

# The Ergodic Hierarchy, Randomness and Hamiltonian Chaos<sup>1, 2</sup>

Joseph Berkovitz

Department of Philosophy, University of Maryland Baltimore County

Email: jberkov@umbc.edu

Roman Frigg

Department of Philosophy, Logic and Scientific Method,

London School of Economics

Email: r.p.frigg@lse.ac.uk

Fred Kronz

Department of Philosophy, University of Texas at Austin

Email: kronz@mail.utexas.edu

## Abstract

Various processes are often classified as both deterministic and random or chaotic. The main difficulty in analysing the randomness of such processes is the apparent tension between the notions of randomness and determinism: what type of randomness could exist in a deterministic process? Ergodic theory seems to offer a particularly promising theoretical tool for tackling this problem by positing a hierarchy, the so-called ‘ergodic hierarchy’ (EH), which is commonly assumed to provide a hierarchy of increasing degrees of randomness. However, that notion of

---

<sup>1</sup> Forthcoming in *Studies in History and Philosophy of Modern Physics* 37(4). The final version of the paper differs from this draft. Please download the PDFfile of the final version from <http://www.romanfrigg.org/writings.htm>.

<sup>2</sup> This work is fully collaborative; the authors are listed alphabetically.

randomness requires qualification. The mathematical definition of EH does not make explicit appeal to randomness; nor does the usual way of presenting EH involve a specification of the notion of randomness that is supposed to underlie the hierarchy. In this paper we argue that EH is best understood as a hierarchy of random behaviour if randomness is explicated in terms of unpredictability. We then show that, contrary to common wisdom, EH is useful in characterising the behaviour of Hamiltonian dynamical systems.

Keywords: Randomness, ergodic hierarchy, chaos, predictability, dynamical system, correlation.

## 1. Introduction

Various deterministic processes are frequently classified as random. The main difficulty in analysing random behaviour in such processes is the apparent tension between the notions of randomness and determinism. Intuitively, in deterministic processes the past determines the future, and the question is then: what type of randomness could exist in such processes? One influential suggestion has it that the so-called ‘ergodic hierarchy’ can elucidate the nature of randomness in dynamical processes that are (by assumption) deterministic.

The ergodic hierarchy (EH henceforth) is a hierarchical classification of dynamical systems, and it is typically presented as consisting of five levels:

Ergodic  $\supset$  Weak Mixing  $\supset$  Strong Mixing  $\supset$  Kolmogorov  $\supset$  Bernoulli.

The diagram above indicates that all B-systems (Bernoulli systems) are K-systems (Kolmogorov systems), all K-systems are SM-systems (strong mixing systems), all SM-systems are WM-systems (weak mixing systems), and all WM-systems are E-systems (ergodic systems). A system

that is an E-system but not a WM-system will be referred to below as ‘merely ergodic’, and similarly for the next three levels.<sup>3</sup>

This hierarchy is often presented as a hierarchy of random behaviour: the higher up in this hierarchy a system is placed the more random its behaviour.<sup>4</sup> But in what sense is it random? With the exception of B-systems, the mathematical definitions of the different levels of the hierarchy do not make explicit appeal to randomness or any notion that can readily be connected to randomness. Nor does the usual way of presenting EH involve a specification of the notion of randomness that is supposed to underlie the hierarchy. So what is the notion of randomness that underlies EH and in what sense exactly is EH a hierarchy of random behaviour? The aim in this paper is to answer these questions and to show how EH can be used to characterise the behaviour of Hamiltonian dynamical systems.

The primary contention of our analysis is that EH can naturally be understood as a hierarchy of randomness if different degrees of randomness are explicated in terms of different degrees of unpredictability, where unpredictability is accounted for in terms of epistemic probabilistic

---

<sup>3</sup> Sometimes EH is presented as having yet another level, namely C-systems (also referred to as Anosov systems or completely hyperbolic systems). Although interesting in their own right, C-systems are beyond the scope of this paper. They do not have a unique place in EH and their relation to other levels of EH depends on details, which we cannot discuss here. Paradigm examples of C-systems are located between K- and B-systems; that is, they are K-systems but not necessarily B-systems. The cat map, for instance, is a C-system that is also a K-system (Lichtenberg & Leibermann, 1992, p. 307); but there are K-systems such as the stadium billiard which are not C-systems (Ott, 1993, p. 262). Some C-systems preserve a smooth measure (where ‘smooth’ in this context means absolutely continuous with respect to the Lebesgue measure), in which case they are Bernoulli systems. But not all C-systems have smooth measures. Nevertheless, it is always possible to find other measures such as SRB (Sinai, Ruelle, Bowen) measures. However, matters are more complicated in such cases, as such C-systems need not be mixing and a fortiori they need not be K- or B-systems (Ornstein & Weiss, 1991, pp. 75-82).

<sup>4</sup> This claim is common in the literature on chaos theory where EH is presented as a hierarchy of random behaviour (see for instance Lichtenberg & Leibermann, 1992; Ott, 1993; Tabor, 1989). Recently Belot and Earman (1997) have used EH to define chaos in dynamical systems.

relevance. Different degrees of probabilistic relevance, in turn, are spelled out in terms of different types of decay of correlation between a system's states at different times.<sup>5</sup>

The structure of the paper is as follows. After a brief introduction to dynamical systems and EH in Section 2, we show in Section 3 that EH can (and should) be interpreted as a hierarchy of degrees of unpredictability. To this end, we first motivate the idea that dynamical randomness can be explicated in terms of unpredictability and that different degrees of unpredictability can be cashed out in terms of different types of decay of correlation. Then we show that each level of EH exhibits a different type of decay of correlation. In Section 4 we address the question of the applicability of EH to Hamiltonian systems. It is frequently argued that EH is useless because almost all Hamiltonian systems of interest are non-ergodic. We reply that, contrary to common wisdom, these arguments do not undermine the usefulness of the ergodic hierarchy. Finally, in Section 5 we draw some conclusions from our analysis of EH for the definition of chaos. In particular, we suggest that chaos comes in degrees rather than being an all or nothing matter.

## **2. Classical dynamical systems and the ergodic hierarchy**

In this section we provide a brief introduction to dynamical systems and the ergodic hierarchy. Intuitive and accessible introductions can be found in Lichtenberg & Liebermann (1992) and Tabor (1989); for detailed discussions the reader may consult Arnold and Avez (1968), Mañé (1983), Nadkarni (1998), Parry (1981), Petersen (1983), Sinai (1980), and Walters (1982), as well as many other monographs on the subject.

---

<sup>5</sup> Within the literature on EH and chaos, the term 'correlation' is often used to denote the integral of the product of two functions, where the argument of one of the functions is shifted by a constant. As we explain in section 3, we use 'correlation' in a different sense, which is close to how the term is used in probability theory.

## 2.1 Classical dynamical systems

The fundamental object in ergodic theory is a dynamical system  $[X, \Sigma, \mu, \tau]$ . It consists of a dynamical law  $\tau$  on a probability space  $[X, \Sigma, \mu]$ .  $X$  is a set of elements, which is sometimes called the phase space.  $\Sigma$  is a  $\sigma$ -algebra of measurable subsets of  $X$ , meaning that:<sup>6</sup>

- (1)  $X \in \Sigma$ ,
- (2)  $A \setminus B \in \Sigma$  for all  $A, B \in \Sigma$ , and
- (3)  $\bigcup_{i=1}^n B_i \in \Sigma$  if  $B_i \in \Sigma$  for  $1 \leq i \leq n \leq \infty$ .

This implies that  $\Sigma$  also contains  $\emptyset$  as well as all (finite or countably infinite) intersections  $\bigcap_{i=1}^n B_i$ , where  $\{B_i\}_{i=1}^n \subset \Sigma$  for  $1 \leq n \leq \infty$ . Furthermore there is a probability measure  $\mu$  on  $\Sigma$ :

- (1)  $\mu: \Sigma \rightarrow [0, 1]$  with  $\mu(X) = 1$ , and
- (2) If  $\{B_i\}_{i=1}^n \subset \Sigma$  and  $B_j \cap B_k = \emptyset$  for  $1 \leq j < k \leq n \leq \infty$ , then  $\mu(\bigcup_{i=1}^n B_i) = \sum_{i=1}^n \mu(B_i)$ .

An automorphism  $T$  is a transformation that maps the probability space  $[X, \Sigma, \mu]$  onto itself; it is measure-preserving iff for all  $B \in \Sigma$ :

- (1)  $T^{-1}B \in \Sigma$  and,
- (2)  $\mu(T^{-1}B) = \mu(B)$ , where  $T^{-1}B = \{x \in X : Tx \in B\}$ .

A dynamical law  $\tau := \{T_t\}_{t \in I}$  is a group of measure-preserving automorphisms  $T_t: X \rightarrow X$  of the probability space  $[X, \Sigma, \mu]$ , where  $I$  is either  $R$  (the real numbers) or  $Z$  ( $\dots -1, 0, 1, \dots$ ).<sup>7</sup> In the discrete case ( $I = Z$ ),  $\tau$  is often generated by the iterative application of the same automorphism  $T$  and accordingly we have  $T_t = T^t$ .

---

<sup>6</sup> In what follows, ' $\in$ ' denotes the set-theoretical membership, ' $A \setminus B$ ' denotes the set-theoretical subtraction of  $A$  from  $B$ , ' $\bigcup$ ' is the set-theoretical union. In the text that follows, ' $\bigcap$ ' is the set-theoretical intersection, ' $\emptyset$ ' is the empty set, and 'iff' is a shorthand for 'if and only if'.

<sup>7</sup> Ergodic theory is not limited to the study maps based on  $R$  and  $Z$ . In particular, a considerable body of mathematics has been collected on amenable groups (e.g. spatial translations). Furthermore, substantial parts of EH have been extended to the quantum realm. However, these developments, interesting and important in their own right, fall outside the scope of this paper, which only deals with classical dynamical systems.

The set  $\alpha = \{\alpha_i : i = 1, \dots, n\}$  is a partition of  $X$  iff

$$(1) \alpha_i \cap \alpha_j = \emptyset \text{ for all } i \neq j$$

$$(2) \mu(X \setminus \bigcup_{i=1}^n \alpha_i) = 0.$$

The  $\alpha_i$  are called ‘atoms of the partition’. It is important to observe that the image of a partition under any element  $T_i$  of  $\tau$  is still a partition. Given two partitions  $\alpha = \{\alpha_i : i = 1, \dots, n\}$  and  $\beta = \{\beta_j : j = 1, \dots, m\}$ , their least common refinement (sometimes also referred to as their sum)  $\alpha \vee \beta$  is defined as follows:  $\alpha \vee \beta = \{\alpha_i \cap \beta_j : i = 1, \dots, n ; j = 1, \dots, m\}$ .

In what follows we will, by and large, confine our attention to discrete dynamical systems involving a measure-preserving automorphism in which the dynamics is induced by an iterative application of this automorphism. Some of these restrictions are more significant than others. Nothing of what we have to say about randomness in this paper hinges in any way on the choice of discrete time and a dynamic that is based on the iterative application of the same map. We focus on these systems just for the sake of simplicity; all our results readily generalise to the case of continuous flows. Other systems we omit from our analysis are ones whose dynamics is based on endomorphisms (non-invertible maps) or non-area preserving mappings, or ones that have a phase space whose measure is not normalisable. Whether or not our results can be carried over to these cases is an interesting question, which for want of space we cannot discuss here.

## 2.2 The ergodic hierarchy

The lowest level of the ergodic hierarchy is ergodicity. Let  $f$  be any complex-valued Lebesgue-integrable function on a probability space. Its space mean  $\bar{f}$  is defined as  $\bar{f} := \int_X f(x) d\mu$ , and its time mean  $f^*(x)$  is defined as  $f^*(x) := \lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=0}^{k-1} f[T^i(x)]$ . The Birkhoff-Khinchin theorem assures that  $f^*(x)$  exists almost everywhere (i.e. except, possibly, on a set of measure zero). A dynamical system  $[X, \Sigma, \mu, \tau]$  is ergodic iff for every complex-valued, Lebesgue-integrable function  $f$  the time mean equals the space mean, i.e.  $f^*(x) = \bar{f}$ , almost everywhere. Furthermore, one can prove that a system is ergodic iff the following holds for all sets  $A, B \in \Sigma$ :

$$(E) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu(T^k B \cap A) = \mu(B)\mu(A)$$

A system is strongly mixing iff for all  $A, B \in \Sigma$ :

$$(SM) \quad \lim_{n \rightarrow \infty} \mu(T^n B \cap A) = \mu(A)\mu(B).$$

One can relax this requirement a bit by allowing for fluctuations while requiring that the system is ‘strongly mixing on average’. This gives weak mixing. A system is weakly mixing iff for all  $A, B \in \Sigma$ :

$$(WM) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} |\mu(T^k B \cap A) - \mu(B)\mu(A)| = 0.$$

Strong mixing implies weak mixing, and weak mixing implies ergodicity. The converse relations do not hold. Hence, strong mixing is a stronger condition than weak mixing, and weak mixing is stronger condition than ergodicity.

