators.

### 3.3 Stationary Densities

We define a fixed point  $f_*$  of the transfer operator as a *stationary density* with  $P^t f_* \equiv f_*$  for all  $t$ . A stationary density is important since it corresponds to a state of thermodynamic equilibrium. Note in particular that there is absolutely no requirement for a stationary density to be uniform across the phase space  $\mathcal{X}$ .

To illustrate how the stationary density would be determined in particular cases consider the following.

1. For the system of ordinary differential equations (8),  $f_*$  is given by the solution of

$$\sum_i \frac{\partial(f_* F_i)}{\partial x_i} = 0. \quad (14)$$

Note that the uniform density  $f_* \equiv 1$  is a stationary density of Equation 9 if and only if

$$\sum_i \frac{\partial F_i}{\partial x_i} = 0. \quad (15)$$

2. For the system of stochastic differential equations (10),  $f_*$  is the solution of

$$-\sum_i \frac{\partial(f_* F_i)}{\partial x_i} + \frac{1}{2} \sum_{i,j} \frac{\partial^2(\sigma^2 f_*)}{\partial x_i \partial x_j} = 0. \quad (16)$$

## 4 Entropy

Having postulated that a thermodynamic system has a state characterized by a density  $f$ , we are now in a position to develop the physically useful concept of entropy.

First we define an observable  $\mathcal{O}(f)$  to be a functional of the thermodynamic state characterizing some aspect of a system, for example the energy, pressure, or temperature. As such, an observable corresponds to a map  $\mathcal{O} : \mathcal{D}(\mathcal{X}) \rightarrow \mathcal{R}$ . The expected, or average, value of the observable  $\mathcal{O}(f)$  is given by weighting

---

<sup>1</sup>Any linear operator  $P : \mathcal{D} \rightarrow \mathcal{D}$  that satisfies  $P^t f \geq 0$  and  $\|P^t f\| = \|f\|$  for all  $t \in \mathcal{R}$  and  $f \in \mathcal{D}$  is called a *Markov operator*.  $\|\cdot\|$  denotes the  $L^1$  norm, and the second condition simply means that any Markov operator  $P$  operating on a density again yields a density.

$\mathcal{O}(f)$  with the system state density  $f$  and integrating over the entire phase space:

$$E(\mathcal{O}) = \langle \mathcal{O} \rangle = \int_{\mathcal{X}} \mathcal{O}(f(x)) f(x) dx. \quad (17)$$

In his seminal work Gibbs [7], assuming the existence of a system state equilibrium density  $f_*$  on the phase space  $\mathcal{X}$ , introduced the concept of the index of probability given by  $\log f_*(x)$  where “log” denotes the natural logarithm. He then identified  $-\log f_*$  with the entropy. We identify the entropy  $H$  in an equilibrium situation with the average of the index of probability

$$H(f_*) = - \int_{\mathcal{X}} f_*(x) \log f_*(x) dx, \quad (18)$$

and call this the Gibbs entropy. It can be shown that  $-\infty < H(f) \leq 0$  for all densities  $f$ .

If entropy is to be an extensive quantity (in accord with practical experience) then this definition is unique up to a multiplicative constant [11, 20]. Furthermore, (18) has repeatedly proven to yield correct results when applied to a variety of equilibrium situations. This is why it is the gold standard for equilibrium computations in statistical mechanics and thermodynamics. It is for this reason that we

*Identify the equilibrium Gibbs entropy  $H(f_*)$  with the thermodynamic entropy  $S_{TD}^*$ .*

Other contenders for the definition of entropy, such as one version of the Boltzmann entropy [9], fail to give proper answers for equilibrium calculations unless all particles are non-interacting.

The uniqueness of the entropy definition (18) (under the assumption that the entropy is an extensive quantity) is so important that it is worthwhile to give the proof. It is short.

Consider two systems  $A$  and  $B$  operating in the phase spaces  $\mathcal{X}_A$  and  $\mathcal{X}_B$  respectively, and each having the densities of states  $f_A$  and  $f_B$ . We combine the two systems to form a new system  $C$  operating in the product space  $\mathcal{X}_C = \mathcal{X}_A \times \mathcal{X}_B$ , so system  $C$  will have a density of states  $f_C(x, y) = f_A(x)f_B(y)$  if  $A$  and  $B$  do not interact. On experimental grounds we require that when the two systems are combined into a larger system  $C$ , then the entropy of system  $C$  should equal the sum of the individual entropies of  $A$  and  $B$ , since entropy is an extensive (additive) system property. We wish to show that the Gibbs choice for the index of probability is the only choice (up to a multiplicative constant) that will ensure this.

Assume that the index of probability is left as an unspecified observable  $\mathcal{O}(f)$ . If the observable  $\mathcal{O}(f)$  is such that it transforms products to sums,  $\mathcal{O}(f_C) = \mathcal{O}(f_A f_B) = \mathcal{O}(f_A) + \mathcal{O}(f_B)$ , then the relation  $H(f_A) + H(f_B) = H(f_C)$

holds. It is clear that picking  $\mathcal{O}(w) = d \log w$ , where  $d$  is any arbitrary non-zero constant, will work but are there any other functions  $\mathcal{O}$  with the requisite property?

