

9. Irreversibility in stochastic dynamics *

Jos Uffink

9.1 Introduction

Over recent decades, some approaches to non-equilibrium statistical mechanics, that differ decidedly in their foundational and philosophical outlook, have nevertheless converged in developing a common unified mathematical framework. I will call this framework ‘stochastic dynamics’, since the main characteristic feature of the approach is that it characterizes the evolution of the state of a mechanical system as evolving under stochastic maps, rather than under a deterministic and time-reversal invariant Hamiltonian dynamics.¹

The motivations for adopting this stochastic type of dynamics come from at least three different viewpoints.

1. “Coarse graining”(cf. van Kampen 1962, Penrose 1970): In this view one assumes that on the microscopic level the system can be characterized as a (Hamiltonian) dynamical system with deterministic time-reversal invariant dynamics. However, on the macroscopic level, one is only interested in the evolution of macroscopic states, i.e. in a partition (or coarse graining) of the microscopic phase space into discrete cells. The usual idea is that the form and size of these cells are chosen in accordance with the limits of our observational capabilities.

On the macroscopic level, the evolution now need no longer be portrayed as deterministic. When only the macrostate of a system at an instant is given, it is in general not fixed what its later macrostate will be, even if the underlying microscopic evolution is deterministic. Instead, one can provide *transition probabilities*, that specify how probable the transition from any given initial macrostate to later macrostates is. Although it is impossible, without further assumptions, to say anything general about the evolution of the macroscopically characterized states, it is possible to describe the evolution of an ensemble or a probability distribution over these states, in terms of a *stochastic process*.

2. “Interventionism”, or “open systems” (cf. Blatt 1959, Davies 1976, Lindblad 1976, Lindblad 1983, Ridderbos 2002): On this view, one assumes that the system to be described is not isolated but in (weak) interaction with the environment. It is assumed that the total system, consisting of the system of interest and the environment can be described as a (Hamiltonian) dynamical system with a time-reversal invariant and deterministic dynamics. If we represent the state of the system by $x \in \Gamma^{(s)}$ and that of the environment by $y \in \Gamma^{(e)}$, their joint evolution is

⁰This paper is an excerpt of a longer discussion in (Uffink 2007).

¹Also, the name has already been used for precisely this approach by Sudarshan and coworkers, cf. (Sudarshan et al. 1961, Mehra & Sudarshan 1972).

given by a one-parameter group of evolution transformations, generated from the Hamiltonian equations of motion for the combined system: $U_t : (x, y) \mapsto U_t(x, y) \in \Gamma^{(s)} \times \Gamma^{(e)}$. The evolution of the state x in the course of time is obtained by projecting, for each t , to the coordinates of $U_t(x, y)$ in $\Gamma^{(s)}$; call the result of this projection x_t . Clearly, this reduced time evolution of the system will generally fail to be deterministic, e.g. the trajectory described by x_t in $\Gamma^{(s)}$ may intersect itself. Again, we may hope that this indeterministic evolution can nevertheless, for an ensemble of the system and its environment, be characterized as a stochastic process, at least if some further reasonable assumptions are made.

3. A third viewpoint is to deny (Mackey, 1992, 2001), or to remain agnostic about (Streater 1995), the existence of an underlying deterministic or time-reversal invariant dynamics, and simply regard the evolution of a system as described by a stochastic process as a new fundamental form of dynamics in its own right.

While authors in this approach thus differ in their motivation and in the interpretation they have of its subject field, there is a remarkable unity in the mathematical formalism adopted for this form of non-equilibrium statistical mechanics. The hope, obviously, is to arrange this description of the evolution of mechanical systems in terms of a stochastic dynamics in such a way that the evolution will typically display ‘irreversible behaviour’: i.e. an ‘approach to equilibrium’, that a Boltzmann-like evolution equation holds, that there is a stochastic analogy of the H -theorem, etc. In short, one would like to recover the autonomy and irreversibility that thermal systems in non-equilibrium states typically display.

We will see that much of this can apparently be achieved with relatively little effort once a crucial technical assumption is in place: that the stochastic process is in fact a homogeneous Markov process, or, equivalently, obeys a so-called master equation. Much harder are the questions of whether the central assumptions of this approach might still be compatible with an underlying deterministic time-reversal invariant dynamics, and in which sense the results of the approach embody time-asymmetry. In fact we shall see that conflicting intuitions on this last issue arise, depending on whether one takes a probabilistic or a dynamics point of view towards this formalism.

From a foundational point of view, stochastic dynamics promises a new approach to the explanation of irreversible behaviour that differs in interesting ways from the more orthodox Hamiltonian or dynamical systems approach. In that approach, any account of irreversible phenomena can only proceed by referring to special initial conditions or dynamical hypotheses. Moreover, it is well-known that an ensemble of such systems will conserve (fine-grained) Gibbs entropy so that the account cannot rely on this form of entropy for a derivation of the increase of entropy. In stochastic dynamics, however, one may hope to find an account of irreversible behaviour that is not tied to special initial conditions, but one that is, so to say, built into the very stochastic-dynamical evolution. Further, since Liouville’s theorem is not applicable, there is the prospect that one can obtain a genuine increase of Gibbs entropy from this type of dynamics.

As just mentioned, the central technical assumption in stochastic dynamics is that the processes described

have the Markov property.² Indeed, general aspects of irreversible behaviour pour out almost effortlessly from the Markov property, or from the closely connected “master equation”. Consequently, much of the attention in motivating stochastic dynamics has turned to the assumptions needed to obtain this Markov property, or slightly more strongly, to obtain a non-invertible Markov process (Mackey 1992). The best-known specimen of such an assumption is van Kampen’s (1962) “repeated randomness assumption”. And similarly, critics of this type of approach (Sklar 1993, Redhead 1995, Callender 1999) have also focused their objections on the question just how reasonable and general such assumptions are (cf. § 9.5).

I believe both sides of the debate have badly missed the target. Many authors have uncritically assumed that the assumption of a (non-invertible) Markov process does indeed lead to non-time-reversal-invariant results. As a matter of fact, however, the Markov property (for invertible or non-invertible Markov processes) is time-reversal invariant. So, any argument to obtain that property need not presuppose time-asymmetry. In fact, I will argue that this discussion of irreversible behaviour as derived from the Markov property suffers from an illusion. It is due to the habit of studying the prediction of future states from a given initial state, rather than studying retrodictions towards an earlier state. As we shall see, for a proper description of irreversibility in stochastic dynamics one needs to focus on another issue, namely the difference between backward and forwards transition probabilities.

In the next sections, I will first (§9.2) recall the standard definition of a homogeneous Markov process from the theory of stochastic processes. Section 9.3 casts these concepts in the language of dynamics, introduces the master equation, and discusses its analogy to the Boltzmann equation. In §9.4, we review some of the results that *prima facie* display irreversible behaviour for homogeneous Markov processes. In section 9.5 we turn to the physical motivations that have been given for the Markov property, and their problems, while §9.6.2 focuses on the question how seemingly irreversible results could have been obtained from a time-symmetric assumptions. Finally, §9.7.2 argues that a more promising discussion of these issues should start from a different definition of reversibility of stochastic processes.

9.2 The definition of Markov processes

To start off, consider a paradigmatic example. One of the oldest discussions of a stochastic process in the physics literature is the so-called ‘dog flea model’ of P. and T. Ehrenfest (1907).

Consider N fleas, labeled from 1 to N , situated on either of two dogs. The number of fleas on dog 1 and 2 are denoted as n_1 and $n_2 = N - n_1$. Further, we suppose there is an urn with N lots carrying the numbers $1, \dots, N$ respectively. The urn is shaken, a lot is drawn (and replaced), and the flea with the corresponding label is ordered to jump to the other dog. This procedure is repeated every second.

It is not hard to see that this model embodies an ‘approach equilibrium’ in some sense: Suppose that initially all or almost all fleas are on dog 1. Then it is very probable that the first few drawings will move fleas from dog

²Some authors argue that the approach can and should be extended to include non-Markovian stochastic processes as well. Nevertheless I will focus here on Markov processes.

1 to 2. But as soon as the number of fleas on dog 2 increases, the probability that some fleas will jump back to dog 1 increases too. The typical behaviour of, say, $|n_1 - n_2|$ as a function of time will be similar to Boltzmann's H -curve, with a tendency of $|n_1 - n_2|$ to decrease if it was initially large, and to remain close to the 'equilibrium' value $n_1 \approx n_2$ for most of the time. But note that in contrast to Boltzmann's H -curve in gas theory, the 'evolution' is here entirely stochastic, i.e. generated by a lottery, and that no underlying deterministic equations of motion are provided.