K-systems mark the next higher level in the ergodic hierarchy after strong mixing. A dynamical system  $[X, \Sigma, \mu, \tau]$  is a K-system iff there is a subalgebra  $\Sigma_0 \subset \Sigma$  such that the following three conditions hold: (1)  $\Sigma_0 \subseteq T\Sigma_0$ , (2)  $\bigvee_{n=-\infty}^{\infty} T^n \Sigma_0 = \Sigma$ , (3)  $\bigwedge_{n=-\infty}^{\infty} T^n \Sigma_0 = N$ . In this definition,  $T^n \Sigma_0$  is the  $\sigma$ -algebra containing the sets  $T^n B$  ( $B \in \Sigma_0$ ),  $N$  is the  $\sigma$ -algebra consisting uniquely of sets of measure one and measure zero,  $\bigvee_{n=-\infty}^{\infty} T^n \Sigma_0$  denotes the sum or the ‘refinement’ of all the  $T^n \Sigma_0$  (as introduced in section 2), and  $\bigwedge_{n=-\infty}^{\infty} T^n \Sigma_0$  denotes the largest subalgebra of  $\Sigma$  which belongs to each  $T^n \Sigma_0$ . There is an important theorem due to Cornfeld *et al.* (1982, p. 283) which states that a dynamical system  $[X, \Sigma, \mu, \tau]$  is a K-system iff it is K-mixing (KM). A system is KM iff for any set  $A_0$ , for any positive integer  $r$  and for any set of measurable subsets  $A_1, \dots, A_r \in \Sigma$

$$(KM) \quad \lim_{n \rightarrow \infty} \sup_{B \in \sigma_{n,r}} |\mu(B \cap A_0) - \mu(B)\mu(A_0)| = 0,$$

where  $\sigma_{n,r}$  is the minimal  $\sigma$ -algebra generated by  $\{T^k A_i : k \geq n ; i = 1, \dots, r\}$ . The algebra  $\sigma_{n,r}$  is obtained as follows: Add  $X$  to the above set, add all differences  $T^k A_i \setminus T^l A_j$ , where  $k, l \geq n$  and  $i, j = 1, \dots, r$ , and finally also add all (finite and infinite) unions of these. The set we get is a  $\sigma$ -

algebra by construction; it is minimal because apart from the  $T^k A_i$  no other set has been used to construct it.

Bernoulli systems mark the highest level of EH. The intuitive idea is that a Bernoulli system is one whose behaviour is as random as a coin toss or a roulette wheel. We first introduce so-called Bernoulli schemes, which then give rise to the definition of a Bernoulli system. Let  $Y$  be a finite set of elements  $Y = \{f_1, \dots, f_n\}$  (sometimes also called the ‘alphabet’ of the system) and let

$\nu(f_i) = p_i$  be a probability measure on  $Y$ :  $0 \leq p_i \leq 1$  for all  $1 \leq i \leq n$ , and  $\sum_{i=1}^n p_i = 1$ . Furthermore,

let  $X$  be the direct product of infinitely many copies of  $Y$ :  $X = \prod_{i=-\infty}^{+\infty} Y_i$ , where  $Y_i = Y$  for all  $i$ . The

elements of  $X$  are doubly-infinite sequences  $x = \{x_i\}_{i=-\infty}^{+\infty}$ , where  $x_i \in Y$  for each  $i \in Z$ . As the  $\sigma$ -algebra  $C$  of  $X$  we choose the  $\sigma$ -algebra generated by all sets of the form  $\{x \in X \mid x_i = k, m \leq i \leq m+n\}$  for all  $m \in Z$ , for all  $n \in N$ , and for all  $k \in Y$  (the so-called

‘cylinder sets’). As a measure on  $X$  we take the product measure  $\prod_{i=-\infty}^{+\infty} \nu_i$ , that is

$\mu\{x_i\}_{i=-\infty}^{+\infty} = \dots \nu(x_{-2})\nu(x_{-1})\nu(x_0)\nu(x_1)\nu(x_2)\dots$ . The triple  $[X, C, \mu]$  is commonly referred to as a

‘stochastic process’. This process is stationary if the chance element is constant in time, that is iff for all cylinder sets  $\mu(y: y_{i+1} = w_i, m \leq i < m+n) = \mu(y: y_i = w_i, m \leq i < m+n)$  holds. An invertible

measure-preserving transformation  $T: X \rightarrow X$ , the so-called shift map, is naturally associated with every stationary stochastic process:  $Tx = \{y_i\}_{i=-\infty}^{+\infty}$  where  $y_i = x_{i+1}$  for all  $i \in Z$ . It is

straightforward to see that the measure  $\mu$  is invariant under  $T$  (i.e. that  $T$  is measure preserving)

and that  $T$  is invertible. This construction is commonly referred to as a ‘Bernoulli Scheme’ and denoted by ‘ $B(p_1, \dots, p_n)$ ’. From this it follows that the quadruple  $[X, C, \mu, T]$  is a dynamical

system.

Naturally the question arises of how all this relates to automorphisms of continuous spaces we have been discussing in connection with all the other levels of the hierarchy. In order to relate these two frameworks to one another we chose a partition  $\alpha = \{\alpha_1, \dots, \alpha_n\}$  of the phase space  $X$  of the continuous system and say that the system is Bernoulli if the coarse grained dynamics



behaves like a random process. More specifically, an automorphism  $T: X \rightarrow X$  of a probability space  $[X, \Sigma, \mu]$  is a Bernoulli automorphism if there exists a partition  $\alpha = \{\alpha_1, \dots, \alpha_n\}$  such that all  $T^i \alpha$ ,  $-\infty < i < \infty$ , are independent of each other,<sup>8</sup> where two partitions  $\alpha$  and  $\beta$  are independent of each other iff  $\mu(\alpha_i \cap \beta_j) = \mu(\alpha_i)\mu(\beta_j)$  for all atoms  $\alpha_i \in \alpha$  and  $\beta_j \in \beta$ . Finally, a dynamical system  $[X, \Sigma, \mu, T]$  is a Bernoulli system if  $T$  is a Bernoulli automorphism.

We would like to mention that there are other types of Bernoulli conditions: Weak Bernoulli and Very Weak Bernoulli processes. These conditions are mathematically complicated and the discussion of their nature and their intimate relationships to B-systems is beyond the scope of this paper (the interested reader may consult Ornstein (1974) or Shields (1973)). Yet, what is important to mention here is that these conditions are of special interest in practice. It is often very difficult to demonstrate that a dynamical system  $[X, \Sigma, \mu, T]$  is a B-system by way of showing that the automorphism  $T$  is a Bernoulli automorphism. It is easier to show that a dynamical system is Very Weak Bernoulli, and satisfaction of this condition is sufficient for the system to be Bernoulli. Ornstein demonstrated this sufficiency condition, and showed that many dynamical systems satisfy are Very Weak Bernoulli.

### 3. Decay of correlations, unpredictability and randomness

#### 3.1. Introducing the central notions

The term ‘randomness’ may refer to many different things. Among them are the absence of order or pattern, the occurrence of events or accident or chance, the lack of structure, being brought about without method, purpose, plan or principle, being computationally complex, the absence of dynamical stability, having positive Kolmogorov-Sinai entropy (KS entropy) the probabilistic

---

<sup>8</sup> A second condition needs to be satisfied:  $\alpha$  must be  $T$ -generating:  $\bigcup_n \bigvee_{i=-n}^n T^i \alpha$  generates  $\Sigma$ . However, what matters for our considerations is the independence condition.

independence of future from past events, and unpredictability.<sup>9</sup> Some of these have been given detailed treatment within well worked-out theories while others have attracted less attention.

Which of these notions of randomness can be used to understand EH as a hierarchy of random behaviour? The sought after notion of randomness has to allow for a uniform characterisation of the ergodic hierarchy. That is, it must be possible to understand each level of the hierarchy as manifesting some degree of the *same* notion of randomness and the higher the level is in the hierarchy the greater is its degree of this kind of randomness.

Several of the above-mentioned conceptions of randomness do not seem to square with EH at all and can be dismissed as inapplicable (e.g. chance process or pattern breaking). Some seemingly obvious candidates such as the KS entropy, algorithmic complexity and exponential divergence of nearby trajectories are ruled out by the above criterion of adequacy. The problem with these notions is that while they can be used to distinguish between the kind of randomness that we find within the higher and the lower levels of EH, they remain silent about how the random properties of different levels at the high or the low end differ from one another. It is a theorem that a system is a K-system iff it has positive KS entropy (Cornfeld *et al.*, 1982, p. 283). From this it follows that merely E, merely WM and merely SM systems have zero KS entropy while K and B systems have positive KS entropy. For this reason the KS entropy fails to distinguish between merely E, merely WM and merely SM systems. In fact, it also fails to distinguish between merely K and B systems, as there are B systems whose entropy is smaller than the one of certain merely K systems. So the KS entropy is blind towards both the differences within the lower and the higher regions of the hierarchy. Other notions ‘inherit’ this problem from the KS entropy. Algorithmic complexity is linked to the presence of positive KS entropy by Brudno’s theorem, which roughly says that the KS-entropy of a system is equal to the algorithmic complexity of almost all its trajectories (Brudno, 1978; for details see Alekseev & Yakobson, 1981). And similarly for the divergence of nearby trajectories, which can only occur in K or B systems. *Pesin’s theorem* states that a dynamical system is dynamically unstable in the sense of having positive Lyapounov

---

<sup>9</sup> There is subtlety here: what we really mean is probability equal to one, which is not the same as certainty. However, the difference between the two does not matter for our purposes so we keep using the two locutions interchangeably.

exponents (i.e. that it has exponential divergence of nearby trajectories) iff it has a positive KS-entropy (Lichtenberg & Liebermann 1992, p. 304 and references therein).

It is the main contention of this paper that unpredictability may be used to characterise the nature of increasing degrees of randomness involved in EH. But before arguing the case for unpredictability in detail, we first have to briefly review the basic notions of unpredictability and randomness we have in mind. At the most basic level, an event is said to be unpredictable if the probability of its occurrence is independent of past events; and a process is unpredictable if the probability of its present and future stages is independent of its past history. This basic notion of unpredictability can be refined by relaxing the requirement of total probabilistic independence and only demanding that past states are of limited relevance for present and future states (we give a precise definition of relevance below). Furthermore, one can discuss the relevance of different parts of the past for our predictions: do our predictions get better when we know more about the past of the system? Does such knowledge make predictions more reliable than they would be in light of total ignorance about what happened earlier on in the process? Intuitively, the less relevance the information of the past history has for the quality of our predictions, the more random the process is. For instance, tossing a coin has a high degree of randomness as we are not able to predict whether the coin lands on heads or tails, and things do not improve in the light of better knowledge of the outcomes of past tosses.

We reach an interpretation of EH as a hierarchy of unpredictability in three steps. First, we introduce the notion of correlation between two subsets of the phase space and then show that different levels of EH exhibit different patterns of decay of correlations. At this stage, our analysis is purely mathematical and does not rely on any philosophical views about randomness, predictability, or probability. Second, we introduce the notions of event and predictability. Third, we introduce an interpretative principle that connects the notions of decay of correlation and predictability, and on its basis demonstrate how the different patterns of decay of correlation we find in the different levels of EH correspond to different types of unpredictability.

## 3.2 EH and decay of correlation

The correlation between two sets  $A, B \in \Sigma$  is defined as

$$(3) \ C(B, A) := \mu(A \cap B) - \mu(A)\mu(B).$$

If the numerical value of  $C(B, A)$  is positive (negative) there is positive (negative) correlation between  $A$  and  $B$ ; if it is zero then  $A$  and  $B$  are uncorrelated. Using this notion of correlation we now turn to analyse the pattern of decay of correlations involved in the different levels of EH.

### 3.2.1. Ergodicity

From (E), (C) and the fact that  $T$  is area-preserving – i.e. that  $\mu(B) = \mu(T^i B)$  – it follows that a system is ergodic iff:

$$(C-E) \quad \lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=1}^{k-1} C(T^i B, A) = 0 \text{ for all } A, B \in \Sigma.$$

Notice that  $\frac{1}{k} \sum_{i=1}^{k-1} C(T^i B, A)$  is the average of the correlations between the  $T^i B$  and  $A$ . So (C-E) says that, as  $k$  tends towards infinity, the average of all  $C(T^i B, A)$  approaches zero. The qualification ‘the average’ is essential since ergodicity is compatible with there being high correlations during the entire process. In fact, there need not be any decay of correlations; that is, there need not be a single  $i$  for which  $C(T^i B, A)$  equals zero. (C-E) is compatible, for instance, with positive and negative correlations trading off against each other in such a way that the average of all  $C(T^i B, A)$  is washed out when  $k$  tends towards infinity.

### 3.2.2 Mixing

From (SM) and the fact that  $T$  is area-preserving, it follows that a system is SM iff the correlations between  $A$  and  $T^k B$  tend to zero as  $k$  approaches infinity:

$$(C\text{-SM}) \quad \lim_{k \rightarrow \infty} C(T^k B, A) = 0 \text{ for all } A, B \in \Sigma.$$

Unlike in a merely ergodic system, where none of the  $C(T^k B, A)$  has to tend towards zero as  $k$  tends towards infinity, in SM all correlations tend towards zero as  $k$  tends towards infinity.

Weak mixing also involves some decay of correlations, albeit of a weaker sort. Again, using the fact that  $T$  is area preserving the condition for weak mixing becomes:

$$(C\text{-WM}) \quad \lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=1}^{k-1} |C(T^i B, A)| = 0 \text{ for all } A, B \in \Sigma.$$

This implies that in the course of its evolution, the correlations between  $A$  and  $T^i B$  asymptotically approach zero.

The important difference between weak mixing and ergodicity is that since we sum over the modulo of the correlations it is no longer possible that positive and negative correlations between states cancel each other out without themselves ever approaching zero. Thus, WM is stronger than E. On the other hand, WM is weaker SM. In a WM system, some correlations may remain (even as  $k$  approaches infinity). Weak mixing is compatible with the presence of correlations from time to time as long as they are infrequent enough to have no significance influence on the weighted average of the correlations between  $A$  and  $T^i B$ .