Assume there exists a second continuous observable  $\tilde{\mathcal{O}}(f)$  such that

$$\tilde{\mathcal{O}}(f_A f_B) = \tilde{\mathcal{O}}(f_A) + \tilde{\mathcal{O}}(f_B). \quad (19)$$

Define two new functions  $v_A(a)$  and  $v_B(b)$  through

$$f_A(a) = e^{v_A(a)} \quad \text{and} \quad f_B(b) = e^{v_B(b)}. \quad (20)$$

Then we have

$$\tilde{\mathcal{O}}(e^{v_A+v_B}) = \tilde{\mathcal{O}}(e^{v_A}) + \tilde{\mathcal{O}}(e^{v_B}), \quad (21)$$

or with  $h(w) \equiv \tilde{\mathcal{O}}(e^w)$  this becomes

$$h(v_A + v_B) = h(v_A) + h(v_B). \quad (22)$$

This is the Cauchy functional equation that has the unique solution  $h(w) = \delta w$  with  $\delta$  an arbitrary constant [12]. This implies that  $\tilde{\mathcal{O}}(e^w) = \delta w$  so

$$\tilde{\mathcal{O}}(w) = \delta \log w. \quad (23)$$

Thus the observable that gives the requisite additive property for the Gibbs entropy is the logarithmic function and it is unique up to a multiplicative constant.

The question of how a non-equilibrium entropy should be defined has plagued investigators for over a century. However, what is clear is that

*The definition of non-equilibrium entropy must agree with the Gibbs definition at equilibrium.*

An examination of the proof of the uniqueness of the Gibbs definition of entropy shows that the proof applies equally well in equilibrium and non-equilibrium situations. It is for this reason that we extend the definition of the equilibrium Gibbs entropy to non-equilibrium situations and say that

*The non-equilibrium Gibbs entropy of a density  $f(t, x)$  is defined by*

$$H_t(f) = - \int_{\mathcal{X}} f(t, x) \log f(t, x) dx. \quad (24)$$

*We identify  $H_t(f)$  with the non-equilibrium thermodynamic entropy  $S_{TD}(t)$ .*

Finally, we extend these notions and define the non-equilibrium conditional entropy of the density  $f(t, x)$  with respect to a stationary density  $f_*(x)$ :

$$H_t^c(f|f_*) = - \int_{\mathcal{X}} f(t, x) \log \left[ \frac{f(t, x)}{f_*(x)} \right] dx. \quad (25)$$

As before it is the case that  $-\infty < H_t^c(f|f_*) \leq 0$  for all densities  $f$  and  $f_*$ . Notice that if the phase space  $\mathcal{X}$  is finite and the stationary density is uniform on  $\mathcal{X}$  so  $f_*(x) = 1/\mu(\mathcal{X})$  for all  $x \in \mathcal{X}$  [this is a generalization of the density of the microcanonical ensemble], then (25) reduces to  $H_t^c(f|f_*) = H(f) - \log \mu(\mathcal{X})$ . If the space  $\mathcal{X}$  is normalized then  $f_* = 1$  and  $H_t^c(f|1) = H_t(f)$  as defined in (24). Furthermore, we can write (25) in the form

$$H_t^c(f|f_*) = H(f) - H(f_*) + \int_{\mathcal{X}} [f(t, x) - f_*(x)] \log f_*(x) dx. \quad (26)$$

Thus it is clear that if there is a convergence  $\lim_{t \rightarrow \infty} f(t, x) = f_*(x)$  in some sense (which we will make totally precise in Section 7) then  $\lim_{t \rightarrow \infty} H_t^c(f|f_*) = 0$ . It is for this reason that:

*We identify the convergence of  $H_t^c(f|f_*)$  to zero [ $\lim_{t \rightarrow \infty} H_t^c(f|f_*) = 0$ ] with the convergence of the entropy difference  $\Delta S(t)$  to zero.*

## 5 Maximal Entropy Principles

Maximal entropy principles are interesting since they allow us to understand the origin of equilibrium thermodynamics. This approach [8, 10] is best illustrated by the following sample theorem that captures the essence of this point of view.

