Philosophical aspects of chaos: definitions in mathematics, unpredictability, and the observational equivalence of deterministic and indeterministic descriptions.

DISSERTATION
Submitted for the degree of Doctor of Philosophy

CHARLOTTE SOPHIE WERNDL
St John's College, University of Cambridge

Cambridge, September 2009

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration. No part of this dissertation has been submitted for any other qualification.

This dissertation does not exceed the word limit laid down by the Faculty of Philosophy.

Summary of the dissertation, Charlotte Sophie Werndl.

Philosophical aspects of chaos: definitions in mathematics, unpredictability, and the observational equivalence of deterministic and indeterministic descriptions.

This dissertation is about some of the most important philosophical aspects of chaos research, a famous recent mathematical area of research about deterministic yet unpredictable and irregular, or even random behaviour. It consists of three parts.

First, as a basis for the dissertation, I examine notions of unpredictability in ergodic theory, and I ask what they tell us about the justification and formulation of mathematical definitions. The main account of the actual practice of justifying mathematical definitions is Lakatos's account on proof-generated definitions. By investigating notions of unpredictability in ergodic theory, I present two previously unidentified but common ways of justifying definitions. Furthermore, I criticise Lakatos's account as being limited: it does not acknowledge the interrelationships between the different kinds of justification, and it ignores the fact that various kinds of justification—not only proof-generation—are important.

Second, unpredictability is a central theme in chaos research, and it is widely claimed that chaotic systems exhibit a kind of unpredictability which is specific to chaos. However, I argue that the existing answers to the question 'What is the unpredictability specific to chaos?' are wrong. I then go on to propose a novel answer, viz. the unpredictability specific to chaos is that for predicting any event all sufficiently past events are approximately probabilistically irrelevant.

Third, given that chaotic systems are strongly unpredictable, one is led to ask: are deterministic and indeterministic descriptions observationally equivalent, i.e., do they give the same predictions? I treat this question for measure-theoretic deterministic systems and stochastic processes, both of which are ubiquitous in science. I discuss and formalise the notion of observational equivalence. By proving results in ergodic theory, I first show that for many measure-preserving deterministic descriptions there is an observation-

ally equivalent indeterministic description, and that for all indeterministic descriptions there is an observationally equivalent deterministic description. I go on to show that strongly chaotic systems are even observationally equivalent to some of the most random stochastic processes encountered in science. For instance, strongly chaotic systems give the same predictions at every observation level as Markov processes or semi-Markov processes. All this illustrates that even kinds of deterministic and indeterministic descriptions which, intuitively, seem to give very different predictions are observationally equivalent. Finally, I criticise the claims in the previous philosophical literature on observational equivalence.

Contents

Acknowledgements

First and foremost, I want to thank my supervisor Jeremy Butterfield. He really has been one of the best supervisors you can ever wish to have. I am indebted for his helpful comments, for his continued support, and, in particular, for getting up at 3am to read my work. I already miss our meetings! I am grateful to my shadow supervisor Peter Smith for fruitful suggestions and for his strategic advice and help. And I want to thank Roman Frigg for stimulating discussions and his encouragement.

For valuable comments I also want to thank Robert Bishop, Adam Caulton, Stephan Hartmann, Cymra Haskell, Franz Huber, Brendan Larvor, Hannes Leitgeb, Mary Leng, Thomas Müller, Donald Ornstein, Amy Radunskaya, Maximilian Thaler, Jos Uffink and Paul Weingartner.

I am grateful to St John's College Cambridge for financial support, which made it possible for me to study at the University of Cambridge.

Finally, I want to express my gratitude to my fiancé Franz for his continued support and for showing me what is really important in life. I also want to thank my sister Kristina and my parents for their understanding and their encouragement.

Chapter 1

Introduction

This dissertation is about some of the most important philosophical aspects of chaos as understood in the mathematical field of chaos research. A system is deterministic just in case the state of the system at one time determines the state of the system at all times. And, intuitively speaking, a chaotic system is deterministic yet still shows unpredictable and irregular, or even random behaviour. Examples of what is now called 'chaotic behaviour' were already discovered at the end of the nineteenth century. However, only from the 1960s onwards, catalysed by the development of electronic computers, chaotic behaviour was systematically investigated. An area of research called 'chaos research' developed, and chaotic behaviour was examined in several branches of mathematics and theoretical physics, such as in ergodic theory and topological dynamical systems theory. At the end of the twentieth century chaos research boomed, and important results continue to be produced. Because systems in Newtonian mechanics and statistical mechanics can show chaotic behaviour, chaos research has led to a renewed interest in these fields. Chaos research is now widely regarded as one of the most important scientific achievements of the second half of the twentieth century (cf. Aubin & Dahan-Dalmedico 2002).

In the sciences chaotic systems are employed to model many phenomena, from the movement of planets, the motion of billiard balls, the motion of gases, the spinning of waterwheels, turbulence, chemical reactions, weather dynamics, climate dynamics and population dynamics to the dynamics of the

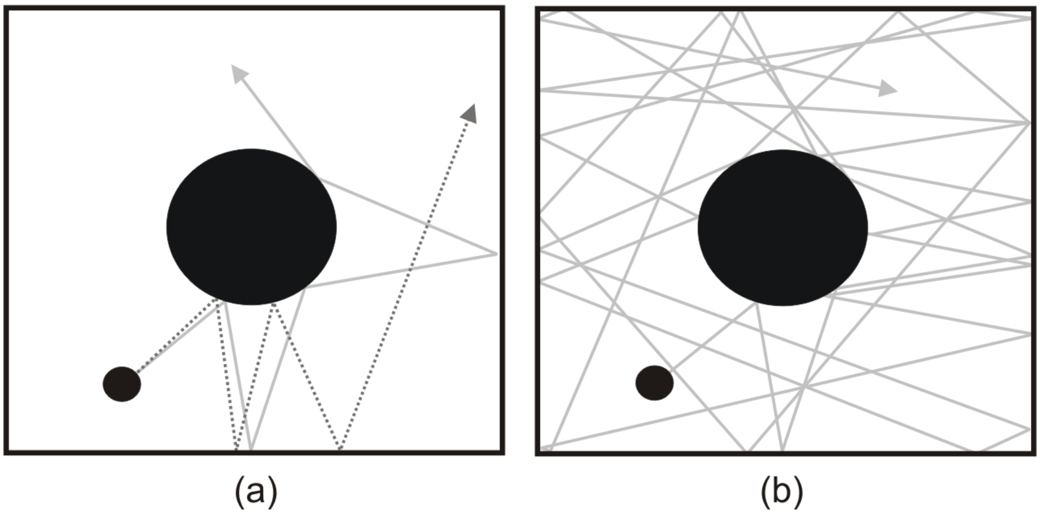

Figure 1.1: A billiard system with a convex obstacle

heartbeat (cf. Chernov & Markarian 2006; Hénon 1976; Kolář & Gumbs 1992; Laskar 1994; Lissauer 1999; Lorenz 1963; Lorenz 1964; May 1976; Ruelle & Takens 1971; Scott 1991; Skinner, Goldberger, Mayer-Kress & Ideker 1997; Sźasz 2000). In some contexts, such as for waterwheels, chaotic descriptions give relatively accurate predictions. Yet often, such as in population ecology and climate dynamics, the phenomena are so complicated that all scientists are able to derive are very simple chaotic models which help us to understand phenomena, but not so much to predict them.

Let me give an example of a chaotic system, namely a so-called billiard system with convex obstacles. This is a system where a ball moves with constant speed on a rectangular table where there are a finite number of convex obstacles with a smooth boundary. It is assumed that there is no friction and that there are perfectly elastic collisions (cf. Ornstein & Galavotti 1974). Figure 1.1 illustrates two key characteristics of chaotic behaviour with help of the example of a billiard system with one convex obstacle. First, Figure 1.1(a) shows that solutions which start close together eventually separate considerably, causing the motion to be unpredictable. Second, Figure 1.1(b) illustrates that the motion exhibits irregular behaviour in the sense that a solution eventually visits every region on the billiard table.