### 3.2.3 *K*-systems

It is a theorem that a dynamical system is a *K*-system iff it is KM (Section 2.2). This equivalence allows us to focus on the study of the KM property rather than the abstract definition of a *K*-system, which is not very telling for the analysis of the nature of dynamical randomness in *K*-systems.

In order to discuss the KM conditions along the lines of our discussion of E, WM and SM, it is convenient to use the standard  $\varepsilon$ -definition of the limit. For any sequence  $\{a_n | n=0,1,2,\dots\}$ ,  $\lim_{n \rightarrow \infty} a_n = 0$  iff for any  $\varepsilon > 0$  there exists a  $n_0 > 0$ , such that, for all  $n > n_0$ ,  $|a_n| < \varepsilon$ . Then the KM condition becomes:<sup>10</sup>

(C-KM) A system is KM iff for any integer  $r$  and any sets  $A_0, A_1, \dots, A_r \in \Sigma$ , and for any  $\varepsilon > 0$  there exists a  $n_0 > 0$ , such that, for all  $n > n_0$  and all  $B \in \sigma_{n,r}(A_1, \dots, A_r)$ ,  $|C(B, A_0)| < \varepsilon$ .

To clarify the kind of decay of correlation involved in K-systems, we first have to spell out the nature of the sigma algebra  $\sigma_{n,r}(A_1, \dots, A_r)$ . From what has been said about sigma algebras in Section 2 it follows that  $\sigma_{n,r}(A_1, \dots, A_r)$  contains, among others, the following sets: all sets having the form  $T^k A_i$  (for all  $k \geq n$  and  $i = 1, \dots, r$ ); all finite and infinite sequences having the forms  $T^n A_{m_1} \cup T^n A_{m_2} \cup T^n A_{m_3} \cup \dots$  and  $T^n A_{m_1} \cup T^{n+1} A_{m_2} \cup T^{n+2} A_{m_3} \cup \dots$  (where the  $m_i$  are indices ranging over  $1, \dots, r$ ); and all finite and infinite intersections having the forms  $T^n A_{m_1} \cap T^n A_{m_2} \cap T^n A_{m_3} \cap \dots$  and  $T^n A_{m_1} \cap T^{n+1} A_{m_2} \cap T^{n+2} A_{m_3} \cap \dots$ .

The elements  $A_1, \dots, A_r$  in the expression  $\sigma_{n,r}(A_1, \dots, A_r)$  are measurable sets in  $X$ . For the purposes of our current discussion it is convenient to choose  $A_1, \dots, A_r$  such that they form a partition of  $X$ . This can be done without any loss of generality, because the sigma algebra  $\sigma_{n,r}(A_1, \dots, A_r)$  always contains a partition based on  $A_1, \dots, A_r$ . This can be seen as follows. First, assume that  $A_1, \dots, A_r$  are disjoint. By definition  $\sigma_{n,r}(A_1, \dots, A_r)$  contains  $A_{r+1} := X - \bigcup_{i=1}^r A_i$  and trivially  $\{A_i\}_{i=1}^{r+1}$  is a partition. Second, if  $A_1, \dots, A_r$  overlap, then the sigma algebra contains (by definition) all intersections and all set-theoretic differences between  $A_1, \dots, A_r$ , which together with  $X - \bigcup_{i=1}^r A_i$  generate a partition. For example, let  $r=2$  and suppose that  $A_1$  and  $A_2$  are measurable overlapping subsets of  $X$  (See Fig. 1a). A partition of  $X$  may then be generated as follows. Let

---

<sup>10</sup> Thanks to David Lavis for suggesting this formulation of (C-KM) to us and for his help with clearing up a confusion in an earlier version.

$P_1 = A_1 \setminus (A_1 \cap A_2)$ ,  $P_2 = (A_1 \cap A_2)$ ,  $P_3 = A_2 \setminus (A_1 \cap A_2)$ ,  $P_4 = X \setminus (A_1 \cup A_2)$ ; then  $P = \{P_1, P_2, P_3, P_4\}$  forms a partition of  $X$  (see Fig. 1b).

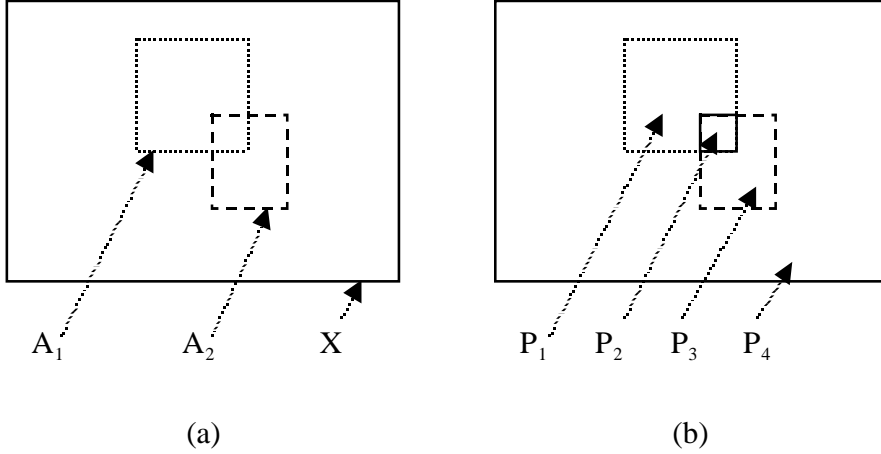


Fig. 1. Generating a partition from overlapping sets.

Based on the above analysis of the sigma algebra  $\sigma_{n,r}(A_1, \dots, A_r)$ , we can now turn to discuss the nature of the decay of correlation in K-systems. For reasons that will become clear as we proceed, we only focus on two types of elements of the sigma algebra and disregard the others.

*Type I: 'Basic' sets.*  $B$  ranges over all sets  $T^k A_i$ ,  $i = 1, \dots, r$  and  $k \geq n$  (which are, by definition, elements of  $\sigma_{n,r}(A_1, \dots, A_r)$  for every  $n$ ). Substituting these sets into (C-KM) we obtain:

(C-KM-1) Given any  $\varepsilon > 0$  there exists a  $n_0 > 0$ , such that, for all  $n > n_0$   $|C(T^k A_i, A_0)| < \varepsilon$  for all  $k \geq n$  and  $i = 1, \dots, r$

This implies that given any  $\varepsilon > 0$  there exists a  $n_0 > 0$ , such that, for all  $n > n_0$  and all  $|C(T^n A_i, A_0)| < \varepsilon$ ,  $i = 1, \dots, r$ . This, by definition, is equivalent to  $\lim_{n \rightarrow \infty} |C(T^n A_i, A_0)| = 0$ ,  $i = 1, \dots, r$ . It is a theorem that for any sequence  $\{a_n | n = 0, 1, 2, \dots\}$ :  $\lim_{n \rightarrow \infty} |a_n| = 0$  iff  $\lim_{n \rightarrow \infty} a_n = 0$ . Hence the above is equivalent to  $\lim_{n \rightarrow \infty} C(T^n A_i, A_0) = 0$ ,  $i = 1, \dots, r$ , which is just the strong mixing condition (C-SM)

since there are no restrictions on the choice of  $A_i$  and  $A_0$ . So KM implies SM; and accordingly everything that has been said about SM above is also valid in the present case.

*Type II: Intersections of different sets.* What marks the difference between SM- and K-systems is that in the case of K-system the set  $B$  in  $C(B, A_0)$  also ranges over all finite and infinite intersections of  $T^k A_i$ ,  $i = 1, \dots, r$  (as opposed to only basic sets). Substituting one of these intersections into (C-KM) we obtain:

$$(C-KM-2) \quad \text{Given any } \varepsilon > 0 \text{ there exists a } n_0 > 0, \text{ such that, for all } n > n_0 \\ \left| C\left(T^k A_{m_1} \cap T^{k+1} A_{m_2} \cap T^{k+2} A_{m_3} \cap \dots, A_0\right) \right| < \varepsilon \text{ for all } k \geq n \text{ and } m_i \text{ ranging over} \\ \{1, \dots, r\}.$$

This condition is equivalent to  $\lim_{n \rightarrow \infty} C\left(T^n A_{m_1} \cap T^{n+1} A_{m_2} \cap T^{n+2} A_{m_3} \cap \dots, A_0\right) = 0$ , where the  $m_i$  range over  $\{1, \dots, r\}$ . Since we are free to choose the  $A_i$  as we like, this is true for any refinement of such infinitely remote (coarse-grained) past history (provided that it is still a coarse-grained history).

#### 3.2.4. Bernoulli systems

Recall that  $[X, \Sigma, \mu, T]$  is a Bernoulli system if there is a partition  $\alpha$  of  $X$  such that  $T^i \alpha$ ,  $-\infty < i < \infty$ , are independent; that is if  $\mu(T^n \alpha_i \cap T^m \alpha_j) = \mu(T^n \alpha_i) \mu(T^m \alpha_j)$  for any  $-\infty < m, n < \infty$  and where  $i, j$  range over the number of atoms in the partition  $\alpha$ . Let  $m = 0$ . Then we get  $\mu(T^n \alpha_i \cap \alpha_j) = \mu(T^n \alpha_i) \mu(\alpha_j)$ . Denoting (for the sake of consistency with the above notation) the relevant sets by 'A' and 'B' we have:

$$(C-B) \quad C(T^n B, A) = 0 \text{ for all integers } n \text{ and for all } A, B \in \Sigma.$$



Since, trivially, (C-B) holds for  $n = 1$ , a Bernoulli process reaches strong mixing after just one step. Furthermore, since *all*  $T^i \alpha$  are independent, we also have:<sup>11</sup>

$$(C-B^*) \quad C(T^n A_{m_1} \cap T^{n+1} A_{m_2} \cap T^{n+2} A_{m_3} \dots, A_0) = 0,$$

for all  $m_i$ , which range over  $\{1, \dots, r\}$ .

(Again, an intuitive account of what it means for a system to satisfy this condition will be given below in section 3.5.)

### 3.3 States, events and unpredictability

The aim of this subsection is to present the notion of unpredictability that underlies our analysis of the kind of randomness involved in EH. To this end, we first introduce the notions of event and probability we are working with.

It is a common to associate with every subset  $A$  of the sigma algebra of the phase space  $X$  a property  $P_A$ . This property obtains if and only if the state  $x$  of the system lies within the set  $A$ . For instance, given the phase space  $X = \{(q, p): q, p \in \mathbb{R}\}$  of a classical particle moving along a straight line, one may associate with the set  $A = \{(q, p): q \in \mathbb{R}, p \in \mathbb{R}^+\}$  the property of ‘having positive momentum.’ This property obtains iff the system’s state lies within  $A$ . In case  $A$  is a point in the phase space, we call  $P_A$  an ‘atomic property.’ It is important to mention that in general  $A$  may be arbitrary and the property corresponding to a set  $A$  will be rather contrived. But this is not a problem for our analysis since nothing in what follows hinges on  $P_A$  being a ‘natural’ or ‘intuitive’ property.

Let us define an ‘event’  $A^t$  as the obtaining of the property  $P_A$  at some time  $t$  (or, equivalently, as the system’s state being in subset  $A$  at time  $t$ ). For instance, assuming that  $A$  is defined as above,  $A^t$  is the event of the particle having positive momentum at time  $t$ . This definition of event,

---

<sup>11</sup> This is intuitively clear. We refer those who desire a rigorous proof of this to Mañé (1983, p. 87), who formulates the independence condition such that (C-B\*) follows immediately from it.

which is somewhat similar to Kim's (1973), is not uncontroversial. However, the typical problems with this notion of event, e.g. problems associated with event causation, are irrelevant to our analysis of the nature of randomness underlying EH.

Now, although at each time the system is assumed to be exactly in one state  $x \in X$  and thus possesses all properties  $P_A$  for which  $x \in A$ , we may not know this state. For this reason it is convenient to introduce a measure that reflects our uncertainty or ignorance about the system's states and properties. A natural choice for such a measure is probability. Let  $p(A^t)$  denote the epistemic probability of the event  $A^t$ ; i.e., a probability that reflects a degree of belief that the state of the system lies within  $A$  at time  $t$ . Hence,  $p(A^t)$  reflects a degree of belief in  $P_A$ 's obtaining at time  $t$ .

We can similarly introduce conditional probabilities. Let  $A^t$  and  $B^{t_1}$  denote the events of the system's state being in  $A$  at  $t$  and in  $B$  at  $t_1$ , where  $t > t_1$ . The expression  $p(A^t | B^{t_1})$  denotes the probability that event  $A^t$  takes place given that event  $B^{t_1}$  takes place. By the common definition of conditional probability,  $p(A^t | B^{t_1})$  can be expressed in terms of unconditional ones:  $p(A^t | B^{t_1}) = p(A^t \& B^{t_1}) / p(B^{t_1})$ . This conditional, epistemic probability reflects the degree of belief in the future event  $A^t$  on the basis of knowledge of, or in the past event  $B^{t_1}$ .

This conditional probability can easily be generalised to the case of more than two events. Let  $B_1, \dots, B_r$  (where  $r$  is any natural number) be measurable but otherwise arbitrary subsets of  $X$  and consider  $r+1$  instants of time  $t_1, \dots, t_r, t$  such that  $t > t_1 > \dots > t_r$ . Then,  $p(A^t | B_1^{t_1} \& \dots \& B_r^{t_r})$  is the degree of belief in the system's state being in  $A$  at time  $t$  given that it was in  $B_1$  at  $t_1$  and in  $B_2$  at  $t_2$  and ... in  $B_r$  at  $t_r$ .