Generally speaking, a stochastic process is defined as a probability measure P on a measure space X , whose elements will here be denoted as ξ , on which there are infinitely many random variables Y_t , with $t \in \mathbb{R}$ (or sometimes $t \in \mathbb{Z}$). Physically speaking, we interpret t as time, and Y as the macroscopic variable(s) characterizing the macrostate —say the number of fleas on a dog, or the number of molecules with their molecular state in some cell of μ -space, etc. Further, ξ represents the total history of the system which determines the values of $Y_t(\xi)$. The collection Y_t may thus be considered as a single random variable Y evolving in the course of time.

At first sight, the name 'process' for a probability measure may seem somewhat unnatural. From a physical point of view it is the *realization*, in which the random variables Y_t attain the values $Y_t(\xi) = y_t$ that should be called a process. In the mathematical literature, however, it has become usual to denote the measure that determines the probability of all such realizations as a 'stochastic process'.

For convenience we assume here that the variables Y_t may attain only finitely many discrete values, say $y_t \in \mathcal{Y} = \{1, \dots, m\}$. However, the theory can largely be set up in complete analogy for continuous variables.

The probability measure P provides, for $n = 1, 2, \dots$, and instants t_1, \dots, t_n definite probabilities for the event that Y_t at these instants attains certain values y_1, \dots, y_n :

$$\begin{aligned}
 & P_{(1)}(y_1, t_1) \\
 & P_{(2)}(y_2, t_2; y_1, t_1) \\
 & \quad \vdots \\
 & P_{(n)}(y_n, t_n; \dots; y_1, t_1) \\
 & \quad \vdots
 \end{aligned} \tag{1}$$

Here, $P_{(n)}(y_n, t_n; \dots; y_1, t_1)$ stands for the joint probability that at times t_1, \dots, t_n the quantities Y_t attain the values y_1, \dots, y_n , with $y_i \in \mathcal{Y}$. It is an abbreviation for

$$P_{(n)}(y_n, t_n; \dots; y_1, t_1) := P(\{\xi \in X : Y_{t_n}(\xi) = y_n \ \& \ \dots \ \& \ Y_{t_1}(\xi) = y_1\}) \tag{2}$$

Obviously the probabilities (1) are normalized and non-negative, and each $P_{(n)}$ is a marginal of all higher-order

probability distributions:

$$P_{(n)}(y_n, t_n; \dots; y_1, t_1) = \sum_{y_{n+m}} \cdots \sum_{y_{n+1}} P_{(n+m)}(y_{n+m}, t_{n+m}; \dots; y_1, t_1). \quad (3)$$

In fact, the probability measure P is uniquely determined by the hierarchy (1).³

Similarly, we may define conditional probabilities in the familiar manner, e.g.:

$$P_{(1|n-1)}(y_n, t_n | y_{n-1}, t_{n-1}; \dots; y_1, t_1) := \frac{P_{(n)}(y_n, t_n; \dots; y_1, t_1)}{P_{(n-1)}(y_{n-1}, t_{n-1}; \dots; y_1, t_1)} \quad (4)$$

provides the probability that Y_{t_n} attains the value y_n , under the condition that $Y_{t_{n-1}}, \dots, Y_{t_1}$ have the values y_{n-1}, \dots, y_1 . In principle, the times appearing in the joint and conditional probability distributions (1,4) may be chosen in an arbitrary order. However, we adopt from now on the convention that they are ordered as $t_1 < \dots < t_n$.

A special and important type of stochastic process is obtained by adding the assumption that such conditional probabilities depend only the condition at the last instant. That is to say: for all n and all choices of y_1, \dots, y_n and $t_1 < \dots < t_n$:

$$P_{(1|n)}(y_n, t_n | y_{n-1}, t_{n-1}; \dots; y_1, t_1) = P_{(1|1)}(y_n, t_n | y_{n-1}, t_{n-1}) \quad (5)$$

This is the *Markov property* and such stochastic processes are called *Markov processes*.

The interpretation often given to this assumption, is that Markov processes have ‘no memory’. To explain this slogan more precisely, consider the following situation. Suppose we are given a piece of the history of the quantity Y : at the instants t_1, \dots, t_{n-1} its values have been y_1, \dots, y_{n-1} . On this information, we want to make a prediction of the value y_n of the variable Y at a later instant t_n . The Markov-property (5) says that this prediction would not have been better or worse if, instead of knowing this entire piece of prehistory, only the value y_{n-1} of Y at the last instant t_{n-1} had been given. Additional information about the past values is thus irrelevant for a prediction of the future value.

For a Markov process, the hierarchy of joint probability distributions (1) is subjected to stringent demands. In fact they are all completely determined by: (a) the specification of $P_{(1)}(y, 0)$ at one arbitrary chosen initial instant $t = 0$, and (b) the conditional probabilities $P_{(1|1)}(y_2, t_2 | y_1, t_1)$ for all $t_2 > t_1$. Indeed,

$$P_{(1)}(y, t) = \sum_{y_0} P_{(1|1)}(y, t | y_0, 0) P_{(1)}(y_0, 0); \quad (6)$$

and for the joint probability distributions $P_{(n)}$ we find:

$$\begin{aligned} P_{(n)}(y_n, t_n; \dots; y_1, t_1) &= P_{(1|1)}(y_n, t_n | y_{n-1}, t_{n-1}) P_{(1|1)}(y_{n-1}, t_{n-1} | y_{n-2}, t_{n-2}) \times \\ &\times \cdots \times P_{(1|1)}(y_2, t_2 | y_1, t_1) P_{(1)}(y_1, t_1). \end{aligned} \quad (7)$$

³At least, when we assume that the σ -algebra of measurable sets in X is the cylinder algebra generated by sets of the form in the right-hand side of (2).

It follows from the Markov property that the conditional probabilities $P_{(1|1)}$ have the following property, known as the *Chapman-Kolmogorov* equation:

$$P_{(1|1)}(y_3, t_3 | y_1, t_1) = \sum_{y_2} P_{(1|1)}(y_3, t_3 | y_2, t_2) P_{(1|1)}(y_2, t_2 | y_1, t_1) \quad \text{for } t_1 < t_2 < t_3. \quad (8)$$

So, for a Markov process, the hierarchy (1) is completely characterized by specifying $P_{(1)}$ at an initial instant and a system of conditional probabilities $P_{(1|1)}$ satisfying the Chapman-Kolmogorov equation. The study of Markov processes therefore focuses on these two ingredients.⁴

A Markov process is called *homogeneous* if the conditional probabilities $P_{(1|1)}(y_2, t_2 | y_1, t_1)$ do not depend on the two times t_1, t_2 separately but only on their difference $t = t_2 - t_1$; i.e. if they are invariant under time translations. In this case we may write

$$P_{(1|1)}(y_2, t_2 | y_1, t_1) = T_t(y_2, y_1) \quad (9)$$

such conditional probabilities are also called *transition* probabilities.

Is the definition of a Markov process time-symmetric? The choice in (5) of conditionalizing the probability distribution for Y_{t_n} on *earlier* values of Y_t is of course special. In principle, there is nothing in the formulas (1) or (4) that forces such an ordering. One might, just as well, ask for the probability of a value of Y_t in the past, under the condition that part of the *later* behaviour is given (or, indeed, conditionalize on the behaviour at both earlier and later instants.)

At first sight, the Markov property makes no demands about these latter cases. Therefore, one might easily get the impression that the definition is time-asymmetric. However, this is not the case. One can show that (5) is equivalent to:

$$P_{(1|n-1)}(y_1, t_1 | y_2, t_2; \dots; y_n, t_n) = P_{(1|1)}(y_1 t_1 | y_2, t_2) \quad (10)$$

where the convention $t_1 < t_2 < \dots < t_n$ is still in force. Thus, a Markov process does not only have ‘no memory’ but also ‘no foresight’.

Some authors (e.g. Kelly 1979) adopt an (equivalent) definition of a Markov process that is explicitly time-symmetric: Suppose that the value y_i at an instant t_i somewhere in the middle of the sequence $t_1 < \dots < t_n$ is given. The condition for a stochastic process to be Markov is then

$$P_{(n|1)}(y_n, t_n; \dots; y_1, t_1 | y_i, t_i) = P_{(n-i|1)}(y_n, t_n; \dots; y_{i+1}, t_{i+1} | y_i, t_i) P_{(i-1|1)}(y_{i-1}, t_{i-1}; y_1, t_1 | y_i, t_i) \quad (11)$$

⁴Note, however, that although every Markov process is fully characterized by (i) an initial distribution $P_{(1)}(y, 0)$ and (ii) a set of transition probabilities $P_{(1|1)}$ obeying the Chapman-Kolmogorov equation and the equations (7), it is *not* the case that every stochastic process obeying (i) and (ii) is a Markov process. (See Van Kampen 1989: 83] for a counterexample). Still, it is true that one can define a unique Markov process from these two ingredients by stipulating (7).

for all $n = 1, 2, \dots$ and all $1 \leq i \leq n$. In another slogan: The future and past are independent if one conditionalizes on the present.