From a philosophical point of view chaotic behaviour is relevant for the following reasons. First, unpredictability is a crucial philosophical theme because we want to be aware of the limitations of our predictions, and chaos research contributes to our understanding of the kinds of unpredictability scientists can encounter. Second, also randomness is a central theme in philosophy, and chaos research has led to a better understanding of the possible randomness of deterministic behaviour. Third, the question of whether the world is deterministic or indeterministic has always been a topic of philosophical debate. And chaos research provides new insights about how deterministic behaviour compares to indeterministic behaviour. Fourth, it is one of the main questions in the philosophy of science and also in metaphysics how probabilities can be understood, and chaos research sheds light on the emergence of probabilities and has suggested new interpretations of probabilities. Finally, fifth, chaotic behaviour is also of interest to foundational problems in physics. In particular, there is the question whether chaos research can contribute to solving some of the vexing problems in statistical mechanics, such as how to derive an analogue of the second law of thermodynamics. Moreover, there is the hope that chaotic behaviour will help us to understand the emergence of classical physics from the quantum world. Much philosophical research will be needed to answer all the philosophical questions raised by chaos research. This dissertation will mainly contribute to our understanding of unpredictability and the topic of whether phenomena are deterministic or indeterministic, that is, the first and the third point, but will also touch the other points.

Chaos research is a part of dynamical systems theory, a general theory of deterministic behaviour. Dynamical systems theory broadly divides into two approaches: measure-theoretic dynamical systems theory, also called 'ergodic theory', and topological dynamical systems theory. This dissertation will be mainly about ergodic theory, although sometimes I will also invoke notions of topological dynamical systems theory. Ergodic theory not only describes chaotic behaviour but a wide class of deterministic behaviour, namely it deals with all those deterministic systems which are endowed with a measure. For instance, all deterministic systems in Newtonian mechanics and statistical mechanics can be described by ergodic theory.

I focus on ergodic theory for two reasons. First and foremost, in ergodic theory deterministic systems are endowed with a measure, which can be interpreted as a probability density. As a consequence, only the measure-theoretic perspective allows for a connection to probability theory, to information theory, to extant probabilistic accounts of randomness, and to the theory of stochastic processes and hence allows for a comparison of deterministic and indeterministic descriptions. The notions of probability, randomness and determinism are central philosophical themes. Thus I believe that ergodic theory is a richer and more interesting field for philosophical investigations than topological dynamical systems theory. Second, in the philosophical literature on chaos there has been little work on the measure-theoretic approach and there has been more work on the topological approach (e.g., Bishop 2003, Bishop 2008, Kellert 1993, Schurz 1996, Smith 1998, Stone 1989). One reason for this might be that ergodic theory is technically harder than topological dynamical systems theory. So even though ergodic theory seems to be a richer field for philosophical investigations, there has been less work on it.

The general outline of this dissertation on some of the most important philosophical aspects of chaos is as follows. First, I will examine mathematical notions of unpredictability in ergodic theory, and this examination will lead me to draw conclusions about the actual practice of how mathematical definitions are justified. On this basis, second, I will tackle the question of what kind of unpredictability is specific to chaotic systems. Finally, third, the fact that deterministic systems can be unpredictable and even random prompts the question of whether deterministic descriptions in ergodic theory and indeterministic descriptions can be observationally equivalent. I will reflect on this question, and, in particular, I will investigate what kind of results on observational equivalence hold for chaotic behaviour.

More specifically, in Chapter 2 of this dissertation I will introduce the basic notions on which the discussion of this dissertation will be based, most notably deterministic systems and stochastic processes. In Chapter 3 I will investigate historically how notions of unpredictability in ergodic were formed and how they have been justified in the literature. We will see that there is hardly any philosophical research on the actual practice of how mathematical definitions are justified apart from Lakatos (1976, 1978). On the basis of my case study of notions of unpredictability in ergodic theory, I will identify novel ways in which mathematical definitions can be justified, and I will criticise Lakatos's account of the justification of definitions. The discussion of notions of unpredictability in ergodic theory also serves the purpose of providing a background for the following chapters, where these definitions will be applied.

With this background I am ready to embark in Chapter 4 on the question of what is the unpredictability specific to chaos. From the beginning of chaos research, the unpredictability of chaotic systems has been of central interest, and so this question is one of the key questions about chaos and unpredictability. I will discuss the existing answers in the literature, and I will argue that they do not fit the bill. This prompts the search for an alternative answer, and I will propose a novel and general answer.

Given that deterministic systems can be unpredictable and even random, one can go a step further and ask: are deterministic and indeterministic descriptions observationally equivalent; that is, is it possible to describe some phenomena by deterministic as well as indeterministic descriptions? I will discuss this question in Chapter 5 with a special emphasis on observational equivalence involving chaotic behaviour. Once ergodic theory and the modern theory of stochastic processes had been developed, it was realised that by combining these two theories one can compare measure-theoretic deterministic descriptions with stochastic processes (the main indeterministic descriptions used in the sciences). Hence some mathematical results have been proven which shed light on the observational equivalence of deterministic and indeterministic descriptions. I will review these results which, surprisingly, have caught hardly any philosophical attention, and I will extend them by proving several new theorems in ergodic theory. Furthermore, I will philosophically assess all these results on observational equivalence. Then in Chapter 6 I will briefly summarise the findings of this dissertation, and I will conclude with an outlook for future research in this area.

Finally, let me point to two issues this dissertation will not be concerned with. I will be concerned with deterministic descriptions in dynamical systems theory, which can be regarded as a special kind of description of classical physics. Therefore, first, I will not be concerned with quantum theory. In particular, I will not treat the question of how the classical realm emerges from the quantum world. There is, of course, a vast literature on this controversial question. Let me just cite two philosophical works that focus on the connection with chaos theory, namely Belot & Earman (1997) and Landsman (2007, section 5-7). Second, deterministic descriptions in dynamical systems theory are mathematical models. I will explain in some detail in this dissertation that these mathematical models are often used in the sciences to model phenomena. But I will not tackle the questions of what constitutes a scientific model and whether scientific models accurately depict reality. Again, there is, of course, a vast literature on this: let me just cite a recent survey Frigg & Hartmann (2006).

Let me now introduce the notions needed for the discussion in this dissertation.

Chapter 2

Setting the stage

In this chapter in section 2.1 I will discuss the deterministic descriptions which will be needed throughout the dissertation, namely measure-theoretic deterministic systems and topological deterministic systems. After that, in section 2.2 I will introduce stochastic processes. Apart from the notion of a Bernoulli process, which will be important throughout the dissertation, the notions introduced in section 2.2 will only be needed in Chapter 5.

2.1 Deterministic systems

In this dissertation I will be mainly concerned with measure-theoretic deterministic systems but a few times also with topological deterministic systems; both deterministic descriptions are descriptions drawn from *dynamical systems theory*. Generally, deterministic systems as described in dynamical systems theory often model natural systems. Typically, a deterministic system is used to model a phenomenon that is only one among many phenomena which take place in the actual world. The assumption is made that the phenomenon under consideration is isolated from its environment. Of course, in the actual world this is not the case. But nevertheless the actual world is such that many phenomena can effectively be treated as isolated, and hence modeling phenomena with deterministic systems has proven to be very successful.

The two main elements of every deterministic system in dynamical systems theory are a set M of all possible states m, the *phase space* of the deterministic system, and a family of functions $T_t : M \to M$ mapping the phase space to itself, called the *evolution functions*. The parameter t is time, and $T_t(m)$ is the state of the system that started in initial state m after t units of time. If t is an integer (i.e., $t \in \mathbb{Z}$), the dynamics of the system is discrete and the system is said to be a *discrete* deterministic system. If t is a real number (i.e., $t \in \mathbb{R}$), the dynamics of the system is continuous and the system is called a *continuous* deterministic system. The family T_t defining the dynamics of the deterministic system must have the structure of a group where $T_{t_1+t_2}(m) = T_{t_2}(T_{t_1}(m))$ for all $m \in M$ and for all t_1, t_2 either in $\mathbb{Z}$ (discrete time) or $\mathbb{R}$ (continuous time). For discrete deterministic systems all T_t are generated as iterative applications of the single bijective map $T = T_1$, $T_1 : M \to M$ because $T_t(m) = T^t(m)$, and I refer to the $T^t(m)$ as *iterates* of m. The *discrete solution* through m, $m \in M$, is the sequence $s_m = (\ldots T^{-1}(m), m, T^1(m) \ldots)$. The *continuous solution* through m, $m \in M$, is the function $s_m : \mathbb{R} \to M$, $s_m(t) = T(t, m)$, where $T(t, m) = T_t(m)$. Continuous deterministic systems are also called *flows*, and they often arise as solutions to differential equations of motion, such as Newton's laws of motion.