We can now make the notion of unpredictability we have in mind more precise. The basic idea is that unpredictability is to be measured in terms of how relevant the information about states of the system at a certain time is to its states at other times. In particular, it is a measure of the extent to which the information that the state of the system is in the subset  $B$  at  $t_1$  is relevant to the probability that the state of the system will be in the subset  $A$  at  $t$  (where  $A, B \in \Sigma$ ). It is common

to measure uncertainty about the event  $A^t$  given the information about past event  $B^{t_1}$  in terms of the relevance that  $B^{t_1}$  has for the degree of belief in  $A^t$ . A natural measure of an agent's inability to predict  $A^t$  on the basis of the information about  $B^{t_1}$  is the *probabilistic relevance* of  $B^{t_1}$  for  $A^t$ , which is abbreviated as  $R_p(B^{t_1}, A^t)$ , where

$$R_p(B^{t_1}, A^t) := p(A^t | B^{t_1}) - p(A^t), \quad \text{provided that } p(B^{t_1}) \neq 0. \quad ^{12}$$

No probabilistic relevance, i.e.  $R_p(A^{t_1}, B^{t_2}) = 0$ , means that the information that the system is in  $B$  at  $t_1$  has no bearing on the agent's prediction of the system being in  $A$  at  $t$ . In other words, the information that the system is in  $B$  at  $t_1$  has no bearing on the agent's degree of belief that the system is in  $A$  at  $t$ . More generally, the weaker the probabilistic relevance of past events for  $A^t$  is, the more unpredictable is  $A^t$  for the agent.

For purely technical reasons, which will become apparent soon, it turns out to be more convenient to work with a slightly different notion of relevance in what follows, which is obtained from the above by multiplying both sides of the equation with  $p(B^{t_1})$ . We call this new quantity simply *relevance* and denote it by the symbol  $R(B^{t_1}, A^t)$ :

$$(R) \quad R(B^{t_1}, A^t) := R_p(B^{t_1}, A^t)p(B^{t_1}) = p(A^t \& B^{t_1}) - p(A^t)p(B^{t_1})$$

As, by assumption  $p(B^{t_1})$  is a real number greater than zero, we (trivially) have:  $R(B^{t_1}, A^t) * 0$  iff  $R_p(B^{t_1}, A^t) * 0$  and  $R(B^{t_1}, A^t) * R(B^{t_1}, C^t)$  iff  $R_p(B^{t_1}, A^t) * R_p(B^{t_1}, C^t)$ , where '\*' is a placeholder that has to be substituted by either '<', or '=' or '>'. In what follows we are only interested in comparing two correlations and in the question of whether the correlation between two events is positive, negative or zero. Because of the biconditional relations just mentioned, these issues are

---

<sup>12</sup>By 'prediction', we typically refer to future events: we predict future events on the basis of information of past events. In other words, in predictions we typically mean cases in which  $t_1$  is earlier than  $t$ . For the sake of convenience, we shall follow this common terminology. It is noteworthy, however, that in deterministic, Hamiltonian systems, the direction of time has no particular significance.

insensitive to whether we use  $R$  rather than  $R_p$ . It will turn out that  $R$  is more useful when it comes to connecting the above notion of unpredictability to dynamical systems and for this reason we choose to work with  $R$  rather than with  $R_p$ .

This basic notion of unpredictability can be refined in various ways. First, the type of unpredictability we obtain depends on the type of ‘events’ to which (R) is applied. For example, the degree of the unpredictability of  $A^t$  increases if its probability is independent not only of  $B^{t_1}$  or other ‘isolated’ past events, but the entire past. For another example, the unpredictability of  $A^t$  increases if its probability is independent not only of the actual past, but also of any possible past. Second, unpredictability of a future event  $A^t$  increases if the probabilistic dependence of that event on past events  $B^{t_1}$  decreases rapidly with the increase of the temporal distance  $t - t_1$  between the events. Third, as we shall see below, there could be various types of probabilistic independence, which may indicate different degrees of unpredictability. For example, the probability of  $A^t$  may be independent of past events *simpliciter*, or it may be independent of such events only ‘on average’.

In sum, our suggestion is to explicate the notion of degrees of randomness in terms of degrees of unpredictability and to explicate unpredictability in terms of probabilistic dependence between states of systems at different times (‘events’). In the next section we show how, given plausible assumptions, epistemic probabilities used so far should be constrained by the dynamical properties of the system.

### 3.4 Epistemic probabilities and dynamical systems

How can we determine the values of the probabilities in the above formulae? If the probabilities are to be useful to understanding randomness in dynamical systems, the probability assignment has to reflect the properties of the system. That is, in order for unpredictability as introduced above to reflect the random behaviour of a dynamical system, the relevant probabilities cannot be merely subjective degrees of belief. In order for subjective probabilities to reflect objective dynamical randomness, the probabilities have to reflect dynamical properties of systems.

Given that we are working with a dynamical system that is by assumption endowed with a probability measure  $\mu$ , it is natural to suggest that the relevant probabilities should be associated with the measure  $\mu$ . Indeed,  $\mu$  *per se* need not reflect ignorance about the system's state. A measure on a phase space can have a purely geometrical interpretation and need not necessarily be interpreted as reflecting our ignorance. But, we are free at the outset to interpret our dynamical system  $[X, B, \mu, \tau]$  as we like. In particular, we may regard  $\mu$  as reflecting our ignorance. Alternatively, if  $\mu$  is already based on facts about the dynamical behaviour of the system (or systems of the same type), we can fix our ignorance according to it.

While this general idea is straightforward, the question of the exact relation between the measure  $\mu$  and the probability  $p$  is non-trivial. The problem is the following: the measure  $\mu$  takes subsets  $A$  of the phase space as arguments, while the probability  $p$  takes events  $A^t$  as arguments. As the notation indicates, the main difference between the two is that the events are time-indexed whereas sets are not. The main idea of how to relate  $\mu$  to  $p$  is to postulate that the probability of an event  $A^t$  is the measure of  $A$ . Formally:

( $\mu$ -Stat) For all instants of time  $t$  and for all  $A \in \Sigma$ :  $p(A^t) = \mu(A)$ .

This principle assumes that the probability measure  $p$  is stationary and that that  $p$  is determined by  $\mu$ . For this reason we refer to this principle as ( $\mu$ -Stat).<sup>13</sup>

( $\mu$ -Stat) can be generalised to joint simultaneous events:

---

<sup>13</sup> It is worth pointing out that we assume stationarity because it is the common way to justify assumptions of the kind we are making here. From a logical point of view, however, the following weaker condition would do for our purposes:

For some particular instant of time  $t$ :  $(\forall A \in \Sigma)[p(A^t) = \mu(A)]$ .

Basically, this condition states that there *is some particular instant of time* – the reference point, if you like – at which it is true for all sets  $A \in \Sigma$  that  $p(A^t) = \mu(A)$ . Our analysis can then also be based on this weaker link, regardless of whether the measure  $\mu$  and the probability distribution  $p$  coincide at other points of time. This weaker condition is implied by stationarity, but it does not itself imply stationarity.

( $\mu$ -Stat\*) For all instants of time  $t$  and for all  $A, B \in \Sigma$ :  $p(A^t \& B^t) = \mu(A \cap B)$ .

We can further generalise ( $\mu$ -Stat) to joint non-simultaneous events:

( $\mu$ -Dyn) For all instants of time  $t$ , for all times  $t_1 \leq t$ , and for all  $A, B \in \Sigma$ :

$$p(A^t \& B^{t_1}) = \mu(A \cap \phi_{t_1 \rightarrow t} B).$$

where  $\phi_{t_1 \rightarrow t} B$  is the evolution of the subset  $B$  forward in time from  $t_1$  to  $t$ . The LHS of this equation denotes the probability that system is in  $B$  at  $t_1$  and then in  $A$  at some later time  $t$  whereas the RHS denotes the measure of the intersection of the set  $A$  and the set of points in  $\phi_{t_1 \rightarrow t} B$ , which is the evolution of the subset  $B$  forward in time from  $t_1$  to  $t$ . The validity of this formula becomes clearer in Fig. 2.

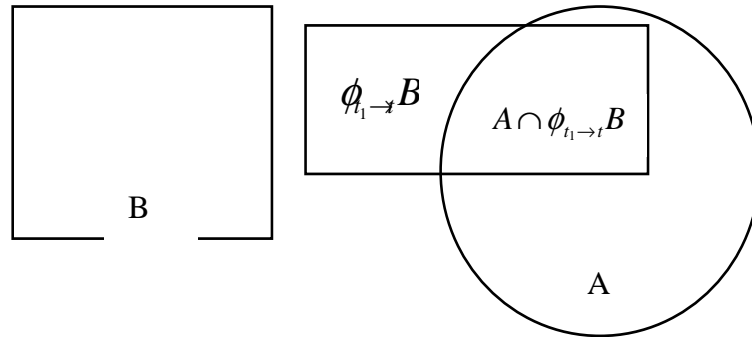


Fig. 2. An illustration of ( $\mu$ -Dyn).

It is worth pointing out that the rule ( $\mu$ -Dyn) is not an independent assumption because it follows from ( $\mu$ -Stat), the law for joint probabilities, and the requirement that our degrees of belief are updated according to the dynamical law of the system (which, recall, is deterministic). This can be seen as follows. From the fact that the time evolution  $\phi$  is deterministic it follows that every point  $x$  of  $X$  which lies within  $\phi_{t_1 \rightarrow t} B$  lies on a trajectory that passed  $B$  at time  $t_1$ , and no other point

does. Therefore the points which lie within  $A \cap \phi_{t_1 \rightarrow t} B$  are exactly those who lie on trajectories that pass  $A$  at time  $t$  and went through  $B$  at time  $t_1 < t$ . In other words,  $A \cap \phi_{t_1 \rightarrow t} B$  contains all points for which  $A^t \& B^{t_1}$  is true. Now we can use the rule for joint probabilities along with ( $\mu$ -Stat) and get  $p(A^t \& B^{t_1}) = \mu(A \cap \phi_{t_1 \rightarrow t} B)$ , which obviously holds for all  $A, B \in \Sigma$  since we have not made any assumptions about either  $A$  or  $B$  and for all  $t_1 < t$ .<sup>14</sup>

The condition ( $\mu$ -Dyn) readily generalises to more than two events. Consider  $r+1$  instants of time:  $t > t_1 > \dots > t_r$  and  $r+1$  sets  $A, B_1, \dots, B_r$  of  $\Sigma$  (where again  $r$  is any natural number). Then we have:

For all instants of time  $t$ , for all times  $t_r \leq \dots \leq t_1 \leq t$ , and for all  $A, B_1 \dots B_r \in \Sigma$ :

$$p(A^t \& B_1^{t_1} \& \dots \& B_r^{t_r}) = \mu(A \cap \phi_{t_1 \rightarrow t} B_1 \cap \dots \cap \phi_{t_r \rightarrow t} B_r).$$

So far we have been dealing with continuous time for heuristic purposes. However, the above results easily carry over to the case of discrete time. In this case the time evolution is induced by a mapping, in which case we have  $\phi_{t_1 \rightarrow t} B_i = T^{(t-t_1)} B_i$ , where  $i = 1, \dots, r$ . Now let  $n_i := t - t_i$  be the time difference between  $t$  ('now') and the past instant  $t_i$ . Then we obtain:

( $\mu$ -DynT) For all instants of time  $t$ , for all natural numbers  $n_r \geq \dots \geq n_1$ , and for all  $A, B_1 \dots B_r \in \Sigma$ :  $p(A^t \& B_1^{t_1} \& \dots \& B_r^{t_r}) = \mu(A \cap T^{n_1} B_1 \cap \dots \cap T^{n_r} B_r)$ .

It is important to note that it follows from ( $\mu$ -DynT) that correlation and relevance are related as follows:

$$(RC) \quad R(B^{t_1}, A^t) = C(T^{t-t_1} B, A).$$

On the basis of (RC) we will formulate below the basic relations between correlation, unpredictability, and dynamical randomness.

---

<sup>14</sup> Trivially, ( $\mu$ -Dyn) implies ( $\mu$ -Stat). Yet, for conceptual clarity it is worth presenting them separately.

Summing up, we suggest that the notion of degrees of randomness is to be explicated in terms of degrees of unpredictability and that unpredictability is to be explicated in terms of epistemic dependence, which is to be informed by objective features of the system through (RC). In the next section, these basic connections between epistemic unpredictability, lack of correlation, and randomness will be used to explicate the degrees of dynamical randomness involved in the different levels of EH.

Before we use (RC) to understand EH as a hierarchy of randomness, let us briefly deal with the issue of how to interpret the ignorance probabilities we are dealing with. So far we have considered the measure  $\mu$  as given. This is in line with the fact that  $\mu$  is a part of the definition of a dynamical system (see Sec. 2). However, one may well want to ask where  $\mu$  comes from and ‘what it means’. A question very similar to this has been extensively discussed in the context of the foundation of statistical mechanics and we therefore only briefly want to hint at the two most prominent answers, the ensemble interpretation and the time-average interpretation of  $\mu$ . We can be brief on this, as our analysis remains valid on both understandings of  $\mu$ .

Imagine that we start only with a triple consisting of a phase space  $X$ , a sigma algebra  $\Sigma$ , and an automorphism  $\tau$ , and then proceed to construct a measure  $\mu$ . According to the so-called ensemble interpretation,  $\mu$  is a measure over tokens with the same phase space  $X$ , sigma algebra  $\Sigma$ , and group  $\tau$  of automorphisms, but different initial conditions. The measure of a certain set  $A \in \Sigma$  at time  $t$  then is defined as the fraction of these items whose state is in set  $A$  at  $t$ ; if this fraction is independent of  $t$ , the measure is stationary.