9.3 Stochastic dynamics

A homogeneous Markov process is for $t > 0$ completely determined by the specification of an initial probability distribution $P_{(1)}(y, 0)$ and the transition probabilities $T_t(y_2|y_1)$ defined by (9). The difference in notation (between P and T) also serves to ease a certain conceptual step. Namely, the idea is to regard T_t as a stochastic evolution operator. Thus, we can regard $T_t(y_2|y_1)$ as the elements of a matrix, representing a (linear) operator T that determines how an initial distribution $P_{(1)}(y, 0)$ will evolve into a distribution at later instants $t > 0$. (In the sequel I will adapt the notation and write $P_{(1)}(y, t)$ as $P_t(y)$.)

$$P_t(y) = (T_t P)(y) := \sum_{y'} T_t(y|y') P_0(y'). \quad (12)$$

The Chapman-Kolmogorov equation (8) may then be written compactly as

$$T_{t+t'} = T_t \circ T_{t'} \quad \text{for } t, t' \geq 0 \quad (13)$$

where \circ stands for matrix multiplication, and we now also extend the notation to include the unit operator:

$$\mathbf{1}(y, y') = T_0(y, y') := \delta_{y, y'} \quad (14)$$

where δ denotes the Kronecker delta.

The formulation (13) can (almost) be interpreted as the group composition property of the evolution operators T . It may be instructive to note how much this is due to the Markov property. Indeed, for arbitrary conditional probabilities, say, if A_i , B_j and C_k denote three families of complete and mutually exclusive events the rule of total probability gives:

$$P(A_i|C_k) = \sum_j P(A_i|B_j, C_k) P(B_j|C_k). \quad (15)$$

In general, this rule can *not* be regarded as ordinary matrix multiplication or a group composition! But the Markov property makes $P(A_i|B_j, C_k)$ in (15) reduce to $P(A_i|B_j)$, and then the summation in (15) coincides with familiar rule for matrix multiplication.

I wrote above: ‘almost’, because there is still a difference in comparison with the normal group property: in the Chapman-Kolmogorov-equation (13) all times must be positive. Thus, in general, for $t > 0$, T_t may not even be defined and so it does *not* hold that

$$T_{-t} \circ T_t = \mathbf{1}. \quad (16)$$

A family of operators $\{T_t, t \geq 0\}$ which is closed under a operation \circ that obeys (13), and for which $T_0 = \mathbf{1}$ is called a *semigroup*. It differs from a group in the sense that its elements T_t need not be *invertible*, i.e., need not have an inverse. The lack of an inverse of T_t may be due to various reasons: either T_t does not possess an inverse, i.e. it is not a one-to-one mapping, or T_t does possess an inverse matrix T_t^{inv} , which however is itself non-stochastic (e.g. it may have negative matrix-elements). We will come back to the role of the inverse matrices in Sections 9.4 and 9.7.

The theory of Markov processes has a strong and natural connection with linear algebra. Sometimes, the theory is presented entirely from this perspective, and one starts with the introduction of a semigroup of *stochastic matrices*, that is to say, m by m matrices T with $T_{ij} \geq 0$ and $\sum_i T_{ij} = 1$. Or, more abstractly, one posits a class of states P , elements of a Banach space with a norm $\|P\|_1 = 1$, and a semigroup of stochastic maps T_t , ($t \geq 0$), subject to the conditions that T_t is linear, positive, and preserves norm: $\|T_t P\|_1 = \|P\|_1$, (cf. Streater 1995).

The evolution of a probability distribution P (now regarded as a vector or a state) is then particularly simple when t is discrete ($t \in \mathbb{N}$):

$$P_t = T^t P_0, \quad \text{where } T^t = \underbrace{T \circ \dots \circ T}_{t \text{ times}}. \quad (17)$$

Homogeneous Markov processes in discrete time are also known as *Markov chains*.

Clearly, if we consider the family $\{T_t\}$ as a semigroup of stochastic evolution operators, or a stochastic form of dynamics, it becomes attractive to look upon $P_0(y)$ as a contingent initial state, chosen independently of the evolution operators T_t . Still, from the perspective of the probabilistic formalism with which we started, this might be an unexpected thought: both $P_{(1)}$ and $P_{(1|1)}$ are aspects of a single, given, probability measure P . The idea of regarding them as independent ingredients that may be specified separately doesn't then seem very natural. But, of course, there is no formal objection against the idea, since every combination of a system of transition probabilities T_t obeying the Chapman-Kolmogorov equation, and an arbitrary initial probability distribution $P_0(y) = P_{(1)}(y, 0)$ defines a unique homogeneous Markov process (cf. footnote 4). In fact, one sometimes even goes one step further and identifies a homogeneous Markov process completely with the specification of the transition probabilities, without regard of the initial state $P_0(y)$; just like the dynamics of a deterministic system is usually presented without assuming any special initial state.

For Markov chains, the goal of specifying the evolution of $P_t(y)$ is now already completely solved in equation (17). In the case of continuous time, it is more usual to specify evolution by means of a differential equation. Such an equation may be obtained in a straightforward manner by considering a Taylor expansion of the transition probability for small times (Van Kampen, 1981: 101–103)—under an appropriate continuity assumption.

The result (with a slightly changed notation) is:

$$\frac{\partial P_t(y)}{\partial t} = \sum_{y'} (W(y|y') P_t(y') - W(y'|y) P_t(y)) \quad (18)$$

Here, the expression $W(y|y')$ is the transition probability from y' to y *per unit of time*. This differential equation,

first obtained by Pauli in 1928, is called the *master equation*.

The interpretation of the equation is suggestive: the change of the probability $P_t(y)$ is determined by making up a balance between gains and losses: the probability of value y increases in a time dt because of the transitions from y' to y , for all possible values of y' . This increase per unit of time is $\sum_{y'} W(y|y')P_t(y')$. But in same period dt there is also a decrease of $P_t(y)$ as a consequence of transitions from the value y to all other possible values y' . This provides the second term.

In this “balancing” aspect, the master equation resembles the Boltzmann equation, despite the totally different derivation, and the fact that $P_t(y)$ has quite another meaning than Boltzmann’s $f_t(v)$. (The former is a probability distribution, the latter a distribution of particles.) Both are first-order differential equations in t . A crucial mathematical distinction from the Boltzmann equation is that the master equation is linear in P , and therefore much easier to solve.

Indeed, any solution of the master equation can formally be written as:

$$P_t = e^{tL} P_0, \tag{19}$$

where L represents the operator

$$L(y|y') := W(y|y') - \sum_{y''} W(y''|y')\delta_{y,y'}. \tag{20}$$

The general solution (19) is similar to the discrete time case (17), thus showing the equivalence of the master equation to the assumption of a homogeneous Markov process in continuous time.

9.4 Approach to equilibrium and increase of entropy?

What can we say in general about the evolution of $P_t(y)$ for a homogeneous Markov process? An immediate result is this: the *relative entropy* is monotonically non-decreasing. That is to say, if we define

$$H(P, Q) := - \sum_{y \in \mathcal{Y}} P(y) \ln \frac{P(y)}{Q(y)} \tag{21}$$

as the relative entropy of a probability distribution P relative to Q , then one can show (see e.g. Moran 1961; Mackey 1992: 30):

$$H(P_t, Q_t) \geq H(P, Q) \tag{22}$$

where $P_t = T_t P$, $Q_t = T_t Q$, and T_t are elements of the semigroup (17) or (19). One can also show that a non-zero relative entropy increase for at least some pair probability distributions P and Q , the stochastic matrix T_t must be non-invertible.

The relative entropy $H(P, Q)$ can, in some sense, be thought of as a measure of how much P and Q “resemble”

each other.⁵ Indeed, it takes its maximum value (i.e. 0) if and only if $P = Q$; it may become $-\infty$ if P and Q have disjoint support, (i.e. when $P(y)Q(y) = 0$ for all $y \in \mathcal{Y}$.) Thus, the result (22) says that if the stochastic process is non-invertible, pairs of distributions P_t and Q_t will generally become more and more alike as time goes by.