It follows that all discrete and continuous deterministic systems are *deterministic* according to the canonical definition: any two solutions that agree at one instant of time agree at all future and past times (Butterfield 2005, Earman 1971, Earman 1986, Montague 1962).

I will mainly be concerned with measure-theoretic deterministic systems, but sometimes I will also need topological deterministic systems. So let me briefly introduce topological deterministic systems and then turn to measure-theoretic deterministic systems.

A topological deterministic system is a one that has a metric defined on M (cf. Petersen 1983, pp. 2–3). More specifically:

Definition 1 *A* discrete topological deterministic system *is a triple* (M, d, T) *where* M *(the phase space) is a set,* d *is a metric on* M*, and* $T : M \to M$ *(the evolution function) is a bijective and continuous function.*

Definition 2 *A* contiuous topological deterministic system *is a triple (M, d, T_t) where M (the phase space) is a set, d is a metric on M, and $T_t : M \to M$ (the evolution functions), $t \in \mathbb{R}$, is a family of continuous functions which have the structure of the above group.*

Assume that a continuous topological deterministic system (M, d, T_t) is given. Then (M, d, T_{t_0}) for $t_0 \in \mathbb{R}$ arbitrary, $t_0 \neq 0$, is a discrete topological deterministic system. The evolution function of this discrete system is $T_{t_0} : M \to M$ which means that you look at the continuous topological deterministic system (M, d, T_t) at points of time nt_0, $n \in \mathbb{Z}$. And I call these discrete deterministic systems (M, d, T_{t_0}) the *discrete versions of the continuous topological deterministic system* (M, d, T_t).[1]

It is generally assumed in the literature (e.g., Devaney 1986, p. 51) that topological deterministic systems provide a possible framework for characterising chaos. This makes intuitive sense because it is often imagined that in case of chaotic behaviour there is some way of measuring the distance between states in the phase space M and thus that there is a metric defined on M. Moreover, to the best of my knowledge, there is always a natural metric for paradigmatic chaotic systems. Often the phase space is simply a subset of $\mathbb{R}^n$, $n \geq 1$, and the metric is the standard Euclidean metric.

A measure-theoretic deterministic system is one whose phase space is endowed with a measure (cf. Cornfeld, Fomin & Sinai 1982, pp. 3–5). Before I can proceed, recall the following canonical definitions. A *measurable space* is a pair (M, Σ_M) where M is a set and Σ_M is a σ-algebra on M. A *measure space* is a triple (M, Σ_M, μ) where M is a set, Σ_M is a σ-algebra on M and μ is a measure on (M, Σ_M). For simplicity and to avoid some technical problems, I assume that any measure space is complete, i.e., every subset of a measurable set of measure zero is measurable. Furthermore, I assume that any measure space (M, Σ_M, μ) is a Lebesgue space;[2] this is standard in the

[1] Alternatively, continuous-time deterministic systems can be discretised by considering the successive hits of a solution on a suitable *Poincaré section*. All I say about discrete versions of continuous deterministic systems also holds true for discrete deterministic systems arising in this way (Berkovitz, Frigg & Kronz 2006, pp. 680–685; Smith 1998, pp. 92–93).

[2] A measure space (M, Σ_M, μ) is called a Lebesgue space if, and only if, there is a measure space (K, Σ_K, ν) where $K = [a, b) \subseteq \mathbb{R}$ is a (possibly nonempty) interval, there

context of measure-theoretic dynamical systems theory.[3]

Now I can define:

Definition 3 *A* discrete measure-theoretic deterministic system *is a quadruple* (M, Σ_M, μ, T) *where* (M, Σ_M, μ) *is a measure space with* $\mu(M) = 1$ *(*M *is the phase space) and* $T : M \to M$ *(the evolution function) is a bijective measurable function such that also* T^{-1} *is measurable.*

Definition 4 *A* continuous measure-theoretic deterministic system *is a quadruple* (M, Σ_M, μ, T_t) *where* (M, Σ_M, μ) *is a measure space with* $\mu(M) = 1$ *(*M *is the phase space) and* $T_t : M \to M$ *(the evolution functions),* $t \in \mathbb{R}$*, is a family of measurable functions which have the structure of the above group such that also* T_t^{-1} *is measurable for all* $t \in \mathbb{R}$.

I follow the common assumption that the measure of measure-theoretic deterministic systems is normalised: $\mu(M) = 1$. The motivation for this is that normalised measures are probability measures, making it possible to use probability calculus. Several interpretations suggest interpreting the measure as probability. This is not one the main topics of this dissertation, but I shall briefly explain at the end of this section some of the most popular interpretations which justify interpreting the measure as probability.

Given a discrete or continuous measure-theoretic deterministic system, when a property holds for all states $m \in \hat{M}$ with $\mu(M \setminus \hat{M}) = 0$, I will say that the property holds *for almost all points in* M or that the property holds *except for a set of measure zero.*

Given a continuous measure-theoretic deterministic system (M, Σ_M, μ, T_t), then $(M, \Sigma_M, \mu, T_{t_0})$ for $t_0 \in \mathbb{R}$ arbitrary, $t_0 \neq 0$, is a discrete measure-

is a countable set $\cup_{i \geq 1} m_i \in M$, there is a $\hat{K} \subseteq K$ with $\nu(\hat{K}) = 1$, there is a $\hat{M} \subseteq M$ with $\mu(\hat{M}) = 1$, and there is a bijective function $\phi : \hat{M} \setminus \cup_{i \geq 1} m_i \to \hat{K}$ such that (i) $\phi(A) \in \Sigma_K$ for all $A \in \Sigma_M$, $A \subseteq \hat{M} \setminus \cup_{i \geq 1} m_i$, $\phi^{-1}(B) \in \Sigma_M$ for all $B \in \Sigma_K, B \subseteq \hat{K}$; and (ii) $\nu(\phi(A)) = \mu(A)$ for all $A \in \Sigma_M$, $A \subseteq \hat{M} \setminus \cup_{i \geq 1} m_i$ (see Petersen 1983, pp. 16–17).

[3] These two assumptions are not restrictive for the following reasons: first, every measure space can easily be made complete. Second, every example of a measure space which is of interest in the applications of dynamical systems theory, and more generally in the development of the mathematical theory of measure-theoretic dynamical systems, is a Lebesgue space (see Petersen 1983, Rudolph 1990).

theoretic deterministic system. And I call these discrete deterministic systems $(M, \Sigma_M, \mu, T_{t_0})$ the *discrete versions of the continuous measure-theoretic deterministic system* (M, Σ_M, μ, T_t).

When observing a measure-theoretic deterministic system (M, Σ_M, μ, T) or (M, Σ_M, μ, T_t), one observes a value functionally dependent on, but maybe different from, the actual state. Hence observations can be modeled by an *observation function*, i.e., a measurable function $\Phi : M \to M_O$ from (M, Σ_M) to (M_O, Σ_{M_O}) where M_O is a set and (M_O, Σ_{M_O}) is a measurable space (cf. Ornstein & Weiss 1991, p. 16).

I will often be concerned with measure-preserving deterministic systems defined as follows (cf. Cornfeld et al. 1982, pp. 3–5):

Definition 5 *A* discrete measure-preserving deterministic system *is a discrete measure-theoretic deterministic system* (M, Σ_M, μ, T) *where the measure* μ *is invariant, i.e.,* $\mu(T(A)) = \mu(A)$ *for all* $A \in \Sigma_M$. *A* continuous measure-preserving deterministic system *is a continuous measure-theoretic deterministic system* (M, Σ_M, μ, T_t) *where the measure* μ *is invariant, i.e.,* $\mu(T_t(A)) = \mu(A)$ *for all* $A \in \Sigma_M$ *and all* $t \in \mathbb{R}$.