On the time-average interpretation, the measure  $\mu$  of a set  $A \in \Sigma$  of the phase space is defined as the long-run average of the fraction of time that the system’s state spends in  $A$ .<sup>15</sup> This measure can then naturally be regarded as the probability that the system’s state is in set  $A$ . In the absence of knowledge of the system’s state, it is natural to base one’s beliefs on  $\mu$  thus construed. Thus,

---

<sup>15</sup> The exact definition of ‘long-term time average’ does not matter for our purposes. However, it is important that the average converges towards a value that remains constant.



the relation between unpredictability, lack of correlations and randomness are exactly the same as in the ensemble interpretation.

On both interpretations, the formal relations between unpredictability, lack of correlations and randomness are the same. Degrees of unpredictability are explicated in terms of decay of correlations, where the basic relation between unpredictability and decay of correlation is expressed by (RC), and degrees of dynamical randomness are explicated in terms of degrees of unpredictability. But the meaning of randomness is not the same. Under the time average interpretation, it is natural to think of the degree of randomness of a system as reflecting an intrinsic property of the system, whereas under the ensemble interpretation of randomness, it is natural to think of randomness as reflecting a property of an ensemble.

### 3.5 The ergodic hierarchy as a hierarchy of degrees of unpredictability

We now turn to explicate how the ergodic hierarchy (EH) can be interpreted as a hierarchy of increasing degrees of randomness. The basic idea is simple: given (RC) the pattern of decay of correlations characteristic of a particular dynamical system is translated into the relevance of past states to the prediction of the future states.

#### 3.5.1. Ergodicity

To understand the kind of randomness we find in ergodic systems, we ‘translate’ (C-E) into epistemic probabilities in the following way. We set (without loss of generality)  $t = 0$ , i.e. we chose  $t$  to be ‘now’ and then apply (RC). (C-E) then becomes  $\lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=1}^{k-1} R(B^i, A^0) = 0$ , where  $t_i = -i$ . This can readily be interpreted as the average of the sum of the relevance of being in  $B$  at  $t = -1$  to the probability of being in  $A$  at  $t = 0$ , the relevance of being in  $B$  at  $t = -2$  to the probability being in  $A$  at  $t = 0$  etc. approaching zero as time tends towards minus infinity. This does not mean, however, that any individual relevance of  $B^i$  to the probability of  $A^0$  has to be zero. The infinite average being zero is compatible with  $R(B^i, A^0) \neq 0$  for all  $t_i$ . Since we can

choose  $A$  and  $B$  as we like, this is true for any current state  $A$  and any past state  $B$ . Hence, ergodicity allows for the past to be highly relevant to prediction of the probabilities of future events. That is why ergodicity *per se* need not involve any randomness.

### 3.5.2. Mixing

By contrast with ergodicity, mixing involves some degree of randomness. Applying (RC) to (C-SM) and again setting  $t = 0$  we obtain:  $\lim_{k \rightarrow \infty} R(B^{t_k}, A^0) = 0$ , where  $t_k = -k$ . Then SM can be interpreted as saying that for any two sets  $A, B \in \Sigma$ , having been in  $B$  at time  $t_k = -k$  becomes irrelevant to the probability of being in  $A$  now ( $t = 0$ ) as  $k$  tends towards infinity. In other words, the past event  $B^{t_k}$  is irrelevant to our ability to predict the probability of the present event  $A^0$ , as the past becomes infinitely remote. More generally, in SM the relevance of the system's states in the remote past for the prediction of the probability of the system's current state decreases as the temporal distance between these states grows; and it vanishes completely in the limit of the temporal distance tending towards infinity. Thus, strong mixing involves a substantial degree of dynamical randomness.

Weak mixing also involves some degree of randomness. Again applying (RC), (C-WM) readily implies  $\lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=1}^{k-1} |R(B^{t_i}, A^0)| = 0$ , where  $t$  and  $t_i$  have been chosen as above. That is, the average of the sum of the modulo of the relevance of being in  $B$  at  $t = -1$  to the probability being in  $A$  at  $t = 0$  and the modulo of the relevance of being in  $B$  at  $t = -2$  to the probability of being in  $A$  at  $t = 0$  etc. approaches zero as time tends towards minus infinity. Since we can choose  $A$  and  $B$  as we like, this intuitively means that almost all past events that have or could have occurred are irrelevant to predicting the probability of present events. Since it is the average of the modulo of the relevances that has to approach zero, WM involves a weaker degree of unpredictability than SM, but a stronger degree than E.

### 3.5.3. K-systems and correlations

In K-systems both (C-KM-1) and (C-KM-2) obtain. As observed in subsection 4.2.3, (C-KM-1) equivalent to strong mixing. The crucial novelties come into play with (C-KM-2). Applying (RC),

(C-KM-2) implies  $\lim_{n \rightarrow \infty} R(A_{m_1}^{t_n} \& A_{m_2}^{t_{n+1}} \& A_{m_3}^{t_{n+2}} \dots, A_0) = 0$ , with  $t_i = -i$ . This means that as  $n$  tends towards infinity the fact that the system's state was in  $A_{m_1}$  at  $t = -n$ , in  $A_{m_2}$  at  $t = -(n+1)$ , and so on becomes irrelevant to our ability to predict the probability of the present state. By assumption  $A_1, \dots, A_r$  form a partition and therefore  $T^n A_{m_1} \cap T^{n+1} A_{m_2} \cap T^{n+2} A_{m_3} \cap \dots$  may be chosen so that it represents the *entire infinitely remote past coarse-grained history* of the system. Thus (C-KM-2) implies that the probability of the present state  $A_0$  becomes independent of the *entire infinitely remote past coarse-grained history* of the system. Moreover, since we are free to choose the  $A_i$  as we like, this will be true for any refinement of such infinitely remote (coarse-grained) past history (provided that it is still a coarse-grained history).

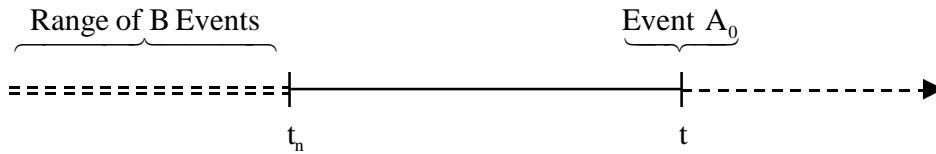


Fig. 3. The remote past of the system

There may be some correlations between a system's remote past history, i.e. between its remote past and present states, but they asymptotically approach zero: the more remote the past history is the weaker the correlations are.

This is the significant difference between K and SM. While SM only requires probabilistic independence of the system's present state from its states at particular instants of time in the remote past, K requires probabilistic independence from the system's entire remote past history.

Since we are free to choose the  $A_{m_j}^{t_i}$  as we wish, we can 'patch together' whatever sequence we like, including any sequences that denote other possible, infinitely remote past histories. Accordingly, in K-systems the system's present state is independent not only of the *actual* remote past history of the system but also of *any* of its possible infinitely remote past histories (i.e. any infinitely remote (coarse-grained) past history that could evolve to the present state  $A_i$ ).

### 3.5.4. Bernoulli systems

While in SM and KM, the decay of correlation occurs only in the infinite limit, in B systems the decay of the correlations is immediate: the present stage of a B-system is uncorrelated with any of its past (and future) stages. For example, in a series of coin tosses the current outcome does not depend on previous outcomes.

Applying (RC) yields, (C-B) implies  $R(A_i^{t_n}, A_j^0) = 0$  for all integers  $n$ , where  $t_n = -n$ ; and similarly for (C-B\*). Hence, the probabilities of the present state are totally independent of whatever happened in the past, even if we have perfect knowledge of the entire (coarse-grained) history of the system. Thus, B-systems involve the highest degree of randomness in EH.<sup>16</sup>

## 4. Using EH: Characterising randomness in physical systems

In the previous sections we have introduced EH and provided a uniform characterisation of the underlying concept of randomness. In the remainder of the paper we shall discuss the relevance of EH for characterising randomness and chaos in classical physics.

### 4.1. The ‘no-application’ charge

A popular objection to ergodic theory is that it is not more than a sophisticated but eventually useless piece of mathematics since Hamiltonian systems are typically non-ergodic. This bit of

---

<sup>16</sup> In a sense, B systems are ‘anomaly’ in EH; for it is the only level of the hierarchy in which correlations decay ‘immediately’ rather than asymptotically. There are systems who lie ‘in-between’ B and K in this sense, namely Markov systems of order  $n$ . In these systems the current state correlates with the previous  $n$  states. Space constraints prevent us from further discussing these systems.

conventional wisdom is backed-up by a theorem by Markus & Mayer (1974) which is based on KAM theory and basically says that generic Hamiltonian dynamical systems are not ergodic.

There are two strategies to counter this objection. The first is to mitigate its force by emphasising that it is not the ‘sheer number’ of the applications that makes a physical concept important. Rather, the question is whether the systems that we are interested in are ergodic. And at least some of them are. Most importantly, hard sphere models of the ideal gas, which after all are paradigm systems of statistical mechanics, are K-systems and hence ergodic. We discuss this case in detail in section 4.2 below.

The second strategy is to shift the focus from continuous to discrete transformations. To this end notice that the aforementioned theorem applies only to continuous Hamiltonian systems of the kind we find in classical mechanics; but it has no force in the case of mappings of the sort we have been dealing with in this paper. Moreover, the great number of ergodic mappings we find in the literature suggests that in this case (unlike in the case of continuous systems) ergodicity is the standard rather than the exception, or at least central enough to justify serious consideration. This shifts the focus of our attention. What we need is an argument that maps of this sort are relevant to physical problems. In Sections 4.3 through 4.5 we argue that they are indeed.

## 4.2 The relevance of ergodic theory in light of the KAM theory

The relevance of ergodic theory (and consequently the ergodic hierarchy) for the effective modelling of actual physical systems has come into serious question due to KAM theory (named after Kolmogorov, Arnold, and Moser). The discussion below shows that the KAM theory does not, as is often alleged, undermine the status of ergodic theory in the manner indicated above.

The term “KAM theory” is used here to refer to a theorem of Kolmogorov and its many consequences. Kolmogorov (1954) formulated the theorem, which concerns Hamiltonian systems and perturbation theory, and provided an outline of a proof. The theorem basically says that if a small perturbation is added to an integrable system with two physical degrees of freedom, then the tori with a sufficiently irrational winding number survive the perturbation. Arnold (1963) and

Moser (1962) independently provided rigorous proofs of it; so, it is now usually referred to as the “KAM theorem”. Markus and Meyer (1974) developed some important consequences of the KAM theorem, and one of these is a theorem that is pertinent to the relevancy issue. For the purposes of this essay, we shall refer it as the “Markus-Meyer theorem.”<sup>17</sup> We turn now to discuss this theorem.

The Markus-Meyer theorem is often informally characterized as the claim that generic Hamiltonian dynamical systems are not ergodic. This is a striking claim and its effect appears to undermine the relevancy of ergodic theory is immediate. But this effect is mitigated as soon as the theorem is stated in full: The set of Hamiltonian flows that are ergodic on the energy hypersurfaces associated with each element of a dense set of energy values is of first category in the set of infinitely differentiable Hamiltonian flows. In short, the class of ergodic Hamiltonians is of first category in the set of generic Hamiltonians. A set  $P$  is of first category in set  $Q$ , if  $P$  can be represented as a countable union of nowhere dense sets in  $P$ . All other sets are of second category. Loosely speaking, a set of first category is the ‘topological counterpart’ to a set of measure zero in measure theory, and it is sometimes referred to as “meager”. The contrast class is “second category”, which is sometimes referred to as “non-meager” and is the ‘topological analogue’ of a set of non-zero measure (see Kelley, 1975, p. 201). It is worth noting, however, that these two notions can (though need not) be diametrically opposed. For example, the real line can be decomposed into two complementary sets  $A$  and  $B$  such that  $A$  is of first category and  $B$  is of measure zero – see theorem 1.6 in Oxtoby (1971).

The key phrase that has the mitigating effect is *infinitely differentiable*.<sup>18</sup> This restriction is substantial. It actually rules out a large class of Hamiltonian systems that are of physical interest particularly for classical statistical mechanics, as will be shown below. Before doing so, it is worth contrasting the Markus-Meyer theorem with a theorem (the first) of Oxtoby & Ulam (1941): In the set of measure-preserving generalized dynamical flows,<sup>19</sup> the set of dynamical

---

<sup>17</sup> It is the second of two important theorems that Markus and Meyer prove in their memoir. The first says that generic Hamiltonian systems are not integrable (in the sense of Liouville).

<sup>18</sup> Markus & Meyer (1974) introduce this restriction on page 4.

<sup>19</sup> They are “generalized” in the sense that they are not necessarily differentiable or derivable from a set of differential equations.

flows that are *not* ergodic is of first category. The class of Hamiltonian flows of physical interest is substantially broader than the class considered in the Markus-Meyer theorem, but substantially narrower than that considered in the Oxtoby-Ulam theorem.

Among the Hamiltonian flows that are excluded from consideration in the Markus-Meyer theorem are those for systems that involve particle collisions. They are excluded because collisions, which are typically modeled using an extremely steep repulsive hard-core potential, result in singularities in the dynamical flow; thus, the flow will be continuous everywhere but not differentiable everywhere. Such models are now referred to as “hard ball systems”. They may be conceptualized concretely as collections of hard balls having a finite radius but no rotational motion that interact by elastic collision, and they (and more sophisticated variants of them) are effective idealizations of the dynamics of gas molecules.