Hence it seems we have obtained a general weak aspect of “irreversible behaviour” in this framework. Of course, the above result does not yet imply that the ‘absolute’ entropy $H(P) := -\sum_y P(y) \ln P(y)$ of a probability distribution is non-decreasing. But now assume that the process has a *stationary distribution*. In other words, there is a probability distribution $P^*(y)$ such that

$$T_t P^* = P^*. \quad (23)$$

The intention is, obviously, to regard such a distribution as a candidate for the description of an equilibrium state. If there is such a stationary distribution P^* , we may apply the previous result and write:

$$H(P, P^*) \leq H(T_t P, T_t P^*) = H(P_t, P^*). \quad (24)$$

In other words, as time goes by, the distribution $T_t P$ will then more and more resemble the stationary distribution than does P . If the stationary distribution is also uniform, i.e.:

$$P^*(y) = \frac{1}{m}, \quad (25)$$

then not only the relative but also the absolute entropy $H(P) := -\sum_y P(y) \ln P(y)$ increases, because

$$H(P, P^*) = H(P) - \ln m. \quad (26)$$

In order to get a satisfactory description of an ‘approach to equilibrium’ the following questions remain:

- (i) is there such a stationary distribution?
- (ii) If so, is it unique?
- (iii) does the monotonic behaviour of $H(P_t)$ imply that $\lim_{t \rightarrow \infty} P_t = P^*$?

Harder questions, which we postpone to the next subsection 9.5, are:

- (iv) how to motivate the assumptions needed in this approach or how to make judge their (in)compatibility with an underlying time deterministic dynamics; and
- (v) how this behaviour is compatible with the time symmetry of Markov processes.

Ad (i). A stationary distribution as defined by (23), can be seen as an eigenvector of T_t with eigenvalue 1, or, in the light of (19), an eigenvector of L for the eigenvalue 0. Note that T or L are not necessarily Hermitian (or, rather, since we are dealing with real matrices, symmetric), so that the existence of eigenvectors is not guaranteed by the

⁵Of course, this is an asymmetric sense of “resemblance” because $H(P, Q) \neq H(Q, P)$.

spectral theorem. Further, even if an eigenvector with the corresponding eigenvalue exists, it is not automatically suitable as a probability distribution because its components might not be positive.

Still, it turns out that, due to a theorem of Perron (1907) and Frobenius (1909), every stochastic matrix indeed has a eigenvector, with exclusively non-negative components, and eigenvalue 1 (see e.g. Gantmacher 1959). But if the set \mathcal{Y} is infinite or continuous this is not always true.

However, for continuous variables with a range that has finite measure, the existence of a stationary distribution is guaranteed under the condition that the probability density ρ_y is at all times bounded, i.e. $\exists M \in \mathbb{R}$ such that $\forall t$ $\rho_t \leq M$; (see Mackey 1992: 36)

Ad (ii). The question whether stationary distributions will be unique is somewhat harder to tackle. This problem exhibits an analogy to that of metric transitivity in the ergodic problem.

In general, it is very well possible that the range \mathcal{Y} of Y can be partitioned in two disjoint regions, say A and B , with $\mathcal{Y} = A \cup B$, such that there are no transitions from A to B or vice versa (or that such transitions occur with probability zero). That is to say, the stochastic evolution T_t might have the property

$$T_t(Y \in A|Y \in B) = T_t(Y \in B|Y \in A) = 0 \quad (27)$$

In other words, its matrix may, (perhaps after a conventional relabeling of the outcomes) be written in the form:

$$\begin{pmatrix} T_A & 0 \\ 0 & T_B \end{pmatrix}. \quad (28)$$

The matrix is then called (completely) *reducible*. In this case, stationary distributions will generally not be unique: If P_A^* is a stationary distribution with support in the region A , and P_B^* is a stationary distribution with support in B , then every convex combination

$$\alpha P_A^*(y) + (1 - \alpha) P_B^*(y) \quad \text{with } 0 \leq \alpha \leq 1. \quad (29)$$

will be stationary too. In order to obtain a unique stationary distribution we will thus have to assume an analogue of metric transitivity. That is to say: we should demand that every partition of \mathcal{Y} into disjoint sets A and B for which (27) holds is ‘trivial’ in the sense that $P(A) = 0$ or $P(B) = 0$.

So, one may ask, is the stationary distribution P^* unique if and only if the transition probabilities T_τ are not reducible? In the ergodic problem, the answer is positive (at least if P^* is assumed to be absolutely continuous with respect to the microcanonical measure). But not in the present case!

This has to do with the phenomenon of so-called ‘transient states’, which has no analogy in Hamiltonian

dynamics. Let us look at an example to introduce this concept. Consider a stochastic matrix of the form:

$$\begin{pmatrix} T_A & T' \\ 0 & T_B \end{pmatrix} \quad (30)$$

where T' is a matrix with non-negative entries only. Then:

$$\begin{pmatrix} T_A & T' \\ 0 & T_B \end{pmatrix} \begin{pmatrix} P_A \\ 0 \end{pmatrix} = \begin{pmatrix} T_A P_A \\ 0 \end{pmatrix}, \quad \begin{pmatrix} T_A & T' \\ 0 & T_B \end{pmatrix} \begin{pmatrix} 0 \\ P_B \end{pmatrix} = \begin{pmatrix} T' P_B \\ T_B P_B \end{pmatrix} \quad (31)$$

so that here transitions of the type $a \rightarrow b$ have probability zero, but transitions of the type $b \rightarrow a$ occur with positive probability. (Here, a, b stand for arbitrary elements of the subsets A and B .) It is clear that in such a case the region B will eventually be ‘sucked empty’. That is to say: the total probability of being in region B (i.e. $\|T^t P_B\|$) will go exponentially to zero. The distributions with support in B are called ‘transient’ and the set A is called ‘absorbing’ or a ‘trap’.

It is clear that these transient states will not play any role in the determination of the stationary distribution, and that for this purpose they might be simply ignored. Thus, in this example, the only stationary distributions are those with a support in A . And there will be more than one of them if T_A is reducible.

A matrix T that may be brought (by permutation of the rows and columns) in the form (30), with T_A reducible is called *incompletely reducible* (van Kampen 1981, : 108). Further, a stochastic matrix is called *irreducible* if it is neither completely or incompletely reducible. An alternative (equivalent) criterion is that all states ‘communicate’ with each other, i.e. that for every pair of $i, j \in \mathcal{Y}$ there is some time t such that $P_t(j|i) > 0$.

The Perron-Frobenius theorem guarantees that as long as T irreducible, there is a unique stationary distribution. Furthermore, one can then prove an analogue of the ergodic theorem:(Petersen 1983, : 52)

ERGODIC THEOREM FOR MARKOV PROCESSES: If the transition probability T_t is irreducible, the time average of P_t converges to the unique stationary solution:

$$\lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_0^\tau T_t P(y) dt = P^*(y). \quad (32)$$

Ad (iii). If there is a unique stationary distribution P^* , will $T_t P$ converge to P^* , for every choice of P ? Again, the answer is not necessarily affirmative. (Even if (32) is valid!) For example, there are homogeneous and irreducible Markov chains for which P_t can be divided into two pieces: $P_t = Q_t + R_t$ with the following properties (Mackey 1992, : 71) (Mackey 1992: 71):

1. Q_t is a term with $\|Q_t\| \rightarrow 0$. This is a transient term.
2. The remainder R_t is periodic, i.e. after some time τ the evolution repeats itself: $R_{t+\tau} = R_t$.

These processes are called *asymptotically periodic*. They may very well occur in conjunction with a unique stationary distribution P^* , and show strict monotonic increase of entropy, but still not converge to P^* . In this case,

the monotonic increase of relative entropy $H(P_t, P^*)$ is entirely due to the transient term. For the periodic piece R_t , the transition probabilities are permutation matrices, which, after τ repetitions, return to the unit operator. Further technical assumptions can be introduced to block examples of this kind, and thus enforce a strict convergence towards the unique stationary distribution, e.g. by imposing a condition of ‘exactness’ (Mackey 1992). However, it would take us too far afield to discuss this in detail.

In conclusion, it seems that a weak aspect of ‘irreversible behaviour’, i.e. the monotonic non-decrease of relative entropy is a general feature for all homogeneous Markov processes, (and indeed for all stochastic processes), and non-trivially so when the transition probabilities are non-invertible. Stronger versions of that behaviour, in the sense of affirmative answers to the questions (i), (ii) and (iii), can be obtained too, but at the price of additional technical assumptions.

9.5 Motivations for the Markov property and objections against them

(ad iv). We now turn to the following problem: what is the motivation behind the assumption of the Markov property? The answer, of course, is going to depend on the interpretation of the formalism that one has in mind, and may be different in the ‘coarse-graining’ and the ‘open systems’ or interventionist approaches (cf. section 9.1). I shall discuss the coarse-graining approach in the next section below, and then consider the similar problem for the interventionist point of view .