**Theorem 1** *Given an observable  $\alpha(x)$  and*

$$\langle \alpha \rangle \equiv \int_{\mathcal{X}} \alpha(x) f(x) dx, \quad (27)$$

*the maximum entropy occurs for the density given by*

$$f_*(x) = \frac{1}{Z} e^{-\nu \alpha(x)} \quad (28)$$

*where*

$$Z = \int_{\mathcal{X}} e^{-\nu \alpha(x)} \quad (29)$$

*The entropy corresponding to this density (28) is*

$$H(f_*) = \log Z + \nu \langle \alpha \rangle. \quad (30)$$



The choice of notation in Theorem 1 was deliberate, and meant to be evocative of concepts from equilibrium thermodynamics in spite of the fact that the statement of the theorem has absolutely no thermodynamic content. Namely if we identify  $Z$  with the partition function,  $\langle \alpha \rangle = U$  with the energy, take  $\nu = 1/kT$  and  $H(f_*) = S_{TD}$  (the thermodynamic entropy) as we have postulated in Section 4, then we immediately obtain a fundamental relation of equilibrium thermodynamics. Namely,

$$\begin{aligned} \text{Gibbs function} &= U - TS_{TD} \\ &= -kT \log Z \\ &= F \\ &= \text{Helmholtz free energy} \end{aligned}$$

Although maximal entropy principles allow us to comprehend a vast body of equilibrium thermodynamic phenomena by a proper identification of  $\langle \alpha \rangle$ ,  $\nu$ , etc. they make no reference to the dynamics leading to these maximal entropy states. This clearly highlights the fact that

*It is essential to understand the dynamics constraints necessary for the evolution of entropy to a maximum as required by the second law of thermodynamics (Section 2).*

## 6 Invertible and Non-invertible Dynamics

In trying to uncover the nature of the dynamics responsible for the behaviour embodied in the second law of thermodynamics, it is necessary to consider some fundamental properties of the dynamics. That is the subject of this section. We start with a consideration of invertible dynamics and then proceed to non-invertible dynamics.

### 6.1 Invertible Dynamics

A dynamics  $S_t$  is *invertible* if a trajectory evolving under the action of  $S_t$  may be run forward and backward in time unambiguously, i.e.  $S_{-t}$  applied to any point is unique. As examples of invertible dynamics we have:

1. Any system of autonomous ordinary differential equations like (8).
2. Any invertible map like  $x_{t+1} = \gamma x_t$ ,  $\mathcal{X} = R$ , so

$$S(x) = \gamma x \quad \text{and} \quad S^{-1}(x) = \frac{x}{\gamma}. \quad (31)$$

3. All of the fundamental (microscopic) laws of physics as currently formulated, e.g. Newton's laws, Hamiltonian dynamics, quantum mechanics.

Most importantly for our considerations here, the conditional entropy of any invertible system is constant and uniquely determined by the method of system preparation. This is formalized in

**Theorem 2** *If  $P^t$  is the transfer operator for an invertible system with stationary density  $f_*$ , then the entropy is constant for all time  $t$ , and equal to the value determined by  $f_*$  and the choice of the initial density  $f_0$ . That is,*

$$H^c(P^t f_0|f_*) \equiv H^c(f_0|f_*) \quad (32)$$

for all  $t$ .

**Proof.** Since  $P$  is invertible, by Voigt's theorem <sup>2</sup> [21] with  $g = f_*$  it follows that

$$H_c(P^{t+t'} f_0|f_*) = H_c(P^{t'} P^t f_0|f_*) \geq H_c(P^t f_0|f_*) \geq H_c(f_0|f_*) \quad (33)$$

for all times  $t$  and  $t'$ . Pick  $t' = -t$  so

$$H_c(f_0|f_*) \geq H_c(P^t f_0|f_*) \geq H_c(f_0|f_*) \quad (34)$$

and therefore

$$H_c(P^t f_0|f_*) = H_c(f_0|f_*) \quad (35)$$

for all  $t$ .

In particular, for the system of ordinary differential equations (8) whose density evolves according to the Liouville equation (9) we can assert that the entropy of the density  $P^t f_0$  will be constant for all time and will have the value determined by the initial density  $f_0$  with which the system is prepared. This result can also be proved directly by noting that from the definition of the entropy we may write

$$H_c(f|f_*) = - \int_{R^d} f(x) \left[ \log \left( \frac{f}{f_*} \right) + \frac{f_*}{f} - 1 \right] dx \quad (36)$$

when the stationary density is  $f_*$ . Differentiating with respect to time gives

$$\frac{dH_c}{dt} = - \int_{R^d} \frac{df}{dt} \log \left[ \frac{f}{f_*} \right] dx \quad (37)$$

or, after substituting from (9) for  $(\partial f/\partial t)$ , and integrating by parts under the assumption that  $f$  has compact support,

$$\frac{dH_c}{dt} = \int_{R^d} \frac{f}{f_*} \sum_i \frac{\partial(f_* F_i)}{\partial x_i} dx. \quad (38)$$

---

<sup>2</sup>Voigt's theorem [21] says that if  $P$  is a Markov operator, then  $H_c(P^t f_0|P^t g) \geq H_c(f|g)$  for all  $f, g \in \mathcal{D}$ .

However, since  $f_*$  is a stationary density of  $P^t$ , it is clear from (9) that

$$\frac{dH_c}{dt} = 0, \quad (39)$$

and we conclude that the conditional entropy  $H_c(P^t f_0|f_*)$  does not change from its initial value when the dynamics evolve in this manner.