There are other models aside from hard ball systems that involve singularities such as billiards and the Lorentz gas, and they are just as important as hard ball systems for the purposes of characterizing real physical systems.<sup>20</sup> But, it will suffice for the purposes at hand to consider hard ball systems. Boltzmann (1871) studied hard ball systems in developing a mathematical foundation for statistical mechanics. He conjectured that such systems are ergodic when the number  $N$  of hard balls is sufficiently large.<sup>21</sup> Whether this is the case is still an open question. However, it turns out that hard ball systems are ergodic for some small  $N$ . Sinai (1970) provided the first rigorous demonstration that a hard ball system is ergodic. He did so for a pair of colliding two-dimensional disks moving on  $T^2$  (the unit two-torus) seven years after he put forward what is now called the “Boltzmann-Sinai ergodic hypothesis” (1963). This hypothesis says that a system of  $N$  hard balls on  $T^2$  or  $T^3$  is ergodic for any  $N \geq 2$ , and it was inspired by Krylov’s observation, made in 1942 that hard ball systems exhibit an instability that is similar to geodesic flows on hyperbolic surfaces, known at the time to be ergodic (Krylov 1979).

---

<sup>20</sup> A *billiard system* is commonly taken to be a system having a finite two-dimensional planar phase space with a piecewise smooth boundary in which a point particle moves with constant velocity and bounces off the wall in an optical manner (i.e. the angle of incidence is assumed to be equal to the angle of reflection). A Lorentz gas is a dynamical system of one point particle moving in the complement of regular lattice of rounded scatters.

<sup>21</sup> The discussion of hard ball systems below is inspired by a brief historical survey by Szász (1996), reprinted in Szász (2000), and a survey by Simányi (2000).

Since Sinai's groundbreaking work, there have been some notable advances in understanding hard ball systems. The ergodic hierarchy plays a central role in characterizing many of them as it does in the ones considered here. Simányi (1992) demonstrated that a system of  $N$  hard balls is K-mixing (and hence ergodic) on  $T^m$  for  $N \geq 2$  and  $m \geq N$ . This result falls short of what is desired since the minimal size of  $m$  is restricted by the size of  $N$ , but it is still significant. Chernov & Haskell (1996) showed that any system of hard balls or disks on a torus or in any container is Bernoulli, provided that it is K-mixing and completely hyperbolic.<sup>22</sup> Recently, Simányi and Szász (1999) demonstrated that a system of  $N$  hard balls on  $T^m$  is completely hyperbolic for  $N \geq 2$  and  $m \geq 2$ . Combining the above one can derive that a system of  $N$  hard balls on  $T^m$  is Bernoulli for  $N \geq 2$  and  $m \geq N$ . Of course, a rectangular box is more realistic than a torus. It is also more difficult to treat.<sup>23</sup> Nevertheless, Simányi (1999) demonstrated that the dynamics of  $N$  hard spheres in an  $m$ -dimensional box is ergodic for  $N = 2$  and  $m \geq 2$ .<sup>24</sup> The same result is expected but has not yet been demonstrated for any  $N > 2$ .

### 4.3 The surface of section approach

Our second line of defence against the claim that EH is useless because most systems are not ergodic is that EH may be used to characterise randomness in systems that are not strictly speaking ergodic. The main idea is that systems that are not ergodic with respect to the relevant phase space may well display ergodic behaviour on some subspace of it.

The reasoning is as follows. So far we have been dealing with abstract dynamical systems whose dynamics is defined on an abstract state space  $X$ . In order to bring to bear this notion of a dynamical system on a 'real' physical system as treated within the context of classical Hamiltonian mechanics, certain contextualisations are needed. The full state space of a classical

---

<sup>22</sup> Complete hyperbolicity means that all Lyapunov exponents are nonzero almost everywhere.

<sup>23</sup> Conceptually speaking, one obtains a torus by replacing the boundary conditions corresponding to the walls of a rectangular box with periodic boundary conditions (meaning that parallel sides of the box are identified), which are much easier to utilise mathematically than the former.

<sup>24</sup> The ratio of the length of the sides to the radius of the balls must be less than  $1/6$ . There are no requirements concerning the equality of the sides of the container, the masses of the balls, or the radii of the balls.



system is the so-called phase space, basically consisting of a position and momentum dimension for every of the  $n$  physical degree of freedom, denoted by  $q = (q_1, \dots, q_n)$  and  $p = (p_1, \dots, p_n)$  respectively. For instance, the phase space of a point mass moving on plane is four dimensional, having two position and two momentum dimensions. The motion of the system's state in this space is governed by Hamilton's equation of motion,  $dq_i / dt = dH(q, p) / dp_i$  and  $dp_i / dt = -dH(q, p) / dq_i$ , where  $H(q, p)$  is the so-called Hamilton function (or Hamiltonian, for short) of the system. Under most circumstances (i.e. if the system's motion is not subject to explicitly time-dependent geometrical constraints on its motion)  $H(q, p)$  is the energy of the system. The solutions of Hamilton's equation are commonly referred to as 'Hamiltonian flow'. For details we refer the reader to Goldstein's (1980) classical book on mechanics.

Due to geometrical constraints on the motion and conservation laws not the entire phase space is accessible. The most important of these restrictions in the context of Hamiltonian systems is the conservation of energy:  $H(q, p) = E$ . This restricts the motion on a hypersurface, which is commonly referred to as the 'energy shell'. It is now natural (and common) to associate the energy shell of a Hamiltonian system with  $X$  of an abstract dynamical system.

It is customary to take the energy shell as the relevant portion of the phase space; and when a system is claimed not to be ergodic it is usually tacitly assumed that this it is not ergodic with respect to this hypersurface. It is important to bring this hidden assumption to the fore because the fact that the dynamics of a system is not ergodic with respect to the energy shell does *not* preclude it from being ergodic with respect to another (more restricted) portion of the phase space. This is what Lichtenberg & Lieberman have in mind when they observe that '[i]n a sense, ergodicity is universal, and the central question is to define the subspace over which it exists' (Lichtenberg & Lieberman 1992, p. 295). So EH can be used to characterise randomness in dynamical systems that are not ergodic with respect to the entire energy shell, the main idea being to consider the system's behaviour on a subspace of this shell where it may well be ergodic.<sup>25</sup>

---

<sup>25</sup> One might even conjecture that this method can be extended beyond Hamiltonian systems to dissipative systems. In this vein Ornstein (1989) writes: "The main example that I used to illustrate our results was billiards. These results, however, apply to a large class of dissipative systems [...] I conjecture that long-term statistical averages exist for dissipative as well as conservative systems."

The crucial question for the success of this line of defence obviously is whether such subspaces exist. The aim of this subsection is to show that they do; and the strategy is to study what happens in a surface of section rather than to focus on the full-fledged continuous dynamics.

Consider a continuous Hamiltonian system with two (physical) degrees of freedom. (For the sake of simplicity and convenience, in what follows we limit our attention to such systems. But, the techniques described below equally apply to systems with a greater number of degrees of freedom.) The phase space of such a system (assuming that it is autonomous) is four-dimensional.<sup>26</sup> Now, observe that since by assumption the energy is conserved, the motion takes place on a three-dimensional hypersurface in the phase space defined by the condition  $H(q, p) = E$ , where  $E$  is some constant. We now choose a two-dimensional surface  $\Sigma$  within this hypersurface and label its two sides  $\Sigma_L$  and  $\Sigma_R$ ;  $L$  and  $R$  denote left and right. We then study the successive points of intersection as the trajectory passes through  $\Sigma$  in a particular direction, say from  $L$  and to  $R$ . The points of intersection are labelled in succession  $x_n, x_{n+1}, x_{n+2}, \dots$  (see Fig. 4).

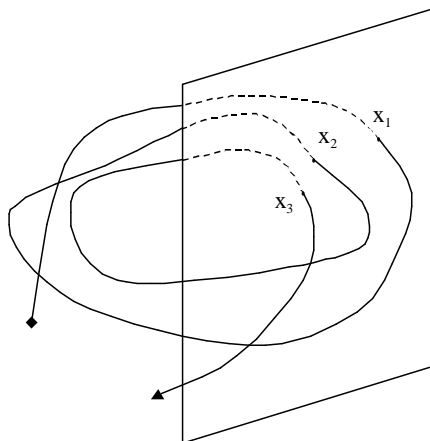


Fig. 4. A particle trajectory intersecting a surface at three points.

---

<sup>26</sup> A Hamiltonian system is autonomous, if the system's Hamiltonian does not explicitly depend on time, meaning that  $\partial H/\partial t = 0$ .

What we get in the surface of section is successive intersections and these can be described by a mapping. That is, in the surface of section the dynamics of the system is represented by a mapping  $T$  that takes a phase point  $x_i$  to  $x_{i+1}$ :  $x_{i+1} = Tx_i$ . Furthermore, one can prove that the discrete dynamics in the surface of section is area preserving (or, to be more specific, symplectic) as well (Tabor 1989, pp. 123-6). Hence, the continuous Hamiltonian flow generates an area-preserving mapping that describes a trajectory's successive crossings of the surface of section. This shows that there is a close connection between continuous Hamiltonian systems with two degrees of freedom and area preserving mappings of a two dimensional surface onto itself.<sup>27</sup>

Also, from the construction of  $T$  it is obvious that there is a close connection between the dynamics of the continuous system and the mapping. In fact, such mappings may display all typical properties of continuous systems such as the breaking up of tori and the occurrence of island chains (we will discuss an example below). For this reason we can use two-dimensional mappings to discuss the behaviour of continuous systems. This has great advantages since mappings are analytically and numerically much easier to handle than continuous flows and, owing to their relative simplicity, many theorems about dynamical systems are more easily proven for mappings than for general Hamiltonian (e.g. the Poincaré Birkhoff fixed-point theorem and the KAM theorem as proven by Moser (1973)).

Yet, although the connection between the continuous flow and the mapping in the surface of section is close, it is still loose enough to allow for mappings to have properties that the continuous flow as a whole does not possess. Most notably, even if the system is not ergodic, some of its trajectories can give rise to ergodic mappings in the surface of section. This insight (which will be illustrated in due course) paves the way to use mappings to discuss the behaviour of a system in some region of the phase space. More specifically, the study of ergodic maps,

---

<sup>27</sup>Two remarks: (1) It is noteworthy that the converse is also true for a broad class of mappings. As Lichtenberg & Lieberman (1992, p. 171) show, radial twist maps can be converted into a continuous Hamiltonian system by expansion of the transformation in a Fourier series. (2) In passing we would like to mention that besides the surface of section, there is another technique by which a continuous system (i.e. a system that is defined by a continuous transformation) can be converted into a mapping: the discretisation of time (sometimes also referred to as 'stroboscopic projection'). The method basically consists in taking 'snapshots' of the system at discrete times (Tabor, 1989, p. 125; Argyris *et al.*, 1994, p. 57).

mixing maps, etc. can help to understand the behaviour of Hamiltonian systems in various distinct regions of the phase space because they can serve as models of the flow when studied in the surface of section. Hence, ergodic maps can be used to ‘characterise’ random behaviour in certain regions of a continuous flow even if the flow itself is *not* ergodic.

To see how all this works we will now discuss KAM-type systems (i.e., ‘nearly integrable’ systems, meaning integrable systems plus a small perturbation). This class of systems is the touchstone for our claim that EH can be used to study the behaviour of dynamical systems even if they are not ergodic. The key feature of a KAM-type system is the presence of both regular and stochastic regions, each closed (so that the system as a whole is not ergodic with respect to the energy shell) but intermeshed with the other.

Consider an integrable Hamiltonian (for the details of what follows see for instance Tabor, 1989, pp. 126ff.; Lichtenberg & Lieberman, 1992, pp. 166ff.; Ott, 1993, pp. 229ff.). In this case the phase space is foliated into tori; that is, every trajectory is confined to one particular torus. When we now focus on one particular trajectory, it is easily seen that in the surface of section all points lie on a smooth curve, which corresponds to the intersection of the torus (on which the trajectory moves) with the surface of section.

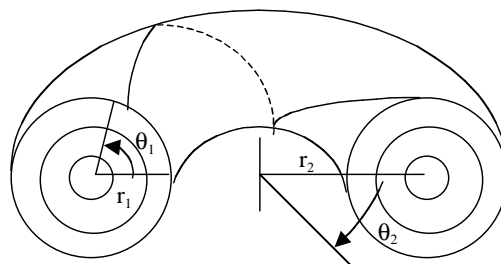


Fig. 5. A particle trajectory on the surface of a torus corresponding to constant energy  $E$ .

The motion of the particle on the surface of section may be represented by the so-called ‘twist map.’ Using polar co-ordinates, the mapping can be written as:

$$T: \begin{cases} \theta(i+1) = \theta(i) + 2\pi\omega_1/\omega_2 \\ r(i+1) = r(i) \end{cases}$$

where  $\omega_1$  and  $\omega_2$  are the angular velocities along  $\theta_1$  and  $\theta_2$  respectively, and  $\omega_1/\omega_2$  is the so-called ‘winding number.’ The elements  $r, \theta$  of the twist map may be respectively associated with  $r_1, \theta_1$  in Fig. 5 above. From Section 3 we know that if the ratio of the two frequencies,  $\omega_1/\omega_2$ , is irrational the sequence  $x_n, x_{n+1}, x_{n+2}, \dots$  fills up the curve ergodically, whereas if the ratio is rational only a finite sequence of iterates, corresponding to a closed orbit, appears.