0.1 Coarse-graining and the repeated randomness assumption

In the present point of view, one assumes that the system considered is really an isolated Hamiltonian system, but the Markov property is supposedly obtained from a partitioning of its phase space. But exactly how is that achieved?

One of the clearest and most outspoken presentations of this view is (van Kampen 1962). Assume the existence of some privileged partition of the Hamiltonian phase space Γ —or of the energy hypersurface Γ_E — into disjoint cells: $\Gamma = \omega_1 \cup \dots \cup \omega_m$. Consider an arbitrary ensemble with probability density ρ on this phase space. Its evolution can be represented by an operator

$$U_t^* \rho(x) := \rho(U_{-t}x), \quad (33)$$

where, we use U_t to denote the Hamiltonian evolution operators. Let transition probabilities between the cells of this partition be defined as

$$T_t(j|i) := P(x_t \in \omega_j | x \in \omega_i) = P(U_t x \in \omega_j | x \in \omega_i) = \frac{\int_{(U_{-t}\omega_j) \cap \omega_i} \rho(x) dx}{\int_{\omega_i} \rho(x) dx}, \quad (34)$$

Obviously such transition probabilities will be homogeneous, due to the time-translation invariance of the Hamiltonian evolution U_t . Further, let $\hat{p}_0(i) := P(x \in \omega_i) = \int_{\omega_i} \rho(x) dx$, $i \in \mathcal{Y} = \{1, \dots, m\}$, be an arbitrary initial coarse-grained probability distribution at time $t=0$.

Using the coarse-graining map defined by:

$$\mathcal{CG} : \rho(x) \mapsto \mathcal{CG}\rho(x) = \sum_i \hat{\rho}(i) \mathbf{1}_{\omega_i}(x), \quad (35)$$

where

$$\hat{\rho}(i) := \frac{\int_{\omega_i} \rho(x) dx}{\int_{\omega_i} dx}, \quad (36)$$

one may also express the coarse-grained distribution at time t as

$$\mathcal{CG}U_t^* \rho(x) = \sum_{ji} T_t(j|i) \hat{p}_0(i) \frac{1}{\mu(\omega_j)} \mathbf{1}_{\omega_j}(x) \quad (37)$$

where μ is the canonical measure on Γ , or the microcanonical measure on Γ_E . This expression indicates that, as long as we are only interested in the coarse grained history, it suffices to know the transition probabilities (34) and the initial coarse grained distributions.

But in order to taste the fruits advertised in the previous sections, one needs to show that the transition probabilities define a Markov process, i.e., that they obey the Chapman-Kolmogorov equation (8),

$$T_{t'+t}(k|i) = T_{t'}(k|j)T_t(j|i); \text{ for all } t, t' > 0. \quad (38)$$

Applying (37) for times t, t' and $t + t'$, it follows easily that the Chapman-Kolmogorov equation is equivalent to

$$\mathcal{CG}U_{t'+t}^* = \mathcal{CG}U_{t'}^* \mathcal{CG}U_t^*, \text{ for all } t, t' > 0. \quad (39)$$

In other words, the coarse-grained probability distribution at time $t + t'$ can be obtained by first applying the Hamiltonian dynamical evolution during a time t , then performing a coarse-graining operation, next applying the dynamical evolution during time t' , and then coarse-graining again. In comparison to the relation $U_{t'+t}^* = U_{t'}^* U_t^*$, we see that the Chapman-Kolmogorov condition can be obtained by demanding that it is allowed to apply a coarse-graining, i.e. to reshuffle the phase points within each cell at any intermediate stage of the evolution.

Of course, this coarse-graining halfway during the evolution erases all information about the past evolution apart from the label of the cell where the state is located at that time; and this ties in nicely with the view of the Markov property as having no memory (cf. the discussion on p. 5).

What is more, the *repeated* application of the coarse-graining does lead to a monotonic non-decrease of the

Gibbs entropy: If, for simplicity, we divide a time interval into m segments of duration τ , we have

$$\rho_{m\tau} = \underbrace{\mathcal{C}\mathcal{G}U_{\tau}^* \mathcal{C}\mathcal{G}U_{\tau}^* \cdots \mathcal{C}\mathcal{G}U_{\tau}^*}_{m \text{ times}} \rho \quad (40)$$

and, since $\sigma(\mathcal{C}\mathcal{G}\rho) \geq \sigma(\rho)$:

$$\sigma[\rho_{m\tau}] \geq \sigma[\rho_{(m-1)\tau}] \geq \dots \geq \sigma[\rho_{\tau}] \geq \sigma[\rho_0]. \quad (41)$$

But since the choice of τ is arbitrary, we may conclude that $\sigma[\rho_t]$ is monotonically non-decreasing.

Thus, van Kampen argues, the ingredient to be added to the dynamical evolution is that, at any stage of the evolution, one should apply a coarse-graining of the distribution. It is important to note that it is not sufficient to do that just once at a single instant. At every stage of the evolution we need to coarse-grain the distribution again and again. Van Kampen (1962: 193) calls this the *repeated randomness* assumption.

What is the justification for this assumption? Van Kampen points out that it is “not unreasonable” (ibid., p. 182), because of the brute fact of its success in phenomenological physics. Thermodynamics and other phenomenological descriptions of macroscopic systems (the diffusion equation, transport equations, hydrodynamics, the Fokker-Planck equation, etc.) all characterize macroscopic systems with a very small number of variables. This means that their state descriptions are very coarse in comparison with the microscopic phase space. But their evolution equations are autonomous and deterministic: the change of the macroscopic variables is given in terms of the instantaneous values of those very same variables. The success of these equations shows, apparently, that the precise microscopic state does not add any relevant information beyond this coarse description. At the same time, van Kampen admits that the coarse-graining procedure is clearly not always successful. It is not difficult to construct a partition of a phase space into cells for which the Markov property fails completely.

Apparently, the choice of the cells must be “just right” (van Kampen 1962, : 183). But there is as yet no clear prescription how this is to be done. Van Kampen (1981: 80) argues that it is “the art of the physicist” to find the right choice, an art in which he or she succeeds in practice by a mixture of general principles and ingenuity, but where no general guidelines can be provided. The justification of the repeated randomness assumption is that it leads to the Markov property and from there onwards to the master equation, providing a successful autonomous, deterministic description of the evolution of the coarse-grained distribution.

It is worth noting that van Kampen thus qualifies the ‘usual’ point of view on the choice of the cells; namely, that the cells are chosen in correspondence to our finite observation capabilities. Observability of the macroscopic variables is not sufficient for the success of the repeated randomness assumption. It is conceivable (and occurs in practice) that a particular partition in terms of observable quantities does not lead to a Markov process. In that case, the choice of observable variables is simply inadequate and has to be extended with other (unobservable) quantities until we (hopefully) obtain an exhaustive set, i.e. a set of variables for which the evolution can be described autonomously. An example is the spin-echo experiment: the (observable) total magnetization of the system does not provide a suitable coarse-grained description. For further discussion of this theme, see: (Blatt 1959, Ridderbos

& Redhead 1998, Lavis 2004, Balian 2005).

Apart from the unsolved problem for which partition the repeated randomness assumption is to be applied, other objections have been raised against the repeated randomness assumption. Van Kampen actually gives us not much more than the advice to accept the repeated randomness assumption bravely, not to be distracted by its dubious status, and firmly keep our eyes on its success. For authors as Sklar (1993) this puts the problem on its head. They request a justification of the assumption that would *explain* the success of the approach. (Indeed, even van Kampen (1981, : 80) describes this success as a “miraculous fact”!). Such a request, of course, will not be satisfied by a justification that relies on its success. (But that does not mean, in my opinion, that it is an invalid form of justification.)

Another point that seems repugnant to many authors, is that the repeated coarse-graining operations appear to be added ‘by hand’, in deviation from the true dynamical evolution provided by U_t . The increase of entropy and the approach to equilibrium would thus apparently be a consequence of the fact that *we* shake up the probability distribution repeatedly in order to wash away all information about the past, while refusing a dynamical explanation for this procedure. Redhead (1995, : 31) describes this procedure as “one of the most deceitful artifices I have ever come across in theoretical physics” (see also Blatt (1959) Sklar (1993) and Callender (1999) for similar objections).

0.2 Interventionism or ‘open systems’

Another approach to stochastic dynamics is by reference to open systems. The idea is here that the system in continual interaction with the environment, and that this is responsible for the approach to equilibrium.