The consequences of Theorem 2 in conjunction with the fact that all of the microscopic laws of physics are formulated in terms of invertible dynamics are far reaching. Namely

*The behaviour of thermodynamic systems, as embodied in the second law of thermodynamics, does not and cannot have an explanation in terms of the microscopic laws of physics as currently formulated.*

## 6.2 Non-invertible Dynamics

We say that a dynamics  $S_t$  is *non-invertible* if a trajectory evolving under the action of  $S_t$  may not be unambiguously run forward and backward in time ( $S_{-t}$  is not unique). Examples of non-invertible dynamics are

1. Stochastic differential equations like (10).
2. Differential delay equations

$$\frac{dx(t)}{dt} = \mathcal{F}(x(t), x(t - \tau)). \quad (40)$$

3. Non-invertible maps  $x_{t+1} = S(x_t)$  like the dyadic map

$$S(x) = 2x \pmod{1} \quad (41)$$

on  $\mathcal{X} = [0, 1]$ .

The importance of systems with non-invertible dynamics is that their entropy *can* increase. That is:

**Theorem 3** *If  $P^t$  is the transfer operator for a non-invertible system with stationary density  $f_*$ , then the conditional entropy satisfies*

$$H^c(f_0|f_*) \leq H^c(P^t f_0|f_*) \leq H^c(f_*|f_*) \equiv 0 \quad (42)$$

for all  $t > 0$ .

**Proof.** The demonstration is a trivial application of Voigt’s theorem in conjunction with the observation that the conditional entropy is maximized by  $f_*$ .

Where have we gotten? We have shown that the Gibbs entropy definition is unique if entropy is to be an extensive quantity (Section 4). We have also shown that the Gibbs entropy of a system cannot change if the underlying dynamics are invertible (Section 6.1, Theorem 2). Further, the Gibbs entropy *may* increase if the dynamics are non-invertible (Section 6.2, Theorem 3). This summarizes the point that matters were at when the following words were written in 1967:

“It is not very difficult to show that the combination of the reversible laws of mechanics with Gibbsian statistics does not lead to irreversibility but that the notion of irreversibility must be added as a special ingredient ...

... the explanation of irreversibility in nature is to my mind still open.”

Bergmann [2]

The result from Theorem 3 that non-invertibility is necessary, though not sufficient, for entropy to increase is well known. It has been used implicitly or explicitly over the past century (or more) in almost every attempt to reconcile the observation at the end of Section 6.1 with the indisputable thermodynamic behaviour of systems as embodied in the second law of thermodynamics. Thus (for example) to achieve non-invertibility one can “coarse grain” the phase space (cf. Section 8.1) which is a fancy name for throwing away information, one can add “noise” (cf. Section 8.2), or one can take the “trace” of the system (cf. Section 8.3) which is another fancy name for throwing away a LOT of information.

All of these attempts have their basis in the implicit rejection of the possibility that the fundamental laws of physics might, in fact, be improperly formulated. Clearly, these formulations could not be grossly in error since we see excellent agreement between them and experimental measurements in *almost all situations*. The one obvious exception is in the behaviour of entropy as systems approach equilibrium.

However, before turning our attention to these various ways of trying to deal with the obvious problem that we have highlighted up to this point, we can do more. Namely, we can specify exactly what conditions the dynamics  $S_t$  must satisfy in order for the second law of thermodynamics to be satisfied.

## 7 Pièce de résistance

We have now arrived at the point where we can state **necessary and sufficient** conditions for the operation of the second law of thermodynamics—and that is the subject of this section. This involves defining a new type of non-invertible dynamics.

## 7.1 Exact Dynamics

We now consider exact dynamics. These types of dynamics have profoundly important properties for our investigation of the origins of the second law of thermodynamics. More precisely:

**Definition.** A dynamics  $S_t$  on a phase space  $\mathcal{X}$  with transfer operator  $P_S^t$  and unique stationary density  $f_*$  is *exact* if and only if

$$\lim_{t \rightarrow \infty} \|P_S^t f(x) - f_*\|_{L^1} = 0 \quad (43)$$

for *every* initial density  $f \in \mathcal{D}$ .

Exact systems have a number of interesting properties that we will merely mention (cf. [13, 14, 16] for more complete details). First, exact systems are non-invertible. Secondly, they always have a unique stationary density  $f_*$ . Finally exactness of system dynamics also implies that they are mixing and, hence, ergodic:

$$\text{Exact} \Rightarrow \text{Mixing} \Rightarrow \text{Ergodic}. \quad (44)$$

It is important to emphasize that the reverse implications *do not* hold (consult Section 8.1 for the definitions of ergodicity and mixing).

As an example, the dyadic map (41) is exact with unique stationary density

$$f_*(x) = 1 \quad \text{for all } x \in [0, 1] \quad (45)$$

Another example is the quadratic map  $x_{t+1} = S(x_t)$  with  $S(x) = 4x(1-x)$  on  $\mathcal{X} = [0, 1]$ , which has a unique stationary density

$$f_*(x) = \frac{1}{\pi \sqrt{x(1-x)}} \quad x \in [0, 1]. \quad (46)$$

Is exactness enough to offer the explanation that Bergmann found missing? Indeed so.