Now we add a small perturbation to the original (integrable) Hamiltonian. This perturbation changes the twist map as follows (Tabor, 1989, p. 128):

$$T_\varepsilon: \begin{cases} \theta(i+1) = \theta(i) + 2\pi\omega_1/\omega_2 + \varepsilon f(\theta(i), r(i)) \\ r(i+1) = r(i) + \varepsilon g(\theta(i), r(i)) \end{cases}$$

where  $\varepsilon$  is a small perturbation parameter;  $f$  and  $g$  are periodic in the angle and are chosen such that the resulting transformation is area-preserving. The natural question to ask about this transformation is what happens to the invariant circles of the unperturbed mapping when the perturbation is ‘switched on’. The KAM theorem answers this question by stating, roughly, that for sufficiently small perturbations circles with a sufficiently irrational winding numbers  $\omega_1/\omega_2$  are preserved.

On the other hand, the Poincaré-Birkhoff fixed-point theorem informs us that the tori with a rational winding number get destroyed under the perturbation and are replaced by an even number of fixed-points. These fixed-points are alternately stable (i.e. elliptical) and unstable (i.e. hyperbolic). Schematically we can represent this as shown in Fig. 6.

Note that it follows immediately from this that KAM-type systems are not ergodic (and a fortiori do not possess any other property of EH either): Ergodic systems are irreducible (see section 2), i.e. their phase space is not separable in two (or more) regions that are mapped onto themselves by the time evolution. But this is exactly what happens in a KAM system. The invariant curves surviving the perturbation divide the phase space into distinct parts to which a trajectory remains

confined for all times because no trajectory can move across an invariant torus and therefore it can never get from one part into the other.

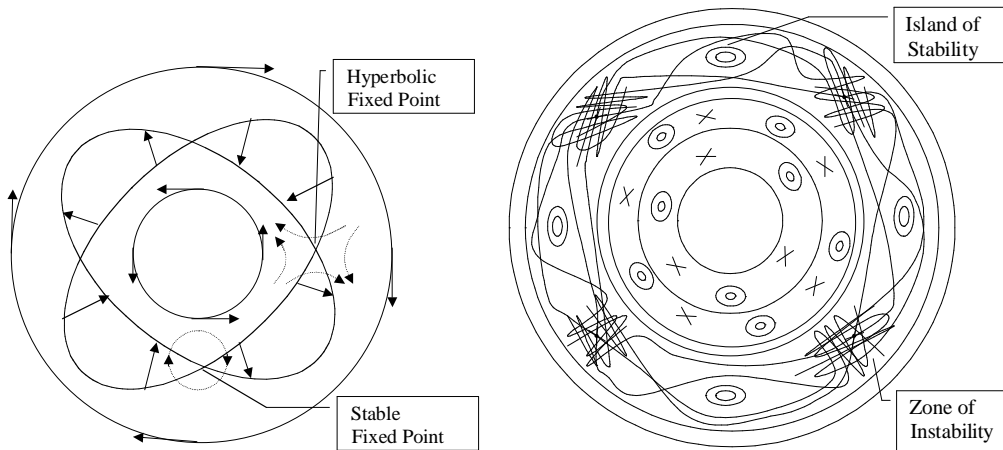


Fig. 6. The diagram on the left exhibits two stable points and two hyperbolic points. The diagram on the right shows in more detail stable points corresponding to islands of stability and hyperbolic points corresponding to zones of instability.

As a concrete example of such mapping consider the (by now famous) Henon system, which involves a quadratic twist map (a thorough discussion is provided by Argyris *et al.*, 1994, pp. 110-26; brief summaries can be found in Lichtenberg & Lieberman, 1992, pp. 191-195 and Tabor, 1989, pp. 129-32). The transformation, now in Cartesian co-ordinates, reads:

$$T: \begin{cases} x(i+1) = x(i)\cos\alpha + (y(i) - (x(i))^2)\sin\alpha \\ y(i+1) = x(i)\cos\alpha + (y(i) - (x(i))^2)\cos\alpha \end{cases}$$

where  $\alpha$  is a parameter and  $T: X \rightarrow X$  where  $X = [-1,1] \times [-1,1]$ .

This is a clear illustration of the above-mentioned ‘mixture’ of regions with chaotic and regular behaviour. There are island chains corresponding to irrational tori surviving the perturbation. The circles correspond to elliptical fixed-points; and the motion in the surrounding of these points is quite regular. This is not the case for the hyperbolic fixed-points. When the scale is enlarged an

incredibly rich, fine structure interspersed in a ‘sea of irregularity’ shows up that is analogous to that exhibited in Fig. 7 having similar zones of instability and islands of stability.

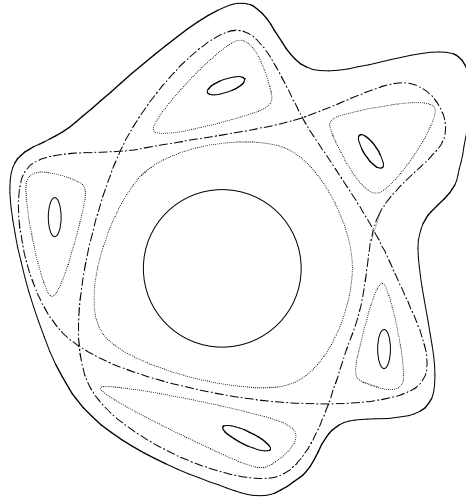


Fig. 7. A phase plane of the Henon map

But what do we mean by ‘quite regular’ or ‘sea of irregularity’? It is at this point where ergodic theory comes into play because it provides notions that can be used to characterise the system’s behaviour in different regions of the surface of section. To begin with, we know from the KAM theorem that the remaining closed lines are the intersections of irrational tori that survived the perturbation with the surface of section. From that it follows that the motion on these curves is ergodic.

Further one can show that near hyperbolic fixed points it is possible to embed a Bernoulli shift (with an infinite alphabet) in the phase space. From a mathematical point of view, all this is rather involved and we cannot go into details here. (The original source is Moser (1973, Ch. 3). A more intuitive description of the situation can be found in Reichl (1991, pp. 76-80).) However, the upshot is that if we choose an arbitrary point near a hyperbolic fixed-point its trajectory will behave like a B-system. For this reason we can say that in the region close to such a fixed-point the system behaves like a Bernoulli system and therefore exhibits all the random properties these systems possess.

Furthermore, there is strong numerical evidence that each connected stochastic region in the phase space, such as a separatrix layer, has positive Kolmogorov Sinai entropy and thus the system exhibits the same behaviour as a K-system in these regions (Lichtenberg & Leiberman 1992, 309).

Since it is possible to characterise the dynamics of a *non*-ergodic continuous flow in certain regions of the phase space can be characterised in terms of maps that are (merely) ergodic or Bernoulli, the fact that the flow as a whole is not ergodic does not undercut the usefulness of ergodic theory as a tool for characterising a system's random behaviour.

We should mention, however, that we have not been able to find any cases in the literature where the dynamics of the system in some region of the phase space has been described in terms of merely mixing. One may just speculate as to what the reasons for this absence are. Looking at the enormous mathematical sophistication that is needed to show that certain regions exhibit Bernoulli behaviour, one could guess that similar proofs for other mixing just may not be available. Another reason might be that due to the lack of conceptual clarity in the ergodic hierarchy as a whole, physicists did not find it useful to characterise dynamic properties using the notions of strong or weak mixing. Be this as it may, given the unified treatment of the ergodic hierarchy we presented in the previous section, we think that our analysis could add to the understanding of the character of randomness in classical ergodic and non-ergodic systems.

#### 4.4 Maps as analogues of continuous flows

Ergodic maps are also used as 'analogue pictures' of continuous flows. An analysis of EH can be used to study the degree of randomness involved in such maps and accordingly the dynamical randomness involved in the corresponding continuous flows. Systems that are commonly considered as chaotic have positive Lyapunov exponents and hence nearby orbits diverge exponentially. But this cannot be the case for all directions since this would not be compatible with the fact that the phase flow is area preserving. Therefore, while nearby trajectories diverge in one direction they converge negative-exponentially in some other direction so that the sum of all Liapunov exponents is zero. More precisely, the phase decomposes at every point into three



subspaces: The first is simply the trajectory of the system through that point. The second is a manifold of trajectories that exponentially diverge from the trajectory of our reference point, and the third is a manifold consisting of trajectories that exponentially converge towards that same trajectory. Hence, what we end up with is that every finite element of the phase space shrinks in one direction while it spreads in another and as a result of this bundles of trajectories get squeezed into something like a thin sheet. However, chaotic motion is characteristically bounded, i.e. it is confined to a certain region of the phase space. As a consequence, this squeezing of trajectories cannot continue forever. What happens is that, in order to secure confinement to such region, a ‘folding back’ of the sheet onto itself is taking place. Schematically, this can be visualised as follows shown in Fig. 8.

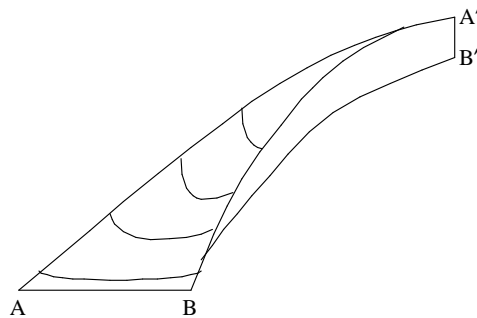


Fig. 8. The folding back can be envisioned by conceptually identifying edge AB with edge A'B'.

It is at this point that ergodic maps enter the scene. Looking at the above picture, what we identify as the basic pattern is a motion consisting of stretching and folding back. And it is precisely this stretching and folding that many of the mappings that are discussed in chaos theory (e.g. the cat map, the baker's transformation or the horseshoe) are supposed to capture. The horseshoe mapping, for instance, is best interpreted as a schematic picture of precisely this mechanism of stretching and folding we find in a continuous phase space. Obviously, these mappings are much simpler than the full-fledged dynamics and their mechanism is not quite the same as in the continuous case. But they capture the basic structure (that is why we refer to these maps as ‘analogue pictures’) and thereby help to elucidate what happens in the more intricate case of a continuous flow. In particular, the baker's map shows that the system involves a high degree of dynamical randomness.

## 4.5 Mappings as tools for analysing physical problems

So far we have used mappings as tools to understand the properties of a continuous phase flow, with which the ‘real physics’ was associated. But maps can also be used in a more direct way to study various physical problems. That is, there are physical problems whose formalisation immediately results in a mapping. In these cases the mapping emerges directly from the physical phenomenon under investigation and no ‘detour’ to a continuous formulation is needed. A case in point is the so-called Fermi accelerator (see Lichtenberg & Lieberman, 1991, pp. 57-9 and pp. 215-30). The system was proposed by Enrico Fermi in 1949 to model the acceleration of cosmic rays by momentum transfer from magnetic field structures. The model consists of a ball bouncing between a fixed and a sinusoidally oscillating wall. Depending on the concrete assumptions about the behaviour of the components of the model one obtains various mappings, including well-known mappings such as the twist map and the Chirikov-Taylor map (also referred to as the ‘standard map’). As a consequence, the whole machinery of ergodic theory can be directly used to discuss such physical problems, and in particular the degrees of randomness that they involve.

In sum, we think that the three uses discussed above – i.e. the use of mappings in surfaces of sections, as analogue pictures and as ‘direct’ models of physical phenomena – forcefully counter the objection that ergodic theory is irrelevant to physics in general and to characterising dynamical randomness in physical processes in particular.

## 5. Summary and concluding remarks

### 5.1. Summary of the main arguments

EH is often regarded as relevant for explicating the nature of randomness in deterministic dynamical systems. It is not clear, however, what notion of randomness this claim invokes. The formal definitions of EH do not make explicit appeal to randomness and the usual ways of presenting EH do not involve specifications of the notion (or notions) of randomness that is (or

are) supposed to underlie EH. After dismissing some seemingly obvious candidates we have suggested that EH can be interpreted as a hierarchy of randomness if degrees of randomness are explicated in terms of degrees of unpredictability, which in turn are explicated in terms of conditional degrees of beliefs. In order for these degrees of belief to be indicative of the system's dynamical properties, they have to be updated according to a system's dynamical law, i.e. according to ( $\mu$ -DynT). In technical terms, these degrees of belief equal the corresponding degrees of decay of correlation which are dictated by the system's dynamical properties.

Note that on our analysis, degrees of beliefs in a system's states at different times reflect randomness not because their (specific) values are fully determined by objective facts about the system. Any degree of belief that is updated according to ( $\mu$ -DynT) would reflect the system's dynamical properties. Thus, our use of degrees of beliefs to explicate randomness in deterministic systems is not subjected to some standard objections against the application of subjective probabilities in classical statistical physics.

## 5.2 Randomness and the Decay of Correlations

In EH, the different degrees of randomness, which correspond to the different levels of the hierarchy, are (generally) defined in asymptotic terms (the exception being Bernoulli processes). Each of these degrees of randomness expresses a different degree of unpredictability that corresponds to a different type of asymptotic decay of correlations between states of systems at different times. This might suggest that a similar pattern can be found in the rates of decay. That is, one might be tempted to think that EH can equally be characterised as a hierarchy of increasing rates of decay of correlations. On this reading, a K system, for instance, which exhibits exponential divergence of trajectories would be characterised by exponential rate of decay of correlations, while a SM systems would exhibit a polynomial rate of decay.

This, unfortunately, does not work.<sup>28</sup> Natural as it may seem, EH cannot be interpreted as a hierarchy of increasing rates of decay of correlations. It is a mathematical fact that there is no

---

<sup>28</sup> Thanks to Viviane Baladi and Dan Rudolph for pointing this out to us.

particular rate of decay associated with each level of EH. For instance, one can construct K systems in which the decay is as slow as one wishes it to be (e.g. polynomial). So the rate of decay is a feature of a particular system and not of a level of EH.