Indeed, it cannot be denied that in concrete systems isolation is an unrealistic idealization. The actual effect of interaction with the environment on the microscopic evolution can be enormous. A proverbial example, going back to Borel (1914), estimates the gravitational effect caused by displacing one gram of matter on Sirius by one centimeter on the microscopic evolution of an earthly cylinder of gas. Under normal conditions, the effect is so large, that, roughly and for a typical molecule in the gas, it may be decisive for whether or not this molecule will hit another given molecule after about 50 intermediary collisions. That is to say: microscopic dynamical evolutions corresponding to the displaced and the undisplaced matter on Sirius start to diverge considerably after a time of about 10^{-6} sec. In other words, the mechanical evolution of such a system is so extremely sensitive for disturbances of the initial state that even the most minute changes in the state of the environment can be responsible for large changes in the microscopic trajectory. But we cannot control the state of environment. Is it possible to regard irreversible behaviour as the result of such uncontrollable disturbances from outside?⁶

Let (x, y) be the state of a total system, where, as before, $x \in \Gamma^{(s)}$ represents the state of the object system and

⁶Note that the term ‘open system’ is employed here for a system in (weak) interaction with its environment. This should be distinguished from the notion of ‘open system’ in other branches of physics where it denotes a system that can exchange particles with its environment.

$y \in \Gamma^{(e)}$ that of the environment. We assume that the total system is governed by a Hamiltonian of the form

$$H_{\text{tot}}(x, y) = H_{(s)} + H_{(e)} + \lambda H_{\text{int}}(x, y), \quad (42)$$

so that the probability density of the ensemble of total systems evolves as

$$\rho_t(x, y) = U_t^* \rho_0(x, y) = \rho(U_{-t}(x, y)) \quad (43)$$

i.e., a time-symmetric, deterministic and measure-preserving evolution.

At each time, we may define marginal distributions for both system and environment:

$$\rho_t^{(s)}(x) = \int dy \rho_t(x, y), \quad (44)$$

$$\rho_t^{(e)}(y) = \int dx \rho_t(x, y). \quad (45)$$

We are, of course, mostly interested in the object system, i.e. in (44). Assume further that at time $t = 0$ the total density factorizes:

$$\rho_0(x, y) = \rho_0^{(s)}(x) \rho_0^{(e)}(y). \quad (46)$$

What can we say about the evolution of $\rho_t^{(s)}(x)$? Does it form a Markov process, and does it show increase of entropy?

An immediate result (see e.g. Penrose & Percival 1962) is this:

$$\sigma[\rho_t^{(s)}] + \sigma[\rho_t^{(e)}] \geq \sigma[\rho_0^{(s)}] + \sigma[\rho_0^{(e)}], \quad (47)$$

where σ denotes the Gibbs fine-grained entropy

$$\sigma[\rho] = - \int \rho(x) \ln \rho(x) dx. \quad (48)$$

This result follows from the fact that $\sigma[\rho_t]$ is conserved and that the entropy of a joint probability distribution is always smaller than or equal to the sum of the entropies of their marginals; with equality if the joint distribution factorizes. This gives a form of entropy change for the total system, but it is not sufficient to conclude that the object system itself will evolve towards equilibrium, or even that its entropy will be monotonically increasing. (Notice that (47) holds for $t \leq 0$ too.)

Actually, this is obviously not to be expected. There are interactions with an environment that may lead the system away from equilibrium. We shall have to make additional assumptions about the situation. A more or less usual set of assumptions is:

- (a). The environment is very large (or even infinite); i.e.: the dimension of $\Gamma^{(e)}$ is much larger than that of $\Gamma^{(s)}$,

and $H_{(s)} \ll H_{(e)}$.

(b). The coupling between the system and the environment is weak, i.e. λ is very small.

(c). The environment is initially in thermal equilibrium, e.g., $\rho^{(e)}(y)$ is canonical:

$$\rho_0^{(e)} = \frac{1}{Z(\beta)} e^{-\beta H^{(e)}} \quad (49)$$

(d). One considers time scales only that are long with respect to the relaxation times of the environment, but short with respect to the Poincaré recurrence time of the total system.

Even then, it is a major task to obtain a master equation for the evolution of the marginal state (44) of the system, or to show that its evolution is generated by a semigroup, which would guarantee that this forms a Markov process (under the proviso of footnote 4). Many specific models have been studied in detail (cf. Spohn 1980). General theorems were obtained (although mostly in a quantum mechanical setting) by (Davies 1974, Davies 1976a, Lindblad 1976, Gorini et al. 1976). But there is a similarity to the earlier approach: it seems that, here too, an analogue of ‘repeated randomness’ must be introduced. (Mehra & Sudarshan 1972, van Kampen 1994, Maes & Netočný 2003).

At the risk of oversimplifying the results obtained in this analysis, I believe they can be summarized as showing that, in the so-called ‘weak coupling’ limit, or some similar limiting procedure, the time development of (44) can be modeled as

$$\rho_t^{(s)}(x) = T_t \rho^{(s)}(x) \quad t \geq 0, \quad (50)$$

where the operators T_t form a semigroup, while the environment remains in its steady equilibrium state:

$$\rho_t^{(e)}(y) = \rho_0^{(e)}(y) \quad t \geq 0. \quad (51)$$

The establishment of these results would also allow one to infer, from (47), the monotonic non-decrease of entropy of the system.

To assess these findings, it is convenient to define, for a fixed choice of $\rho_0^{(e)}$ the following linear map on probability distributions of the total system:

$$\mathcal{TR} : \rho(x, y) \mapsto \mathcal{TR}\rho(x, y) = \int \rho(x, y) dy \cdot \rho_0^{(e)}(y) \quad (52)$$

This map removes the correlation between the system and the environment, and projects the marginal distribution of the environment back to its original equilibrium form.

Now, it is not difficult to see that the Chapman-Kolmogorov equation (which is equivalent to the semigroup property) can be expressed as

$$\mathcal{TR}U_{t+t'}^* = \mathcal{TR}U_{t'}^* \mathcal{TR}U_t^* \quad \text{for all } t, t' \geq 0 \quad (53)$$

which is analogous to (39).

There is thus a strong formal analogy between the coarse-graining and the open-systems approaches. Indeed, the variables of the environment play a role comparable to the internal coordinates of a cell in the coarse graining approach. The exact microscopic information about the past is here translated into the form of correlations with the environment. This information is now removed by assuming that at later times, effectively, the state may be replaced by a product of the form (46), neglecting the back-action of the system on the environment. The mappings \mathcal{CG} and \mathcal{TR} are both linear and idempotent mappings, that can be regarded as special cases of the projection operator techniques of Nakajima and Zwanzig, which allows for a more systematical and abstract elaboration, sometimes called *subdynamics*.

Some proponents of the open systems approach, (e.g. Morrison 1966, Redhead 1995), argue that in contrast to the coarse-graining approach, the present procedure is ‘objective’. Presumably, this means that there is supposed to be a fact of the matter about whether the correlations are indeed ‘exported to the environment’. However, the analogy between both approaches makes one suspect that any problem for the coarse-graining approach is translated into an analogous problem of the open systems approach. Indeed, the problem of finding a privileged partition that we discussed in the previous section is mirrored here by the question where one should place the division between the ‘system’ and ‘environment’. There is no doubt that its practical applications this choice is also arbitrary.

9.6 Can the Markov property explain irreversible behaviour?

Ad (v). Finally, I turn to what may well be the most controversial and surprising issue: is the Markov property, or the repeated randomness assumption offered to motivate it, responsible for the derivation of time-reversal non-invariant results?

We have seen that every non-invertible homogeneous Markov process displays “irreversible behaviour” in the sense that different initial probability distributions will tend to become more alike in the course of time. Under certain technical conditions, one can obtain stronger results, e.g. an approach to a unique equilibrium state, monotonic non-decrease of absolute entropy, etc. All these results seem to be clearly time-asymmetric. And yet we have also seen that the Markov property is explicitly time symmetric. How can these be reconciled?

To start off, it may be noted that it has often been affirmed that the Markov property is the key towards obtaining time-asymmetric results. For example, Penrose writes:

“... the behaviour of systems that are far from equilibrium is not symmetric under time reversal: for example: heat always flows from a hotter to a colder body, never from a colder to a hotter. If this behaviour could be derived from the symmetric laws of dynamics alone there would, indeed, be a paradox; we must therefore acknowledge the fact that some additional postulate, non-dynamical in character and asymmetric under time reversal must be adjoined to the symmetric laws of dynamics before the theory can become rich enough to explain non-equilibrium behaviour. In the present theory, this additional postulate is the Markov postulate” (Penrose 1970: 41).

Many similar statements, e.g. that the repeated randomness assumption is “the additional element by which sta-

tistical mechanics has to be supplemented in order to obtain irreversible equations” (van Kampen 1962, : 182), or that the non-invertibility of a Markov process provides the origin of thermodynamic behaviour (Mackey 1992) can be found in the works of advocates of this approach.