Measure-preserving deterministic systems are important models in physics but are also important in other sciences such as biology, geology etc. For first, all deterministic Hamiltonian systems and deterministic statistical-mechanical systems, and their discrete versions, are measure-preserving; and the relevant invariant measure is the Lebesgue-measure or a close cousin of it (Petersen 1983, pp. 5–6). A measure-preserving deterministic system is called *volume-preserving* if, and only if, the Lebesgue measure or a normalised Lebesgue measure is the invariant measure. A measure-preserving deterministic system which fails to be volume-preserving is called *dissipative*. Dissipative systems can also often be modeled as measure-preserving deterministic systems. More precisely, if $(M, \Sigma_M, \lambda, T)$ or $(M, \Sigma_M, \lambda, T_t)$ is dissipative (where λ is the Lebesgue measure), then often there exists a measure $\mu \neq \lambda$ such that (M, Σ_M, μ, T) or (M, Σ_M, μ, T_t) is measure-preserving. The Lorenz system is a case in point (see Example 3 which will be introduced later in this section) (Luzzatto, Melbourne & Paccaut 2005). Generally, the long-term behaviour of a large class of deterministic systems can be modeled by measure-preserving deterministic systems (Eckmann & Ruelle 1985), and the potential scope of measure-preserving deterministic systems is quite wide: although some evolution functions cannot be modeled by invariant measures, for very wide classes of evolution functions invariant measures have been proven to exist. For instance, if T is a continuous function on a compact metric space, then there exists at least one invariant measure (Mañé 1987, p. 52).[4]

It is generally agreed in the literature that measure-preserving deterministic systems provide a possible framework for characterising chaos (e.g., Eckmann & Ruelle 1985). As already pointed out, for volume-preserving deterministic systems the relevant invariant measure is the Lebesgue measure or a normalized Lebesgue measure. For dissipative deterministic systems, to the best of my knowledge, all systems that have ever been identified as chaotic have, or are believed to have, a relevant invariant measure—in the light of the following considerations.

Many chaotic systems have attractors. For a discrete topological deterministic system (M, d, T) the set $\Lambda \subset M$ is an *attractor* if, and only if, (i) $T(\Lambda) = \Lambda$; (ii) there is a neighbourhood $U \supset \Lambda$, called a 'basin of attraction', such that all solutions are attracted by Λ, i.e., for all y in U $\lim_{t\to\infty} \inf\{d(T^t(y), x) \,|\, x \in \Lambda\} = 0$; and (iii) no proper subset of Λ satisfies (i) and (ii). For a continuous topological deterministic system (M, d, T_t) the set $\Lambda \subset M$ is an *attractor* if, and only if, (i) $T_t(\Lambda) = \Lambda$ for all $t \in \mathbb{R}$; (ii) there is a neighbourhood $U \supset \Lambda$, called a 'basin of attraction', such that for all y in U $\lim_{t\to\infty} \inf\{d(T_t(y), x) \,|\, x \in \Lambda\} = 0$; and (iii) no proper subset of Λ satisfies (i) and (ii). Liouville's theorem implies that only dissipative systems can have attractors (Schuster & Just 2005, p. 162).[5] As I will show in section 4.3, for chaotic systems the evolution of any bundle of initial con-

[4] Topological deterministic systems and measure-theoretic deterministic systems are usually related in the following way: the σ-algebra Σ_M of a measure-theoretic deterministic system is or at least includes the Borel σ-algebra of the metric space (M, d) of the topological deterministic system. The Borel σ-algebra of (M, d) is the σ-algebra generated by all open sets of M (cf. Mañé 1987, pp. 2–3). Intuitively, it is the σ-algebra which arises from the metric space (M, d).

[5] Some other definitions of 'attractor' allow that volume-preserving deterministic systems can have attractors; yet these definitions are not standard in our context.

ditions eventually enters every region of phase space. This is impossible for the motion approaching an attractor Λ since the attracted solutions never return arbitrarily close to where they originated. Hence chaotic behaviour can only occur on Λ. The chaotic motion is described by a deterministic system with phase space Λ, and the invariant measure is only defined on Λ. Generally, an attractor on which the motion is chaotic is called a '*strange attractor*'.

Of course, in practice one is often concerned with solutions approaching a strange attractor. Yet after a sufficiently long duration either the solutions enter the attractor or come arbitrarily near to the attractor. In the latter case, since the dynamics is typically continuous, when the solutions are sufficiently near to the attractor, they essentially behave like the solutions on the attractor. And in applications such solutions which are sufficiently near to a strange attractor are considered to be chaotic for practical purposes. In particular, in the latter case, the unpredictability or randomness of solutions very near to the attractor is practically indistinguishable from the unpredictability or randomness on the attractor. Consequently, for characterising the unpredictability or randomness of motion dominated by strange attractors, it is widely acknowledged that *it suffices to consider the dynamics on attractors*, where relevant invariant measures can be defined (cf. Eckmann & Ruelle 1985).

The following examples of a discrete measure-preserving deterministic system and the following two examples of a continuous measure-preserving deterministic system will accompany us throughout the dissertation. They are all also paradigmatic examples of chaotic systems.

Example 1: The baker's system.

On the set $M = [0,1] \times [0,1] \setminus D$ where $D = \{(x,y) \in [0,1] \times [0,1] \mid x = j/2^n$ or $y = j/2^n$, $n \in \mathbb{N}$, $0 \leq j \leq 2^n\}$ consider

$$T(x,y) = (2x, \frac{y}{2}) \text{ if } 0 \leq x < \frac{1}{2};\ (2x-1, \frac{y+1}{2}) \text{ if } \frac{1}{2} \leq x \leq 1. \tag{2.1}$$

I exclude the set D from $[0,1] \times [0,1]$ in order to be able to define a bijective function T. Figure 1 illustrates that the baker's system first stretches the set M to twice its length and half its width; then it cuts the rectangle obtained

Figure 2.1: The baker's system on $0 \leq y \leq 1/2$

in half and places the right half on top of the left. For the Lebesgue σ-algebra Σ_M on M and the Lebesgue measure μ one obtains the measure-preserving deterministic system (M, Σ_M, μ, T). This system also has physical meaning. It describes a particle which moves in a part of three-dimensional space which contains M. It starts out in initial position (x, y) in M. The particle moves with constant speed in three-dimensional space. There it bounces on several mirrors, causing it to return to M at $T(x, y)$ (cf. Pitowsky 1995, p. 166).

Example 2: A billiard system with convex obstacles.
Our first example of a continuous measure-preserving deterministic system is a billiard system with convex obstacles as discussed in the Introduction (Chapter 1, see Figure 1.1). This is a system where a ball moves with constant speed on a rectangular table with a finite number of convex obstacles. It is assumed that there is no friction and that there are perfectly elastic collisions. Here M is the set of all possible positions and directions of the ball, Σ_M is the Lebesgue σ-algebra on M, μ is the Lebesgue measure, and $T_t(m)$, where $m = (p, q)$, gives the position and the direction after t time units of the ball that starts out in initial position q and initial direction p (for details, see Ornstein & Galavotti 1974).

Example 3: The Lorenz system.
Our second example of a continuous measure-preserving deterministic system

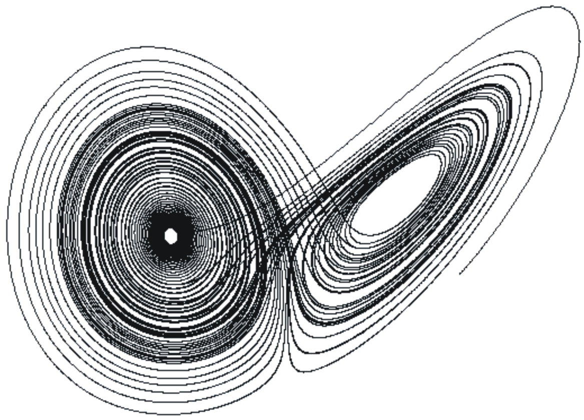

Figure 2.2: Numerical solution of the Lorenz equations for $\sigma = 10$, $r = 28$, $b = 8/3$

is the Lorenz system. Consider the Lorenz equations

$$\begin{aligned} \frac{dx(t)}{dt} &= \sigma(y(t) - x(t)) \\ \frac{dy(t)}{dt} &= rx(t) - y(t) - x(t)z(t) \\ \frac{dz(t)}{dt} &= x(t)y(t) - bz(t), \end{aligned} \tag{2.2}$$

for the parameter values $\sigma = 10$, $r = 28$ and $b = 8/3$. These are the parameters Lorenz (1963) considered when proposing the Lorenz system as a simplified model of weather dynamics. The Lorenz equations have also been used to model waterwheels, and it has been found that the Lorenz system gives relatively accurate predictions of waterwheels (cf. Hilborn 2000; Kolář & Gumbs 1992; Strogatz 1994). For these parameter values it is proven that there is a strange attractor of Lebesgue measure zero such that all solutions originating in the basin of attraction U, which is of positive Lebesgue measure, approach but never enter the attractor. Hence the dynamics is modeled by a measure-preserving deterministic system, the phase space of which is the attractor (Luzzatto et al. 2005). Figure 2.2 shows a numerical solution of these equations; one can vaguely discern the shape of the strange attractor, known as the Lorenz attractor, because the solution spirals toward it.