### 5.3 Chaos is a matter of degree

We mentioned at the beginning that it has become customary among physicists and philosophers alike to characterise the behaviour of chaotic systems in terms of certain levels of EH, typically K- and B-systems. More recently, it was suggested by Belot & Earman (1997, p. 155) that being merely strong mixing (SM) is a necessary condition and being a K-system is a sufficient condition for a system to be chaotic. This view faces two difficulties: the first is technical and the second is conceptual.

Let's begin with the technical difficulty. Intuitively, the claim that SM is a necessary condition for chaos appears appealing, because SM systems involve a significant degree of 'disorder,' which seems necessary for chaos. However, a closer look at chaotic systems soon reveals that, if no further provisos are made, this claim is false. Take, for a notable example, KAM-type systems. Clearly, these systems with their intermingled regions of irregular and regular motion exhibit a kind of behaviour that we would like to call 'chaotic' (for a discussion of such systems, see section 4.3). But according to the proposed criterion KAM-type systems are not chaotic since it is a straightforward consequence of the KAM theorem that these systems are not mixing. (In fact, as we have seen in Section 4.2, KAM systems are not even ergodic.) For this reason it is too restrictive to make SM a necessary condition for chaos.

One may try to rescue the idea that SM is a necessary condition for chaos by restricting the scope to more favourable cases (i.e. ones in which no disturbing KAM-curves are present). This can be achieved, for instance, by making the perturbation so strong that all invariant curves vanish, or by restricting the focus to the 'right' regions of the phase space (see Section 4.3). The claim is then: Given that a system is merely ergodic, SM is a necessary condition and KM is a sufficient condition for it to be chaotic. Or similarly, if the system is not ergodic, then focusing on the

regions where the system displays an ergodic behaviour, SM is a necessary condition and KM is a sufficient condition for chaos.

From a technical point of view, this is a workable suggestion. However, there is a problem with the basic mindset behind it. The search for necessary and sufficient conditions for chaos presupposes that there is a clear-cut divide between chaotic and non-chaotic cases. We believe that our analysis of EH challenges this view. EH provides a hierarchy of random behaviour and every attempt to draw a line somewhere to demarcate chaotic from non-chaotic systems is bound to be somewhat arbitrary. Ergodic systems are pretty regular, mixing systems are less regular and the higher positions in the hierarchy exhibit still more haphazard behaviour. But is there one particular point where the transition from 'non-chaos' to chaos takes place? Based on our argument that EH is a hierarchy of increasing degrees of randomness and degrees of randomness correspond to different degrees of unpredictability, we suggest that chaos may well be viewed as a matter of degree rather than an all-or-nothing affair. Bernoulli systems are very chaotic, K-systems are slightly less chaotic, SM-systems are still less chaotic, and ergodic systems are non-chaotic. This suggestion connects well with the idea that chaos is closely related to unpredictability. For if chaos in deterministic systems is closely related to unpredictability, there seems to be no good reason to claim that SM is a necessary condition for chaos and that K is a sufficient condition for it. First, SM already involves a significant degree of unpredictability. And, secondly, if there were a clear point where unpredictability starts to reflect chaotic behaviour, why should it be K- rather than SM- or B-systems?

The view that being a K-system is the mark of chaos and that any lower degree of randomness is not chaotic is frequently motivated by two ideas. The first is the idea that chaotic behaviour involves dynamical instability in the form of exponential divergence of nearby (possible) trajectories. Thus, since a system involves an exponential divergence of nearby trajectories only if it is a K-system, it is concluded that (merely) ergodic and mixing systems are not chaotic whereas K- and B-systems are. It is noteworthy, however, that SM is compatible with there being polynomial divergence of nearby trajectories and that such divergence sometimes exceeds exponential divergence in the short run. Thus, if chaos is to be closely associated with the rate of divergence of nearby trajectories, there seems to be no good reason to deny that SM- systems exhibit chaotic behaviour.

The second common motivation for the view that being a K-system is the mark of chaos is the idea that the shift from zero to positive KS-entropy marks the transition from a ‘regular’ to ‘chaotic’ behaviour. This may suggest that having positive KS-entropy is both necessary and sufficient condition for chaotic behaviour. Thus, since a system has positive KS-entropy if and only if it is a K-system, it is concluded that K-systems are chaotic whereas SM-systems are not.

Why is KS-entropy a mark of chaos? The answer to this question depends on the exact interpretation of KS-entropy. Three interpretations suggest themselves. One interpretation is based on Pesin’s theorem, which states that a dynamical system has positive KS-entropy if, and only if, it is dynamically unstable in the sense of having exponential divergence of nearby trajectories (see Lichtenberg & Liebermann 1992). As we have just argued above, this interpretation does not seem to support the suggestion that having a KS-entropy is a necessary condition for chaotic behaviour even if we accept the idea that chaos is closely related to the rate of divergence of nearby trajectories.

The second interpretation of KS-entropy is based on Brudno’s theorem, which connects it to algorithmic complexity (see again Lichtenberg & Liebermann 1992). Though elegant in its own right, this does not seem to get us very far in the context of our investigation because it does not connect to physical intuitions about randomness. Admittedly, ‘randomness’ is a vague notion that allows for different interpretations, but algorithmic complexity is just too far off the mark for the needs of physics. A physical notion of randomness must have something to say about unpredictability, which is not the case with algorithmic complexity.

Finally, in a recent paper Frigg (2004) relates the KS-entropy to a generalised version of Shannon’s information theoretic entropy. According to this approach, positive KS-entropy implies a certain degree of unpredictability. The idea is that this degree of unpredictability is sufficiently high to deserve the title chaotic. But this takes us back to the original question. Why draw the line here rather than somewhere else? Why should the line between non-chaotic and chaotic behaviour be drawn between unpredictability that implies zero KS-entropy and unpredictability that implies positive KS-entropy? Also, even if we suppose that positive KS-

entropy is a necessary condition for chaotic behaviour, the question is why should we draw the line at K-systems rather B-systems.

All these problems do not arise for our suggestion that chaotic behaviour of a system/process should be measured in terms of degrees of unpredictability, which are measured by degrees of beliefs that are constrained by the relevant dynamical laws and can be characterised precisely by EH. On this suggestion, no sharp line between chaotic and non-chaotic behaviour is called for: Chaos is conceived as a matter of degree.

## **Acknowledgements**

For helpful discussions we would like to thank the audiences at the conference ‘New Directions in the Foundation of Physics’ (Washington/DC, May 2001), the Quantum Club (Utrecht, March 2003), and the members of the Foundations of Physics Group in Barcelona (March 2003). For helpful discussions and/or communications we would like to thank Viviane Baladi, David Lavis, Janneke van Lith, Donald Ornstein, Dan Rudolph, and two anonymous referees. We also would like to thank the Centre for Philosophy of Social and Natural Science at LSE for providing us with a stimulating environment in which much of the research for this paper has been done. For financial support, JB would like to thank the Faculty of Arts and Sciences or the University of Maryland Baltimore County.

## Bibliography

Alekseev, V. M. and M. V. Yakobson (1981). Symbolic dynamics and hyperbolic dynamical systems. *Physics Reports*, 75, 287-325.

Argyris, J., G. Faust, and M. Haase (1994). *An exploration of chaos*. Amsterdam: Elsevier.

Arnol'd, V. I. (1963). Proof of a theorem of A. N. Kolmogorov on the preservation of conditionally periodic motions under a small perturbation of the Hamiltonian. (Russian) *Uspekhi Matematicheskikh Nauk*, 18 no. 5 (113), 13-40.

– and A. Avez, (1968). *Ergodic problems of classical mechanics*. New York: New York and Amsterdam: W. A. Benjamin.

Belot, G. and J. Earman (1997). Chaos out of order: Quantum mechanics, the correspondence principle and chaos. *Studies in the History and Philosophy of Modern Physics*, 28, 147-82.

Boltzmann, L. (1871). Einige allgemeine Sätze über Wärmegleichgewicht. (German) *Wiener Berichte*, 63, 670-711.

Brudno, A. A. (1978). The complexity of the trajectory of a dynamical system. *Russian Mathematical Surveys*, 33, 197-98.

Chernov, N. I. and C. Haskell, (1996). Nonuniformly hyperbolic K-systems are Bernoulli. *Ergodic Theory and Dynamical Systems*, 16, 19-44.

Cornfeld, I. P., S. V. Fomin, and Y. G. Sinai, (1982). *Ergodic theory*. Berlin and New York: Springer.



- Frigg, R. (2004). In what sense is the Kolmogorov-Sinai entropy a measure for random behaviour? Bridging the gap between dynamical systems theory and information theory. *British Journal for the Philosophy of Science*, 55, 2004, 411-434.
- Goldstein, H. (1980). *Classical mechanics*. Reading/Ma: Addison and Wesley.
- Kelley, J. L. (1975). *General topology*. Graduate Texts in Mathematics Vol. 27. New York: Springer-Verlag.
- Kim, J. (1973). Causation, nomic subsumption and the concept of event. *Journal of Philosophy*, 70, 217-236.
- Kolmogorov, A. N. (1954). On conservation of conditionally periodic motions for a small change in Hamilton's function. (Russian) *Doklady Akademii Nauk SSSR (N.S.)* 98, 527-530.
- Krylov, N.S. (1979). The processes of relaxation of statistical systems and the criterion of mechanical instability. Thesis 1942. In *Works on the Foundations of Statistical Physics*, translated by A. B. Migdal, Y. G. Sinai and Y. L. Zeeman. Princeton: Princeton University Press, 193-238.
- Lichtenberg, A. J. and M. A. Liebermann (1992). *Regular and chaotic dynamics*. 2<sup>nd</sup> ed., Berlin and New York: Springer.
- Mañé, R. (1983). *Ergodic theory and differentiable dynamics*. Berlin and New York: Springer.
- Markus, L. and K. R. Meyer (1974). Generic Hamiltonian dynamical systems are neither integrable nor ergodic. *Memoirs of the American Mathematical Society*, Providence, Rhode Island.
- Moser, J. (1962). On invariant curves of area-preserving mappings of an annulus. *Nachrichten der Akademie der Wissenschaften in Göttingen: II. Mathematisch-Physikalische Klasse*, 1-20.

– (1973). *Stable and random motions in dynamical systems*. New Haven: Yale University Press.

Nadkarni, M. G. (1998). *Basic ergodic theory*. Basel: Birkhäuser.

Ornstein, D. S. (1974). *Ergodic theory, randomness, and dynamical systems*. New Haven: Yale University Press.

– (1989). Ergodic theory, randomness, and ‘chaos’. *Science*, 243, 182-187. 182-187.

– and B. Weiss (1991). Statistical properties of chaotic systems. *Bulletin of the American Mathematical Society* (New Series), 24, 11-115.

Ott, E. (1993). *Chaos in dynamical systems*. Cambridge: Cambridge University Press.

Oxtoby, J. C. (1971). *Measure and category. A survey of the analogies between topological and measure spaces*. Graduate Texts in Mathematics Vol. 2. New York and Berlin: Springer.

– and S. M. Ulam (1941). Measure-preserving homeomorphisms and metrical transitivity. *Annals of Mathematics* (Second Series), 42, 874-920.

Parry, W. (1981). *Topics in ergodic theory*. Cambridge: Cambridge University Press.

Petersen, K. (1983). *Ergodic theory*. Cambridge: Cambridge University Press.

Reichl, L. E. (1992). *The transition to chaos in conservative systems: Quantum manifestations*. New York and Berlin: Springer.

Shields, P. (1973). *The theory of Bernoulli shifts*. Chicago: Chicago University Press.

- Sinai, Y. G. (1970). Dynamical systems with elastic reflections. Ergodic properties of dispersing billiards', (Russian) *Uspekhi Matematicheskikh Nauk*, 25 no. 2 (152), 141-192.
- (1963). On the foundations of the ergodic hypothesis for a dynamical system of statistical mechanics. *Doklady Akademii Nauk SSSR*, 153, 1261-1264 (Russian); translated as *Soviet Mathematics Doklady*, 4, 1818-1822.
- (ed.) (1980). *Dynamical systems II. Ergodic theory with applications to dynamical systems and statistical mechanics*. Berlin and Heidelberg: Springer.
- Simányi, N. (1992). The K-property of  $N$  billiard balls. *Inventiones Mathematicae* 108, 521-548.
- (1999). Ergodicity of hard spheres in a box. *Ergodic Theory and Dynamical Systems* 19, 741-766.
- (2000). Hard ball systems and semi-dispersive billiards: Hyperbolicity and ergodicity. Hard ball systems and the Lorentz gas', in: *Encyclopaedia of Mathematical Sciences* 101 (pp. 51-88). Berlin: Springer.
- and D. Szász (1999). Hard ball Systems are completely hyperbolic. *Annals of Mathematics* (Second Series), 149, 35-96.
- Szász, D. (1996). Boltzmann's ergodic hypothesis, a conjecture for centuries? *Studia Scientiarum Mathematicarum Hungarica*, 31, 299-322.
- (ed.) (2000). Hard Ball Systems and the Lorentz gas. *Encyclopaedia of Mathematical Sciences, 101. Mathematical Physics, II*. Berlin: Springer.
- Tabor, M. (1989). *Chaos and integrability in nonlinear dynamics. An introduction*. New York: John Wiley & Sons.
- Walters, P. (1982). *An introduction to ergodic theory*. New York: Springer.