But how can this be, given that the Markov property is explicitly time-symmetric? In order to probe this problem, consider another question. How does a given probability distribution $P(y, 0)$ evolve for negative times? So, starting again from (6), let us now take $t \leq 0$. We still have:

$$P(y, t) = \sum_{y'} P(y, t, |y', 0)P(y', 0). \quad (54)$$

These conditional probabilities $P(y, t, |y', 0)$ satisfy the ‘time-reversed’ Markov property (10), that says that extra specification of later values is irrelevant for the retrodiction of earlier values. As a consequence, we get for $t \leq t' \leq t'', 0$:

$$P(y, t|y'', t'') = \sum_{y'} P(y, t|y', t')P(y', t'|y'', t'') \quad (55)$$

i.e., a time-reversed analogue of the Chapman-Kolmogorov equation.

We may thus also consider these conditional probabilities for negative times as backward evolution operators. If we could assume their invariance under time translation, i.e. that they depend only on the difference $\tau = t - t'$, we could write

$$S_\tau(y|y') := P(y, t|y', t') \quad \text{with } \tau = t - t' \leq 0, \quad (56)$$

and obtain a second semigroup of operators S_τ , obeying

$$S_{\tau+\tau'} = S_\tau \circ S_{\tau'} \quad \tau, \tau' \leq 0 \quad (57)$$

that generate stochastic evolutions towards the past.

These backward conditional probabilities are connected to the forward conditional probabilities by means of Bayes’ theorem:

$$P_{(1|1)}(y, t|y', t') = \frac{P_{(1|1)}(y', t'|y, t)P(y, t)}{P(y', t')}; \quad (58)$$

and if the process, as before, is homogeneous this becomes

$$P_{(1|1)}(y, t|y', t') = \frac{T_{-\tau}(y'|y)P_t(y)}{P_{t'}(y')} \quad ; \quad \tau = t - t' < 0. \quad (59)$$

The matrix $P_{(1|1)}(y, t|y', t')$ always gives for $t < t'$ the correct ‘inversion’ of T_t . That is to say:

$$\sum_{y'} P(y, t|y', t')(T_{t'-t}P_t)(y') = P_t(y) \quad (60)$$

Note firstly that (59) is *not* the matrix-inverse of T_t ! Indeed, the right-hand side of (59) depends on P_t and $P_{t'}$ as well as T . Even if the matrix-inverse $T^{(\text{inv})}$ does not exist, or is not a bona fide stochastic matrix, the evolution towards the past is governed by the Bayesian inversion, i.e. by the transition probabilities (59).

Note also that if the forward transition probabilities are homogeneous, this is not necessarily so for the backward transition probabilities. For example, if in (59) one translates both t and t' by δ , one finds

$$P(y, t + \delta | y', t' + \delta) = \frac{T_{-\tau}(y' | y) P(y, t + \delta)}{P(y', t' + \delta)}.$$

Here, the right-hand side generally still depends on δ . In the special case that the initial distribution is itself stationary, the backward transition probabilities are homogeneous whenever the forward ones are. If $P(y, t)$ is not stationary, we might still reach the same conclusion, as long as the non-stationarity is restricted to those elements y or y' of \mathcal{Y} for which $T_t(y | y') = 0$ for all t . Otherwise, the two notions become logically independent.

This gives rise to an unexpected new problem. Usually, an assumption of homogeneity (or time translation invariance) is seen as philosophically innocuous, as compared to time reversal invariance. But here we see that assuming time translation invariance for a system of *forward* transition probabilities is not equivalent to assuming the same invariance for the *backward* transition probabilities. If one believes that one of the two is obvious, how will one go about explaining the failure of the other? And how would one explain the preference for which one of the two is obvious, without falling into the “double standards” accusation of the kind raised by (Price 1996)?

But what about entropy increase? We have seen before that for every non-invertible Markov process the relative entropy of the distribution P with respect to the equilibrium distribution increases, and that the distribution evolves towards equilibrium. (Homogeneity of the process is not needed for this conclusion.) But the backward evolution operators form a Markov process too, for which exactly the same holds. This seems paradoxical. If $T_t P_0 = P_t$, we also have $P_t = S_{-t} P_0$. The entropy of P_t can hardly be both higher and lower than that of P_0 ! An example may clarify the resolution of this apparent problem: namely, the stationary solutions of S are not the same as the stationary solutions of T !

Example Consider a Markov chain with $\mathcal{Y} = \{1, 2\}$ and let

$$T = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}. \quad (61)$$

Choose an initial distribution $P_0 = (\alpha, 1 - \alpha)$. After one step we already get: $P_1 = T P_0 = (\frac{1}{2}, \frac{1}{2})$ which is also the (unique) stationary distribution P^* . The backward transition probabilities are given by Bayes' theorem, and one finds easily:

$$S = \begin{pmatrix} \alpha & \alpha \\ 1 - \alpha & 1 - \alpha \end{pmatrix}. \quad (62)$$

The stationary distribution for this transition probability is $\tilde{P}^* = (\alpha, 1 - \alpha)$. That is to say: for the forward evolution operator the transition

$$\begin{pmatrix} \alpha \\ 1 - \alpha \end{pmatrix} \xrightarrow{T} \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \quad (63)$$

is one for which a non-stationary initial distribution evolves towards a stationary one. The relative entropy increases: $H(P_0, P^*) \leq H(P_1, P^*)$. But for the backward evolution, similarly:

$$\begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \xrightarrow{S} \begin{pmatrix} \alpha \\ 1 - \alpha \end{pmatrix} \quad (64)$$

represents an evolution from a non-stationary initial distribution to the stationary distribution \tilde{P}^* and, here too, relative entropy increases: $H(P_1, \tilde{P}^*) \leq H(P_0, \tilde{P}^*)$.

The illusion that non-invertible Markov processes possess a built-in time-asymmetry is (at least partly) due to the habit of regarding T_τ as a fixed evolution operator on an independently chosen distribution P_0 . Such a view is of course very familiar in other problems in physics, where deterministic evolution operators generally *do* form a group and may be used, at our heart's desire, for positive and negative times.

Indeed, the fact that these operators in general have no inverse might seem to reflect the idea that Markov processes have no memory and 'lose information' along the way and that is the cause of the irreversible behaviour, embodied in the time-asymmetric master equation, increase of relative or absolute entropy or approach to equilibrium. But actually, every Markov process has apart from a system of forward, also a system of backward transition probabilities, that again forms a semigroup (when they are homogeneous). If we had considered *them* as given we would get all conclusions we obtained before, but now for negative times.

I conclude that irreversible behaviour is not built into the Markov property, or in the non-invertibility of the transition probabilities, (or in the repeated randomness assumption⁷, or in the Master equation or in the semigroup property). Rather the appearance of irreversible behaviour is due to the choice to rely on the forward transition probabilities, and not the backward. A similar conclusion has been reached before (Edens 2001) in the context of proposals of Prigogine and his coworkers. My main point here is that the same verdict also holds for more 'mainstream' approaches as coarse-graining or open systems.

⁷In recent work, van Kampen acknowledges that the repeated randomness assumption by itself does not lead to irreversibility: "This repeated randomness assumption [...] breaks the time symmetry by explicitly postulating the randomization *at the beginning* of the time interval Δt . There is no logical justification for this assumption other than that it is the only thing one can do and that it works. If one assumes randomness at the end of each Δt coefficients for diffusion, viscosity, etc. appear with the wrong sign; if one assumes randomness at the midpoint no irreversibility appears" (van Kampen 2002, : 475, original emphasis).

9.7 Reversibility of stochastic processes

In order not to end this paper on a destructive note, let me emphasize that I do not claim that the derivation of irreversible behaviour in stochastic dynamics is impossible. Instead, the claim is that motivations for desirable properties of the forward transition probabilities are not enough; one ought also show that these properties are lacking for the backward transitions.

In order to set up the problem of irreversibility in this approach to non-equilibrium statistical mechanics for a more systematic discussion, one first ought to provide a reasonable definition for what it means for a stochastic process to be (ir)reversible; a definition that would capture the intuitions behind its original background in Hamiltonian statistical mechanics.