## 7.2 Exact Dynamics and Entropy Evolution

The core of this section, and indeed this entire paper, is contained in the following theorem.

**Theorem 4 (Lasota & Mackey)** [13, 14] *The conditional entropy evolves to its maximum value of zero for all system preparations  $f_0 \in \mathcal{D}$ ,*

$$\lim_{t \rightarrow \infty} H_t^c(f_0|f_*) = 0 \quad (47)$$

**if and only if**  $S_t$  is exact.

**Proof.** The proof is too long and would require the introduction of too many other details to make a presentation here sensible. If you are interested you can consult ([16], Theorem 7.7, pp. 98-100).

Thus:

*In constructing a microscopic explanation of the second law of thermodynamics Theorem 4 directs our attention solely to systems with exact dynamics.*

### 7.3 Implications

Since the second law of thermodynamics requires (non-invertible) exact system dynamics and none of the currently formulated microscopic dynamics in physics have this property, there is clearly a major problem. What are the possible solutions? There are at least two:

- Possibility 1. Some other process, in combination with the microscopic laws of physics, makes them exact and produces non-invertible (irreversible) thermodynamic behaviour.
- Possibility 2. The microscopic laws of physics are incorrectly formulated.

We will explore Possibility 1 in the next section, considering three ways out of the box that Theorem 4 has put us in. None of these three are new—indeed they were grasped at early on as a way out of the problems that physics had justifying the second law. However, what was not recognized was that they can, under certain circumstances, provide that element of exactness that Theorem 4 requires. I give all three of these without editorial comment. Possibility 2 will be dealt with even more briefly in Section 8.4.

## 8 Potential Resolutions

In this section we consider the effects of coarse graining, noise, and taking a trace on the behaviour of entropy and their possible bearing on the origin of the second law of thermodynamics.

### 8.1 Coarse Graining

Coarse graining was introduced by Boltzmann [3] in his search for a mechanism whereby entropy behaviour could be made to conform to the second law. It was later considered by Gibbs [7] and the Ehrenfests [6].

Coarse graining is carried out by first partitioning the phase space  $\mathcal{X}$  into cells  $A_i$  that satisfy

$$\bigcup_i A_i = \mathcal{X} \quad \text{and} \quad A_i \cap_{i \neq j} A_j = \emptyset. \quad (48)$$

There is no unique way this partition may be formed, but we require that each cell of the partition have positive measure,  $\mu_L(A_i) > 0$ . For every density  $f$ ,

within each cell of this partition we denote the average of  $f$  over  $A_i$  by  $\langle f \rangle_i$ ,

$$\langle f \rangle_i = \frac{1}{\mu_L(A_i)} \int_{A_i} f(x) dx, \quad (49)$$

so the density  $f$  coarse grained with respect to the partition  $A_i$  is given by

$$f^{cg}(x) = \sum_i \langle f \rangle_i 1_{A_i}(x). \quad (50)$$

Thus,  $f^{cg}$  is constant within each cell  $A_i$ . Clearly  $\sum_i \langle f \rangle_i \mu_L(A_i) = 1$ . Given this partition, a density  $f$ , and a coarse grained density  $f^{cg}$  then the Gibbs entropy of the coarse grained density  $f^{cg}$  is given by

$$H(f^{cg}) = - \sum_i \langle f \rangle_i \mu_L(A_i) \log \langle f \rangle_i, \quad (51)$$

and it is easy to show ([16], Theorem 8.1) that  $H(f) \leq H(f^{cg})$ .

It is also easy to show that the conditional entropy of  $f^{cg}$  with respect to a stationary density  $f_*$ , also coarse grained with respect to the same partition is given by

$$H_c(f^{cg}|f_*) = - \sum_i \langle f \rangle_i \mu_L(A_i) \log \left[ \frac{\langle f \rangle_i}{\langle f_* \rangle_i} \right], \quad (52)$$

and that  $H(f|f_*) \leq H_c(f^{cg}|f_*)$  [16].

Thus, coarse graining will either increase the entropy or leave it unchanged. Is there any circumstance in which we can assert something stronger about the behaviour of the entropy after coarse graining? Indeed there is, and this leads us to consider mixing systems which appeared briefly and without definition in Section 7.1.

Mixing is a dynamical property stronger than ergodicity and weaker than exactness, and it is equivalent to a weak convergence property [13, 14]. Namely: **Definition.** A dynamics  $S_t$  on a phase space  $\mathcal{X}$  with transfer operator  $P_S^t$  and unique stationary density  $f_*$  is mixing if and only if

$$\lim_{t \rightarrow \infty} \langle P_S^t f_0, g \rangle = \langle f_*, g \rangle, \quad (53)$$

where  $g$  is any bounded measurable function and

$$\langle f, g \rangle = \int_{\mathcal{X}} f(x)g(x)dx \quad (54)$$

denotes the scalar product.