I have pointed out above that the measure of measure-theoretic deterministic systems is commonly interpreted as a *probability density*. This deep issue has been discussed in statistical mechanics but is not one of the main topics of this dissertation. But let me mention two interpretations that naturally suggest interpreting measures as probability. Namely, according to the *time-average interpretation*, the measure of a set A is the fraction of the proportion of time the deterministic system spends in A; and according to the *ensemble interpretation*, the measure of a set A at time t is the fraction of solutions starting from some given set of initial conditions that are in A at t (see Falconer 1990, p. 254; Lavis 2010).

Let me say more about the time-average interpretation. For a discrete measure-preserving deterministic system (M, Σ_M, μ, T) the *long-run time-average* of a solution starting at m relative to A, $m \in M$, $A \in \Sigma_M$, is:

$$L_A(m) = \lim_{t\to\infty} \frac{1}{t} \sum_{i=0}^{t-1} \chi_A(T^i(m)), \tag{2.3}$$

where $\chi_A(m)$ is the characteristic function of A.[6] For a continuous measure-preserving deterministic system (M, Σ_M, μ, T_t) the *long-run time-average* of a solution starting at m relative to A, $m \in M$, $A \in \Sigma_M$, is:

$$L_A(m) = \lim_{t\to\infty} \frac{1}{t} \int_0^t \chi_A(T_\tau(m)) d\tau, \tag{2.4}$$

where $\chi_A(m)$ is the characteristic function of A and the measure on the time axis $\tau \in \mathbb{R}_0^+$ is the Lebesgue measure. For discrete and continuous time it follows from Birkhoff's (1931) so called pointwise ergodic theorem that $L_A(m)$ exists for almost all states $m \in M$.

Now from an observational viewpoint it is natural to demand that the long-run time-averages of almost all solutions (relative to the Lebesgue measure) of a deterministic system approximate the measure of the system. Such measures are called *'physical measures'*. And, clearly, physical measures can be interpreted as probability densities in terms of the time average interpretation of probability. Let us look at physical measures in more detail. We need to distinguish two methods by which they can be specified.

[6] That is, $\chi_A(m) = 1$ for $m \in A$ and $\chi_A(m) = 0$ for $m \in M \setminus A$.

For discrete measure-preserving deterministic systems (M, Σ_M, μ, T) with $\lambda(M) > 0$ or continuous measure-preserving deterministic systems (M, Σ_M, μ, T_t) with $\lambda(M) > 0$, where λ is the Lebesgue measure, the following method identifies physical measures. (M1): (i) Take any $A \in \Sigma_M$. (ii) Take an initial condition $m \in M$. (iii) Consider $L_A(m)$, the long-run time-average of a solution starting at m relative to A. (iv) Consider $G_A = \{m \in M \mid L_A(m)$ exists and $L_A(m) = \mu(A)\}$. Then μ is a physical measure if, and only if, for any $A \in \Sigma_M$, Lebesgue-almost all initial conditions approximate the measure of A, i.e., $\lambda(G_A) = \lambda(M)$. If such a measure exists, it is unique (cf. Eckmann & Ruelle 1985, Young 2002).

What are physical measures for attractors (see the definition on p. 20)? I will be concerned with two kinds of attractors: first, the case where all solutions eventually enter an attractor Λ with $\lambda(\Lambda) > 0$. Clearly, here method (M1) can be applied directly, i.e., for $M = \Lambda$. Second, it can be that the solutions approach but never enter an attractor Λ with $\lambda(\Lambda) = 0$ but $\lambda(U) > 0$, where U is the basin of attraction of Λ. Here the method has to be slightly modified. (M2): (i) Take any measurable region $A \subseteq \Lambda$. (ii) Take an initial condition $m \in U$. (iii) Consider $\bar{L}_A(m)$, the long-run time-average of the solution originating at m which is *close to* A. (iv) Consider $\bar{G}_A = \{m \in U \mid \bar{L}_A(m)$ exists and $\bar{L}_A(m) = \mu(A)\}$. Then μ is a physical measure if, and only if, for all $A \in \Sigma_M$ Lebesgue-almost all initial conditions in U approximate the measure of A, i.e., $\lambda(\bar{G}_A) = \lambda(U)$. If such a measure exists, it is unique (for more details, see Eckmann & Ruelle 1985, Young 2002).

To illustrate the time-average interpretation for chaotic systems, consider the baker's system (Example 1). Now choose an initial condition m in the phase space M and draw a histogram of the fraction of iterates of m (up to an iterate $T^t(m), t \geq 1$) which are in a particular part on M. Then, for Lebesgue-almost all initial conditions we chose in M, we obtain what is illustrated in Figure 2.3: as t goes to infinity and the histogram becomes finer, the histograms approximate the uniform measure on M, that is, the Lebesgue measure. Hence this measure is physical according to method (M1).

Also, recall Example 3 and Figure 2.2 of the Lorenz system. Recall that here there is a strange attractor of Lebesgue measure zero such that all solutions in the basin of attraction U (of the attractor), which is of positive

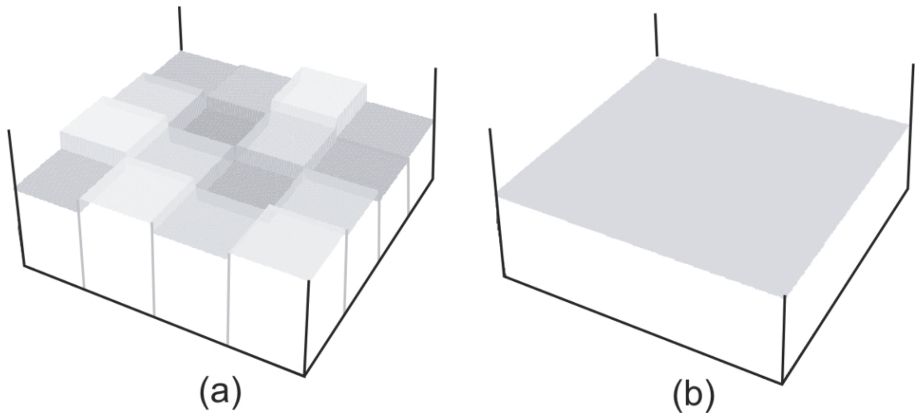

Figure 2.3: (a) histogram and (b) natural measure of the baker's system

Lebesgue measure, approach but never enter the attractor. According to the method (M2), the physical measure, which is the natural invariant measure on the attractor, is the unique measure with the following property: for Lebesgue-almost-all initial conditions in the basin of attraction the long-run time-average that the solution spends close to a set A on the attractor approximates the measure of A (cf. Luzzatto et al. 2005).

These two examples illustrate what is generally true, namely that for deterministic systems proven to be chaotic physical measure exist. For first, as I will show in section 4.3, chaotic systems are ergodic.

Definition 6 *A discrete measure-preserving deterministic system* (M, Σ_M, μ, T) *is* ergodic *if, and only if, for all* $A \in \Sigma_M$ *with* $\mu(A) > 0$*:*

$$\mu(\cup_{t\geq 0} T^{-t}(A)) = 1. \tag{2.5}$$

Now for ergodic volume-preserving deterministic systems method (M1) yields that the Lebesgue-measure is the physical measure (Eckmann & Ruelle 1985). Second, as I will explain in more detail in section 4.3, for dissipative systems proven to be chaotic, physical measures can be proven to exist. Moreover, for systems only conjectured to be chaotic, numerical evidence generally favours the existence of physical measures (Lyubich 2002; Young 1997; Young 2002).[7]

[7] Also for nonergodic deterministic systems the time-average interpretation can be used to justify interpreting the measure as probability (see Lavis 2010).

Finally, let me introduce the definition of a partition which I will need throughout the dissertation. Intuitively speaking, a partition of (M, Σ_M, μ) is a collection of non-empty, non-intersecting sets that cover M.