One general definition that seems to be common (cf. Kelly 1979: 5) is to call a stochastic process reversible iff, for all n and t_1, \dots, t_n and τ :

$$P_{(n)}(y_1, t_1; \dots; y_n, t_n) = P_{(n)}(y_1, \tau - t_n; \dots; y_n, \tau - t_1). \quad (65)$$

See Grimmett & Stirzaker 1982: 219) for a similar definition restricted to Markov processes) The immediate consequence of this definition is that a stochastic process can only be reversible if the single-time probability $P_{(1)}(y, t)$ is stationary, i.e. in statistical equilibrium. Indeed, this definition seems to make the whole problem of reconciling irreversible behaviour with reversibility disappear. As Kelly (1979: 19) notes in a discussion of the Ehrenfest model: “there is no conflict between reversibility and the phenomenon of increasing entropy — reversibility is a property of the model in equilibrium and increasing entropy is a property of the approach to equilibrium”

But clearly, this view trivializes the problem, and therefore it is not the appropriate definition for non-equilibrium statistical mechanics. Recall that the Ehrenfest dog flea model (§9.2) was originally proposed in an attempt of showing how a tendency of approaching equilibrium from a initial non-equilibrium distribution (e.g. a probability distribution that gives probability 1 to the state that all fleas are located on the same dog) could be reconciled with a stochastic yet time-symmetric dynamics.

If one wants to separate considerations about initial conditions from dynamical considerations at all, one would like to provide a notion of (ir)reversibility that is associated with the stochastic dynamics alone, independent of the initial distribution is stationary.

It seems that an alternative definition which would fulfill this intuition is to say that a stochastic process is reversible if, for all y and y' and $t' > t$,

$$P_{(1|1)}(y, t|y', t') = P_{(1|1)}(y, t'|y', t). \quad (66)$$

In this case we cannot conclude that the process must be stationary, and indeed, the Ehrenfest model would be an example of a reversible stochastic process. I believe this definition captures the intuition that if at some time state

y' obtains, the conditional probability of the state one time-step earlier being y is equal to that of the state one time-step later being y .

According to this proposal, the aim of finding the “origin” of irreversible behaviour or “time’s arrow”, etc. in stochastic dynamics must then lie in finding and motivating conditions under which the forward transition probabilities are different from the backwards transition probabilities, in the sense of a violation of (66). Otherwise, irreversible behaviour would essentially be a consequence of the assumptions about initial conditions, a result that would not be different in principle from conclusions obtainable from Hamiltonian dynamics.

9.8 References

References

- Balian, R. (2005). Information in statistical physics. *Studies In History and Philosophy of Modern Physics*, 36, 323–353.
- Blatt, J.M. (1959). An alternative approach to the ergodic problem. *Progress in Theoretical Physics*, 22, 745–756.
- Borel, E. (1914). *Le Hasard*. Paris: Alcan.
- Brush, S.G. (2003), *The kinetic theory of gases*. London: Imperial College Press.
- Callender, C. (1999). Reducing thermodynamics to statistical mechanics: the case of entropy. *Journal of Philosophy*, 96, 348–373.
- Callender, C. (2001). Taking Thermodynamics Too Seriously. *Studies In History and Philosophy of Modern Physics*, 32, 539–553.
- Davies, E.B. (1974). Markovian master equations. *Communications in Mathematical Physics*, 39 91–110.
- Davies, E.B. (1976a). Markovian master equations II *Mathematische Annalen*, 219 147–158.
- Davies, E.B (1976b). *Quantum theory of open systems*. New York: Academic Press
- Edens, B. (2001). Semigroups and symmetry: an investigation of Prigogine’s theories. <http://philsci-archive.pitt.edu/archive/00000436/>.
- Emch, G.G. (1965). On the Markov character of master-equations. *Helvetica Physica Acta*, 38, 164–171.
- Emch, G.G. & Liu, C. (2001). *The Logic of thermostistical physics*. Berlin: Springer.
- Gantmacher, F.R. (1959). *Matrizenrechnung* Vol 2. Berlin: Deutscher Verlag der Wissenschaften.
- Gorini, V., Kossakowski, A. & Sudarshan, E.C.G. Completely positive dynamical semigroups of N -level systems. *Journal of Mathematical Physics* 17 8721–825.

- Grimmett, G. R. & Stirzaker, D.R. (1982). *Probability and random processes*. Oxford: Clarendon Press.
- Grünbaum, A. (1973). Is the coarse-grained entropy of classical statistical mechanics an anthropomorphism? In *Philosophical problems of space and time*, 2nd ed., Dordrecht: Reidel, pp. 646–665.
- van Kampen, N.G. (1962). Fundamental problems in the statistical mechanics of irreversible processes. In E.G.D. Cohen (Ed.), *Fundamental problems in statistical mechanics* Amsterdam: North-Holland, pp. 173–202.
- van Kampen, N.G. (1981). *Stochastic processes in chemistry and physics*. Amsterdam: North-Holland.
- van Kampen, N.G. (1994) Models for dissipation in quantum mechanics. In J.J.Brey, J. Marro, J.M. Rubi, M. San Miguel (Eds.) *25 years of non-equilibrium statistical mechanics* Berlin: Springer.
- van Kampen, N.G. (2002) The road from molecules to Onsager. *Journal of Statistical Physics*, 109, 471–481.
- van Kampen, N.G. (2004). A new approach to noise in quantum mechanics *Journal of Statistical Physics*, 115, 1057–1072.
- Kelly F.P. (1979). *Reversibility and stochastic networks* Chichester: Wiley. Also at http://www.statslab.cam.ac.uk/~afrb2/kelly_book.html.
- Lavis, D.A. (2004). The spin-echo system reconsidered. *Foundations of Physics*, 34, 669–688.
- Lindblad, G. (1976). On the generators of quantum dynamical semigroups. *Communications in Mathematical Physics*, 48, 119–130.
- Lindblad, G. (1983). *Non-equilibrium entropy and irreversibility* Dordrecht: Reidel.
- Maes, C. & Netočný, K. (2003). Time-reversal and entropy. *Journal of Statistical Physics*, 110, 269–310.
- Mackey, M.C. (1992). *Time's arrow: the origins of thermodynamic behavior*. New York: Springer.
- Mackey, M.C. (2001). Microscopic dynamics and the second law of thermodynamics. In: *Time's Arrows, quantum Measurements and Superluminal Behavior*. C. Mugnai, A. Ranfagni & L.S. Schulman (Eds.) (Roma: Consiglio Nazionale delle Ricerche).
- Mehra, J. & Sudarshan, E.C.G. (1972). Some reflections on the nature of entropy, irreversibility and the second law of thermodynamics, *Nuovo Cimento B*, 11, 215–256.
- Moran, P.A.P. (1961). Entropy, Markov processes and Boltzmann's H-theorem. *Proceedings of the Cambridge Philosophical Society*, 57, 833–842.
- Morrison, P. (1966). Time's arrow and external perturbations. In A. de Shalit, H. Feshbach, & L. van Hove (Eds.), *Preludes in Theoretical Physics in honor of V. F. Weisskopf*. Amsterdam: North Holland, pp. 347–.
- Penrose, O. (1970). *Foundations of statistical mechanics: a deductive treatment*. Oxford: Pergamon Press.

- Penrose, O. (1979). Foundations of statistical mechanics. *Reports on Progress in Physics*, 42, 1937–2006.
- Penrose, O., & Percival I. (1962). The direction of time. *Proceedings of the Physical Society*, 79, 605–616.
- Petersen, K. (1983). *Ergodic theory*. Cambridge: Cambridge University Press.
- Price, H. (1996). *Time's arrow and Archimedes' point*. New York: Oxford University Press.
- Redhead, M. (1995). *From physics to metaphysics*. Cambridge: Cambridge University Press.
- Ridderbos, T.M. (2002). The coarse-graining approach to statistical mechanics: how blissful is our ignorance? *Studies in History and Philosophy of Modern Physics*, 33, 65–77.
- Ridderbos, T.M & Redhead, M.L.G. (1998). The spin-echo experiment and the second law of thermodynamics. *Foundations of Physics*, 28, 1237–1270.
- Schrödinger, E., 1950. Irreversibility. *Proceedings of the Royal Irish Academy* 53, 189–195.
- Shenker, O. (2000). Interventionism in statistical mechanics: some philosophical remarks. <http://philsci-archive.pitt.edu/archive/00000151/>.
- Sklar, L. (1993). *Physics and Chance. Philosophical Issues in the Foundations of Statistical Mechanics*. Cambridge: Cambridge University Press.
- Spohn, H. (1980). Kinetic equations from Hamiltonian dynamics: Markovian limits. *Reviews of Modern Physics* 52, 569–615.
- Streater, R.F. (1995). *Statistical dynamics; a stochastic approach to non-equilibrium thermodynamics*. London: Imperial College Press.
- Sudarshan, E.C.G., Mathews, P.M., & Rau, J. (1961). Stochastic dynamics of quantum-mechanical systems, *Physical Review*, 121, 920–924.
- Uffink, J. (2007) Compendium of the foundations of classical statistical physics. In J. Butterfield and J. Earman (eds.) *Handbook of the Philosophy of Physics*. Amsterdam: North-Holland, pp. 923–1074.