As an example of a mixing dynamics, consider the baker transformation:

$$S(x, y) = \begin{cases} (2x, \frac{1}{2}y) & 0 \leq x \leq \frac{1}{2} \\ (2x-1, \frac{1}{2} + \frac{1}{2}y) & \frac{1}{2} < x \leq 1. \end{cases} \quad (55)$$

The baker transformation has a unique stationary density that is uniform,

$$f_*(x, y) \equiv 1 \quad \text{for all } x, y \in [0, 1] \times [0, 1], \quad (56)$$

is mixing, and is invertible. Thus we know from Theorem 2 that the entropy  $H^c(P^t f_0 | f_*)$  is absolutely constant and equal to the entropy with which the system is prepared. However, coarse graining the baker transformation does wonderful things for the entropy behaviour, and this is why it has been so popular ever since Boltzmann [3] first introduced the concept (Ruelle [18] has an especially lucid description of Boltzmann's approach which is far more comprehensible than much of what is written about Boltzmann's ideas). These wonderful things are shown in the next theorem.

**Theorem 5** *If  $P^t$  is the transfer operator corresponding to an invertible mixing dynamics with a unique stationary density  $f_*$ , then for all non-trivial coarse graining of the phase space*

$$\lim_{t \rightarrow \pm\infty} H_t^c(f_0^{cg} | f_*^{cg}) = 0 \quad (57)$$

for all initial densities  $f_0$ .

As long as we have gotten as far as defining exactness and mixing, we might as well finish the job and define ergodicity.

**Definition.** A dynamics  $S_t$  on a phase space  $\mathcal{X}$  with transfer operator  $P_S^t$  and unique stationary density  $f_*$  is ergodic if and only if  $\{P^t f_0\}$  is Cesàro convergent to  $f_*$  for all initial densities  $f_0$ , i.e., if

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \langle P^t f_0, g \rangle dt = \langle f_*, g \rangle \quad (58)$$

where  $g$  is any bounded measurable function.

Note that:

- The property of ergodicity is *weaker* than the property of mixing; and
- There are no *known* formulations of microscopic physical laws that are ergodic, much less mixing.

## 8.2 Noise

Because of the difficulty of dealing with stochastic systems, I'm not going to go into the details of the effects of noise. I simply point out that the injection of noise into a system with continuous or discrete time invertible dynamics *can* induce second law like behaviour [16].

## 8.3 Traces

As an alternative to coarse graining or the addition of noise, I now turn to the consequences of having an invertible dynamics in which not all dynamical variables are observable. This means that we have a dynamical system operating in an  $n$ -dimensional space, but are able to observe only  $m < n$  of these variables. That is, we observe only a *trace* of its operation in an  $m$ -dimensional space



because  $(n - m)$  of the variables are hidden to us, *e.g.* because either we do not know about them, or do not have the technology to measure them. Here I give just two illustrations, and [16] can be consulted for more details.

### 8.3.1 x Trace of the Baker Transformation

The baker transformation was introduced in Equation 55 as an example of an invertible dynamics that was mixing. Suppose we cannot measure the  $y$  variable in the baker transformation. Then we just measure a sequence of  $x$  values given by the dyadic map defined in Equation 41. This is an example of taking a trace.

In Section 7.1 I asserted that the dyadic map was exact with a uniform stationary density. Therefore by Theorem 4 we know that the Gibbs entropy will smoothly approach zero. Thus by taking a trace in this fashion we have passed from a situation in which the entropy is absolutely constant (the full baker transformation) to one in which the entropy evolution has second law behaviour.

### 8.3.2 y Trace of the Baker Transformation

Now suppose we cannot measure the  $x$  variable in the baker transformation (55). Then we have a sequence of  $y$  values given by

$$y_{t+1} = \frac{1}{2}y_t + \xi_t \quad (59)$$

where

$$\xi_t = \frac{1}{2}1_{[\frac{1}{2}, 1]}(x_t) \quad (60)$$

and

$$x_{t+1} = 2x_t \pmod{1}. \quad (61)$$

This trace system is mathematically equivalent to adding noise to the stable system  $y_{t+1} = \frac{1}{2}y_t$ , and in this case  $\xi_t$  can be interpreted as playing the role of a hidden variable. From ([16], Theorem 10.9) the entropy will monotonically approach zero in this case. Thus, if we lived in a  $y$  universe we would conclude that the entropy always goes to zero regardless of how we prepare the system.

### 8.3.3 Two Amusing (but true) Asides Concerning Traces

**Theorem 6** (*God Theorem [13, 14, 16]*) *Every trajectory in a space  $\mathcal{X}$  is the trace of a single system operating in another space  $\mathcal{Y}$ .*

**Theorem 7** (*Rochlin's Theorem [13, 14, 16, 17]*) *Every "sub-universe" in which the second law of thermodynamics holds is the trace of a "master universe" in which the entropy is constant.*

## 8.4 Possibility 2

The other possibility I raised at the end of Section 7.3 was that all of the microscopic laws of physics are incorrectly formulated. If true, then the violation of these laws is very minute. I don't have much to say on this matter except to point out that experimentally the dynamics of the neutral kaon system are not invertible [1]. However we have no evidence one way or the other concerning the possibility that they are exact.