Definition 7 *$\alpha = \{\alpha_1, \ldots, \alpha_n\}$, $n \in \mathbb{N}$, is a* partition *of (M, Σ_M, μ), where (M, Σ_M, μ) is a measure space, if, and only if, $\alpha_i \in \Sigma_M$ and $\mu(\alpha_i) > 0$ for all i, $1 \leq i \leq n$, $\alpha_i \cap \alpha_j = \emptyset$ for all $i \neq j$, $1 \leq i, j \leq n$, and $M = \bigcup_{i=1}^{n} \alpha_i$.*

The α_i are called *atoms*. A partition is *nontrivial* if, and only if, it has more than one element. For a discrete measure-preserving deterministic system (M, Σ_M, μ, T), if α is a partition, then $T^t\alpha = \{T^t(\alpha_1), \ldots, T^t(\alpha_n)\}$, $t \in \mathbb{Z}$, is also a partition. Likewise, for a continuous measure-preserving deterministic system (M, Σ_M, μ, T_t): if α is a partition, then $T_t\alpha = \{T_t(\alpha_1), \ldots, T_t(\alpha_n)\}$, $t \in \mathbb{R}$, is also a partition. Given two partitions $\alpha = \{\alpha_1, \ldots, \alpha_n\}$ and $\beta = \{\beta_1, \ldots, \beta_m\}$ of (M, Σ_M, μ), the *least common refinement* $\alpha \vee \beta$ is defined as the partition $\{\alpha_i \cap \beta_j \,|\, i = 1, \ldots, n; j = 1, \ldots, m\}$ of (M, Σ_M, μ).

2.2 Stochastic processes

Let me now introduce stochastic processes. Apart from Bernoulli processes which will be important throughout the dissertation, the notions introduced in this section will only be needed to follow the discussion in Chapter 5.

A stochastic process is a process governed by probabilistic laws. Hence there is usually indeterminism in the time-evolution: if the process yields a specific outcome, there are different outcomes that might follow; and a probability distribution measures the likelihood of them. I call a sequence which describes a possible time-evolution of the stochastic process a realisation. Nearly all, but not all, the indeterministic descriptions in science are stochastic processes.[8]

Let me formally define stochastic processes. A *random variable* is a measurable function $Z : \Omega \to \bar{M}$ from a *probability space* $(\Omega, \Sigma_\Omega, \nu)$, that is, a

[8]For instance, Norton's dome (which satisfies Newton's laws) is indeterministic because the time evolution fails to be bijective. Nothing in Newtonian mechanics requires us to assign a probability measure on the possible states of this system. It is possible to assign a probability measure, but the question is whether it is natural (cf. Norton 2003, pp. 8–9).

measure space $(\Omega, \Sigma_\Omega, \nu)$ with $\nu(\Omega) = 1$, to a measurable space $(\bar{M}, \Sigma_{\bar{M}})$. The probability measure $P_Z(A) = P\{Z \in A\} = \nu(Z^{-1}(A))$ for all $A \in \Sigma_{\bar{M}}$ on $(\bar{M}, \Sigma_{\bar{M}})$ is called the *distribution* of Z. If A consists of one element, i.e., $A = \{a\}$, I often write $P\{Z = a\}$ instead of $P\{Z \in A\}$.

Definition 8 *A* discrete stochastic process $\{Z_t;\, t \in \mathbb{Z}\}$ *is a one-parameter family of random variables Z_t, $t \in \mathbb{Z}$, which are defined on the same probability space $(\Omega, \Sigma_\Omega, \nu)$ and take values in the same measurable space $(\bar{M}, \Sigma_{\bar{M}})$.*

Definition 9 *A* continuous stochastic process $\{Z_t;\, t \in \mathbb{R}\}$ *is a one-parameter family of random variables Z_t, $t \in \mathbb{R}$, which are defined on the same probability space $(\Omega, \Sigma_\Omega, \nu)$ and take values in the same measurable space $(\bar{M}, \Sigma_{\bar{M}})$ such that $Z(t, \omega) = Z_t(\omega)$ is jointly measurable in (t, ω).*

The set $\bar{M}$ is called the *outcome space* of the stochastic process. In the case of discrete time, a bi-infinite sequence $r_\omega = (\ldots Z_{-1}(\omega), Z_0(\omega), Z_1(\omega)\ldots)$, for $\omega \in \Omega$ arbitrary, is called a *realisation* of the stochastic process. For continuous time the function $r_\omega : \mathbb{R} \to \bar{M}$, $r_\omega(t) = Z(t, \omega)$, for $\omega \in \Omega$ arbitrary, is called a *realisation* (cf. Doob 1953, pp. 4–46). Intuitively, t represents time; so that each $\omega \in \Omega$ represents a possible history in all its details, and r_ω represents the description of that history by giving the 'score' at each t.

Assume a stochastic process $\{Z_t;\, t \in \mathbb{Z}$ or $\mathbb{R}\}$ with outcome space $\bar{M}$ is given. There can be situations when one observes a value which is dependent on, but maybe different from, the actual outcome of the stochastic process. Such situations can be modeled by an *observation function* Γ, i.e., a measurable function $\bar{M} \to \bar{M}_O$, where M_O is a set and $(\bar{M}_O, \Sigma_{\bar{M}_O})$ is a measurable space. Clearly, the resulting observed stochastic process is $\{\Gamma(Z_t);\, t \in \mathbb{Z}$ or $\mathbb{R}\}$.

I will often deal with stationary stochastic processes:

Definition 10 *A discrete stochastic process $\{Z_t;\, t \in \mathbb{Z}\}$ is* stationary *if, and only if, the distribution of the multi-dimensional random variable $(Z_{t_1+h}, \ldots, Z_{t_n+h})$ is the same as the one of $(Z_{t_1}, \ldots, Z_{t_n})$ for all $t_1, \ldots, t_n \in \mathbb{Z}$, $n \in \mathbb{N}$, and all $h \in \mathbb{Z}$. A continuous stochastic process $\{Z_t;\, t \in \mathbb{R}\}$ is* stationary *if, and only if, the distribution of the multi-dimensional random* *variable* $(Z_{t_1+h}, \dots, Z_{t_n+h})$ *is the same as the one of* $(Z_{t_1}, \dots, Z_{t_n})$ *for all* $t_1, \dots, t_n \in \mathbb{R}$, $n \in \mathbb{N}$, *and all* $h \in \mathbb{R}$ *(Doob 1953, p. 94).*

It is perhaps needless to stress the importance of stochastic processes, and stationary processes in particular: both are ubiquitous in science.

The following examples of discrete stochastic processes and of continuous stochastic processes will be important in this dissertation. Example 4 of a Bernoulli process will be important throughout the dissertation; the other examples will be important later in Chapter 5. Let me first introduce the examples of discrete stochastic processes.

Example 4: Bernoulli processes.
A Bernoulli process is a process where, intuitively, at each time point a (possibly biased) N-sided die is tossed where the probability for obtaining side s_k is p_k, $1 \leq k \leq N$, $N \in \mathbb{N}$, with $\sum_{k=1}^{N} p_k = 1$, and each toss is independent of all the other ones. Bernoulli processes are important in all sciences, from physics and biology to the social sciences.

The mathematical definition proceeds as follows. The random variables $X_1, \dots, X_n$, $n \in \mathbb{N}$, are probabilistically independent if, and only if, $P\{X_1 \in A_1, \dots, X_n \in A_n\} = P\{X_1 \in A_1\} \dots P\{X_n \in A_n\}$ for all $A_1, \dots, A_n \in \Sigma_{\bar{M}}$. The random variables $\{Z_t;\, t \in \mathbb{Z}\}$ are probabilistically independent if, and only if, any finite number of them is probabilistically independent.

Definition 11 *The discrete stochastic process* $\{Z_t;\, t \in \mathbb{Z}\}$ *is a* Bernoulli process *if, and only if, (i) its outcome space is a finite number of symbols* $\bar{M} = \{s_1, \dots, s_N\}, N \in \mathbb{N}$, *and* $\Sigma_{\bar{M}} = \mathbb{P}(\bar{M})$, *where* $\mathbb{P}(\bar{M})$ *is the power set of* $\bar{M}$*; (ii) there is a set of numbers* p_k, $0 \leq p_k \leq 1$, $1 \leq k \leq N$, *with* $\sum_{i=1}^{N} p_k = 1$ *such that* $P\{Z_t = s_k\} = p_k$ *for all* $t \in \mathbb{Z}$ *and all* k*; and (iii)* $\{Z_t;\, t \in \mathbb{Z}\}$ *are probabilistically independent.*

Clearly, a Bernoulli process is stationary.

In this definition the probability space Ω is not explicitly given. I now give a representation of Bernoulli processes where Ω is explicitly given. The idea is that Ω is the set of all possible realisations of the process. For a Bernoulli process with outcomes $\bar{M} = \{s_1, \dots, s_N\}$ which have probabilities $p_1, \dots, p_N$,

$N \in \mathbb{N}$, let Ω be the set of all bi-infinite sequences $\omega = (\ldots \omega_{-1}, \omega_0, \omega_1 \ldots)$ with $\omega_i \in \bar{M}$ corresponding to one of the possible outcomes of the i-th trial in a doubly infinite sequence of trials. Let Σ_Ω be the σ-algebra on Ω generated[9] by the semi-algebra of cylinder-sets

$$C_{i_1 \ldots i_n}^{A_1 \ldots A_n} = \{\omega \in \Omega \mid \omega_{i_1} \in A_1, .., \omega_{i_n} \in A_n,\ i_j \in \mathbb{Z},\ i_1 < \ldots < i_n,\ A_j \subseteq \bar{M}, 1 \leq j \leq n\}. \tag{2.7}$$

Since the outcomes are probabilistically independent, these sets have probability $\bar{\nu}(C_{i_1 \ldots i_n}^{A_1 \ldots A_n}) = P\{Z_{i_1} \in A_1\} \ldots P\{Z_{i_n} \in A_n\}$. Let ν be defined as the unique extension of $\bar{\nu}$ to a measure on Σ_Ω. Finally, define $Z_t(\omega) = \omega_t$ (the t-th coordinate of ω). Then $\{Z_t;\ t \in \mathbb{Z}\}$ is the Bernoulli process we started with in Definition 11.

Example 5: Markov processes.

Markov processes are discrete stochastic processes where the next outcome depends only on the previous outcome; and I will also assume that they have only finitely many possible outcomes and that the stochastic process is stationary. Markov processes are widely used to model phenomena in all sciences, from physics and biology to the social sciences.

Technically:

Definition 12 *A discrete stochastic process* $\{Z_t;\ t \in \mathbb{Z}\}$ *is a* Markov process *if, and only if, (i) its outcome space consists of a finite number of symbols* $\bar{M} = \{s_1, \ldots, s_N\}$, $N \in \mathbb{N}$, *and* $\Sigma_{\bar{M}} = \mathbb{P}(\bar{M})$*; (ii)* $P\{Z_{t+1} = s_j \mid Z_t, Z_{t-1} \ldots, Z_k\} = P\{Z_{t+1} = s_j \mid Z_t\}$ *for any* t*, any* $k \in \mathbb{Z}$, $k \leq t$*, and any* $s_j \in \bar{M}$*; and (iii)* $\{Z_t;\ t \in \mathbb{Z}\}$ *is stationary.*

Define $P^k(s_i, s_j) = P\{Z_{t+k} = s_i \mid Z_t = s_j\}$ for $k \in \mathbb{Z}$. A Markov process is *irreducible* exactly if it cannot be split into two processes because each outcome can be reached from all other outcomes; formally: for every $s_i, s_j \in \bar{M}$ there is a $k \in \mathbb{N}$ such that $P^k(s_i, s_j) > 0$. A Markov process is aperiodic

[9]The σ-algebra on M generated by a set $E \subseteq \mathbb{P}(M)$ is the smallest σ-algebra on M containing E, that is, the σ-algebra (cf. Ash 1972):

$$\bigcap_{\text{All } \sigma\text{-algebras } \Sigma \text{ on } M,\ E \subseteq \Sigma} \Sigma. \tag{2.6}$$

exactly if for every possible outcome there is no periodic pattern in which the process can revisit that outcome. To be precise: the *period* d_{s_i} of an outcome $s_i \in \bar{M},\ 1 \leq i \leq N$, is defined by $d_i = gcd\{k \geq 1 \,|\, P^k(s_i, s_i) > 0\}$ where '*gcd*' denotes the greatest common divisor. An outcome $s_i \in \bar{M}$ is *aperiodic* if, and only if, $d_i = 1$, and the Markov process is *aperiodic* if, and only if, all its possible outcomes are aperiodic.

Example 6: Multi-step Markov processes.

Multi-step Markov processes are Markov processes of order n, $n \in \mathbb{N}$, and are a generalisation of Markov processes. For Markov processes of order n the next outcome depends on the previous n outcomes but no other outcomes. I will also assume that a Markov process of order n has finitely many possible outcomes and is stationary; (hence Markov processes are Markov processes of order 1). Again, multi-step Markov processes are widely used in science to model phenomena.

Definition 13 *A discrete stochastic process* $\{Z_t;\, t \in \mathbb{Z}\}$ *is a* Markov process of order n*,* $n \in \mathbb{N}$*, if, and only if, (i) its outcome space consists of a finite number of symbols* $\bar{M} = \{s_1, \ldots, s_N\}$*,* $N \in \mathbb{N}$*, and* $\Sigma_{\bar{M}} = \mathbb{P}(\bar{M})$*; (ii)* $P\{Z_{t+1} = s_j \,|\, Z_t, Z_{t-1} \ldots, Z_k\} = P\{Z_{t+1} = s_j \,|\, Z_t, \ldots, Z_{t-n+1}\}$ *for any* t*, any* $k \in \mathbb{Z}$*,* $k \leq t - n + 1$*, and any* $s_j \in \bar{M}$*; and (iii)* $\{Z_t;\, t \in \mathbb{Z}\}$ *is stationary.*

That a Markov process of order n is irreducible is defined exactly as for Markov processes; also, that an outcome s_i, $1 \leq i \leq N$, of a Markov process of order n is aperiodic and that the Markov process of order n itself is aperiodic is defined exactly as for Markov processes.

Let me now introduce the examples of continuous-time stochastic processes.

Example 7: Semi-Markov processes.

Intuitively, a semi-Markov process is a continuous stochastic process with finitely many possible outcomes s_i; it takes the outcome s_i for a time $u(s_i)$, and which outcome follows s_i depends only on s_i and no other past outcomes.[10] Semi-Markov processes are widely used in the sciences to model phenomena, from physics and biology to the social sciences. In particular, they play an important role in queuing theory (cf. Janssen & Limnios, 1999).

A semi-Markov process is defined with help of a discrete stochastic process $\{(S_k, T_k);\ k \in \mathbb{Z}\}$. $\{S_k; k \in \mathbb{Z}\}$ describes the successive outcomes s_i visited by the semi-Markov process, and at time 0 the outcome of the semi-Markov process is S_0. T_0 is the time interval after which there is the first jump of the semi-Markov process after the time 0, T_{-1} is the time interval after which there is the last jump of the process before time 0, and all other T_k similarly describe the time-intervals between jumps of the stochastic process. Because at time 0 the semi-Markov process takes the outcome S_0 and the process takes the outcome S_0 for the time $u(S_0)$, it follows that $T_{-1} = u(S_0) - T_0$.

Technically, $\{Y_k;\ k \in \mathbb{Z}\} = \{(S_k, T_k),\ k \in \mathbb{Z}\}$ is a stochastic process which satisfies the following conditions: (i) $S_k \in S = \{s_1, \ldots, s_N\},\ N \in \mathbb{N}$; $T_k \in U = \{u_1, \ldots, u_{\bar{N}}\},\ \bar{N} \in \mathbb{N},\ \bar{N} \leq N$ for $k \neq 0, -1$, where $u_i \in \mathbb{R}^+$, $1 \leq i \leq \bar{N}$; $T_0 \in (0, u(S_0)]$, $T_{-1} \in [0, u(S_0))$, where $u : S \to U, s_i \to u(s_i)$, is a surjective measurable function; and hence $\bar{M} = S \times [0, \max_i u_i]$; (ii) $\Sigma_{\bar{M}} = \mathbb{P}(S) \times L([0, \max_i u_i])$, where $L([0, \max_i u_i])$ is the Lebesgue σ-algebra on $[0, \max_i u_i]$; (iii) $\{S_k;\ k \in \mathbb{Z}\}$ is a Markov process with outcome space S (as defined in Example 5); $p_{s_i} = P\{S_0 = s_i\} > 0$, for all i, $1 \leq i \leq N$; (iv) $T_k = u(S_k)$ for $k \geq 1$, $T_k = u(S_{k-1})$ for $k \leq -2$, and $T_{-1} = u(S_0) - T_0$; (v) for all i, $1 \leq i \leq N$, $P(T_0 \in A \,|\, S_0 = s_i)$ has a uniform density over $(0, u(s_i)]$, i.e., we have $P(T_0 \in A \,|\, S_0 = s_i) = \int_A 1/u(s_i) d\lambda$ for all $A \in L((0, u(s_i)])$, where $L((0, u(s_i)])$ is the Lebesgue σ-algebra on $(0, u(s_i)]$ and λ is the Lebesgue measure on $(0, u(s_i)]$.