## 9 Opinions

In keeping with my attempt to relegate all opinions (as opposed to what can be proven) to a separate section I have collected them here. As far as I know there is no scientific evidence to validate or invalidate my opinions any more or less than contrary opinions held by others.

### 9.1 Coarse Graining is Not the Origin of Second Law Behaviour

Notice in Theorem 5 that the entropy approach to a maximum occurs for  $t \rightarrow \pm\infty$ . Thus coarse graining an invertible mixing system causes the entropy to go to zero independent of the direction of the time. Hence, the entropy increase induced by coarse graining:

- Fails to single out any preferred direction of time;
- May not be monotone [16]; and
- Has a rate of convergence that depends on the way the coarse graining is carried out.

This latter feature implies that if entropy increases to a maximum because we have invertible mixing dynamics but there is coarse graining then the rate of convergence of the entropy (and all other thermodynamic variables) to equilibrium should become slower as measurement techniques improve. This behaviour has not been observed to the best of my knowledge.

### 9.2 Noise is Not the Origin of Second Law Behaviour

If we attribute second law behaviour to the presence of noise, we must inquire about the origin of the noise.

- If the dynamics of the universe are a composition of both deterministic and stochastic elements, then there is no further discussion.
- However I believe that the universe dynamics are deterministic.

- If universe dynamics are purely deterministic, then we must interpret this “noise” as the signature of the deterministic dynamics of a portion of the universe, which we do not observe, impinging on the portion we are studying.
- If the universe dynamics are not only deterministic but also invertible, then the entropy of the universe should be constant.
- However, if the entropy of the universe is constant but we measure an entropy increase in an observed portion, then we must conclude that there is a corresponding entropy decrease in the unobserved portion.
- If we continue to see second law type behaviour as we expand the size of the observed portion of the universe, we arrive at a contradiction.

### 9.3 Second Law Behaviour is Not the Result of a Trace as Discussed Here

If second law like behaviour were to be the result of taking a trace, what would the most likely requirements be? I would insist that the entropy of the master system dynamics obey the second law of thermodynamics. I take this position because I find it inconceivable that the behaviour we observe in our everyday world is only induced by taking a trace—and that the entropy of the entire universe is constant (cf. the argument in Section 9.4 below). Thus although Rochlin’s theorem is true I don’t think it’s applicable.

Rather I think that if traces are important to understanding second law behaviour then:

- The dynamics that we *think* are operating should appear to be invertible within experimental error just as our own microscopically formulated dynamics appear to be.
- However there would have to be some tiny and minute process (experimentally undetected at the moment) that renders these apparently invertible dynamics **noninvertible and exact**. Thus, for example, we would be looking for a situation somewhat like the  $y$  trace of the baker transformation that was considered in Section 8.3.2.
- These dynamics should be the trace of a universal dynamics that are also non-invertible and exact so that the entropy under the influence of the universal dynamics also obeys the second law.

I would also apply precisely the same arguments against the possibilities raised in the previous Sections 9.1 and 9.2 on coarse graining and noise respectively.

### 9.4 A Thought Experiment

Consider the following thought experiment. Put a cat in a **completely sealed** box one meter on a side without food or water (Schulman [19] has remarked

that cats seem to often be the target of bizarre thought experiments). After one month, return and open the box. What will you find? The cat will, of course, be dead. The processes that led to this state of affairs are completely in accord with the operation of the second law of thermodynamics.

Are we seriously to believe that this death was a consequence of coarse graining? Of a noisy environment? Of taking a trace? To me this is so patently ridiculous that I reject it out of hand. Animals do not die in such circumstances because of our ignorance of dynamics (coarse graining or traces) or because of noise.

## 9.5 What Do I Think Is REALLY Going On?

I think that the microscopic laws of physics are incorrectly formulated in the following sense.

- They omit something so minute that it is experimentally undetected (at the moment) but that nevertheless forces second law of thermodynamics compliance which is manifested daily in our observation of the world around us.
- If this opinion is correct, then identifying the nature and source of this deviation is a major challenge for the physics of the future.

## Acknowledgments

I would like to thank Profs. D. Mugnai, A. Ranfagni, and L.S. Schulman for the invitation to participate in the “Conference on Time’s Arrows, Quantum Measurement and Superluminal Behavior” in Napoli and to the other participants there for their interest and collegiality. It was a most enjoyable encounter.

This work was supported by the Hanse Wissenschaftskolleg, the Mathematics of Information Technology and Complex Systems (MITACS, Canada), the Natural Sciences and Engineering Research Council (NSERC grant OGP-0036920, Canada), the Alexander von Humboldt Stiftung, and Le Fonds pour la Formation de Chercheurs et l’Aide à la Recherche (FCAR grant 98ER1057, Québec). I thank the Hanse Wissenschaftskolleg, Delmenhorst, Germany and the Institute für Theoretische Physik, Universität Bremen, Germany for their hospitality during the time this work was completed. I have especially benefited from conversations with Prof. Dr. Helmut Schwegler, Bremen, on these topics for more than 15 years.

## References

- [1] A. Angelopoulos, A. Apostolakis, E. Aslanides, G. Backenstoss, P. Bargaña, O. Behnke, A. Benelli, V. Bertin, F. Blanc, P. Bloch, P. Carlson, M. Carroll, E. Cawley, S. Charalambous, M.B. Chertok, M. Danielsson,

- M. Dejardin, J. Derre, A. Ealet, C. Eleftheriadis, L. Faravel, W. Fetscher, M. Fidecaro, A. Filipi, D. Francis, J. Fry, E. Gabathuler, R. Gamet, H.-J. Gerber, A. Go, A. Haselden, P.J. Hayman, F. Henry-Couannier, R.W. Hollander, K. Jon-And, P.-R. Kettle, P. Kokkas, R. Kreuger, R. Le Gac, F. Leimgruber, I. Mandi, N. Manthos, G. Marel, M. Miku, J. Miller, F. Montanet, A. Muller, T. Nakada, B. Pagels, I. Papadopoulos, P. Pavlopoulos, A. Policarpo, G. Polivka, R. Rickenbach, B.L. Roberts, T. Ruf, C. Santoni, M. Schfer, L.A. Schaller, T. Schietinger, A. Schopper, L. Tauscher, C. Thibault, F. Touchard, C. Touramanis, C.W.E. Van Eijk, S. Vlachos, P. Weber, O. Wigger, M. Wolter, D. Zavrtanik, and D. Zimmerman. First direct observation of time reversal non invariance in the neutral kaon system. *Phys. Lett. B*, 444:43–51, 1998.
- [2] P.G. Bergman. Foundations research in physics. In M. Bunge, editor, *Delaware Seminar in the Foundations of Physics*, pages 1–14, New York, 1967. Springer-Verlag.
- [3] L. Boltzmann. *Wissenschaftliche Abhandlungen*. Barth, Leipzig, 1909.
- [4] P.C.W. Davies. *The Physics of Time Asymmetry*. University of California Press, Berkeley and Los Angeles, 1974.
- [5] A. Eddington. *The Nature of the Physical World*. E.P. Dutton, New York, 1935.
- [6] P. Ehrenfest and T. Ehrenfest. *The Conceptual Foundations of the Statistical Approach in Mechanics*. Cornell University Press, Ithaca, 1959.
- [7] J.W. Gibbs. *Elementary Principles in Statistical Mechanics*. Yale University Press, New Haven, 1902.
- [8] E.T. Jaynes. Information theory and statistical mechanics. *Phys. Rev.*, 106:620–630, 1957.
- [9] E.T. Jaynes. Gibbs vs Boltzmann Entropies. *Amer. J. Physics*, 33:391–398, 1965.
- [10] A. Katz. *Principles of Statistical Mechanics*. Freeman, San Francisco, 1967.
- [11] A.I. Khinchin. *Mathematical Foundations of Statistical Mechanics*. Dover, New York, 1949.
- [12] M. Kuczma. *An Introduction to the Theory of Functional Equations and Inequalities*. Silesian University Press, Katowice (Poland), 1985.
- [13] A. Lasota and M.C. Mackey. *Probabilistic Properties of Deterministic Systems*. Cambridge University Press, Cambridge, 1985.
- [14] A. Lasota and M.C. Mackey. *Chaos, Fractals and Noise: Stochastic Aspects of Dynamics*. Springer-Verlag, New York, 1994.

- [15] M.C. Mackey. The dynamic origin of increasing entropy. *Rev. Mod. Phys.*, 61:981–1016, 1989.
- [16] M.C. Mackey. *Time's Arrow: The Origins of Thermodynamic Behavior*. Springer-Verlag, New York, 1992.
- [17] V.A. Rohlin. Exact endomorphisms of Lebesgue spaces. *Amer. Math. Soc. Transl. (2)*, 39:1–36, 1964.
- [18] D. Ruelle. Smooth dynamics and new theoretical ideas in nonequilibrium statistical mechanics. *J. Stat. Phys.*, 95:393–468, 1999.
- [19] L. S. Schulman. *Time's Arrows and Quantum Measurement*. Cambridge University Press, Cambridge, 1997.
- [20] B.S. Skagerstam. On the mathematical definition of entropy. *Z. Naturforsch.*, 29a:1239–1243, 1974.
- [21] J. Voigt. Stochastic operators, information, and entropy. *Commun. Math. Phys.*, 81:31, 1981.
- [22] H.D. Zeh. *The Physical Basis of the Direction of Time*. Springer-Verlag, Heidelberg, 1992.