Definition 14 *The continuous stochastic process $\{Z_t;\ t \in \mathbb{R}\}$ with outcome space S and $\Sigma_S = \mathbb{P}(S)$ constructed via a process $\{(S_k, T_k);\ k \in \mathbb{Z}\}$ as follows is called a* semi-Markov process*:*

$Z_t = S_0$ for $-T_{-1} \leq t < T_0$,

$Z_t = S_k$ for $T_0 + \ldots + T_{k-1} \leq t < T_0 + \ldots + T_k; k \geq 1$ and thus $t \geq T_0$,

$Z_t = S_{-k}$ for $-T_{-1} - \ldots - T_{-k-1} \leq t < -T_{-1} - \ldots - T_{-k}(\omega); k \geq 1$ and thus $t < -T_{-1}$,

[10] The term 'semi-Markov process' is not used unambiguously in the literature. Our use of this term follows Ornstein & Weiss (1991).

and for all i, $1 \leq i \leq N$,

$$P(Z_0 = s_i) = \frac{p_{s_i} u(s_i)}{p_{s_1} u(s_1) + \ldots + p_{s_N} u(s_N)}. \tag{2.8}$$

It can be proven that semi-Markov processes thus defined are stationary stochastic processes (Ornstein 1970*b*; Ornstein 1974, pp. 56–61). I will be concerned later with semi-Markov processes where the Markov process $\{S_k;\ k \in \mathbb{Z}\}$ is irreducible and aperiodic and where the elements of the set U are irrationally related (u_i and u_j are called *irrationally related* if, and only if, $\frac{u_i}{v_j}$ is not a rational number; and a set of elements $\{u_1, \ldots, u_{\bar{N}}\}$ is called *irrationally related* if, and only if, for all i, j, $i \neq j$, u_i and u_j are irrationally related). I will call those stochastic processes *irrationally related* semi-Markov processes.

Example 8: Multi-step semi-Markov processes.
Multi-step semi-Markov processes are semi-Markov processes of order n, $n \in \mathbb{N}$, and are a generalisation of semi-Markov processes. A semi-Markov process of order n is a continuous stochastic process with a finite number of possible outcomes s_i; it takes the outcome s_i for a time $u(s_i)$, and which outcome follows s_i depends only of the past n outcomes (hence semi-Markov processes are semi-Markov processes of order 1).[11] Again, multi-step semi-Markov processes are widely used to model phenomena in science (cf. Janssen & Limnios, 1999).

Definition 15 Semi-Markov processes of order n *are defined as semi-Markov processes except that for the discrete stochastic process* $\{(S_k, T_k);\ k \in \mathbb{Z}\}$ *condition (iii) is replaced by the following condition: (iii') $\{S_k;\ k \in \mathbb{Z}\}$ is a Markov process of order n with outcome space S (as defined in Example 6) and* $p_{s_i} = P\{S_0 = s_i\} > 0$, *for all* i, $1 \leq i \leq N$.

Again, it can be proven that multi-step semi-Markov processes are stationary stochastic processes (Park 1982). In Chapter 5 I will be concerned with multi-step semi-Markov processes where the multi-step Markov process

[11] The term 'multi-step semi-Markov process' is not used unambiguously in the literature, and I follow the usage of Ornstein & Weiss (1991).

$\{S_k;\ k \in \mathbb{Z}\}$ is irreducible and aperiodic and where the elements of U are irrationally related. I will call those stochastic processes *irrationally related* multi-step semi-Markov processes.

After setting the stage, we are now ready to turn to the first substantial chapter of this dissertation, where I will historically investigate how notions of unpredictability in ergodic theory were formed and how they are justified in the mathematical literature.

Chapter 3

Justifying definitions in mathematics—going beyond Lakatos

3.1 Introduction

Mathematical practice suggests that mathematical definitions are not arbitrary: for definitions to be worth studying there have to be good reasons. Moreover, definitions are often regarded as important mathematical knowledge (cf. Tappenden 2008*a* and 2008*b*). Reasoning and knowledge are classical philosophical issues; hence reflecting on the reasons given for definitions is philosophically relevant.

These considerations motivate the guiding question of this chapter: *in what ways are definitions in mathematics justified, and are these kinds of justification reasonable?* By a justification of a definition I mean a reason provided for the definition. I will concentrate on *explicit* definitions, which introduce a new expression by stipulating that it be semantically equivalent to the definiens consisting of already-known expressions. I will not deal with their complement, *implicit* definitions, which assign meaning to expressions by imposing constraints on how to use sentences (or other longer expressions) containing them (Brown 1999, p. 97).

Generally, attempting to justify definitions is reasonable: as we will see, if definitions were not justified, the mathematics involving these definitions would be much less meaningful to us than mathematics involving definitions which were justified. Thus given our limited resources, it is better to concentrate on definitions which we can justify.[1]

When a mathematician formulates a definition she or he has not known before, I speak of a formulation of the definition. The way a formulation of a definition is guided usually corresponds to the way the definition is justified when it is formulated. Thus all that will be said about the justification of definitions has a natural counterpart in terms of the guidance of the formulation of definitions. Since the guidance of the formulation of definitions derives from the justification, the latter is the main issue, and in what follows I will focus on the justification of definitions.[2]

In this chapter, in section 3.2, I will first discuss the state of the art of philosophical theorising about the actual mathematical practice of how definitions are justified in articles and books. There is hardly any philosophical discussion on this issue apart from Lakatos's ideas on proof-generated definitions, and hence I will concentrate on them. While Lakatos's ideas are important, this chapter aims to show how they are limited. My criticism of Lakatos will be based on a case study of notions of unpredictability in ergodic theory, which will be introduced in section 3.3. In section 3.4 I will discuss how notions of unpredictability in ergodic theory have been justified. And based on this, I will introduce three other ways in which definitions are commonly justified: natural-world justification, condition justification and redundancy justification; the latter two, to my knowledge, have not been discussed before. In section 3.5 I will clarify the interrelationships between the different kinds of justification, an issue which also has not been addressed before. In particular, I argue that in different arguments the same definition

[1]What this means for the ontology of mathematical definitions depends on the ontology adopted: platonists may hold that the entity defined by a definition is real regardless of whether we can justify the definition or not. Constructivists may hold that only those definitions that have been justified a constructed by us.

[2]Strictly speaking, the justification and the guidance of formulation are conceptually distinct. For instance, it could be that a definition which captures an important preformal idea was randomly formulated by a computer; then there was no way the formulation of the definition was guided, but there is a convincing initial justification.

can be justified in different ways. In section 3.6 I point out how Lakatos's ideas are limited: his ideas fail to show that often, and in particular for notions of unpredictability in ergodic theory, various kinds of justification are found and that various kinds of justification can be reasonable. Furthermore, they fail to acknowledge the interplay between the different kinds of justification. Finally, in section 3.7 I summarise the findings of this chapter.

The research of this chapter is in the spirit of 'phenomenological philosophy of mathematics' as recently characterised by Larvor (2001, pp. 214–215) and Leng (2002, pp. 3–5): it looks at mathematics 'from the inside' and on this basis asks philosophical questions.

3.2 Lakatos's proof-generated definitions

In the relatively recent literature Larvor (2001, p. 218) at least mentions the importance of researching the justification of mathematical definitions. Corfield (2003, chapter 9) discusses the related issue of what makes concepts fundamental but does not provide conceptual reflection on our question. Tappenden (2008*a*, 2008*b*) treats the related issues of naturalness of definitions and how to decide between different definitions. In our context Tappenden's (2008*a*) conclusion is relevant: namely that judgments about definitions mainly depend not on the rules of logic but on detailed knowledge about the mathematics involved. Furthermore, several philosophers have argued that mathematical definitions should capture a valuable preformal idea (cf. Brown 1999, p. 109).

Apart from this, the main philosopher who has written on our guiding question in the light of mathematical practice is Lakatos (1976, 1978). Lakatos develops an approach of informal mathematics, which includes an account of mathematical progress called *proofs and refutations*. Most importantly, Lakatos is also concerned with how definitions are justified. His key idea is the notion of a *proof-generated definition*. Here his main example are definitions of polyhedron which are justified because they are needed to make the proof of the Eulerian conjecture work: viz. that for every polyhedron the number of vertices minus the number of edges plus the number of faces equals two ($V - E + F = 2$).

What is a proof-generated definition? Unfortunately, Lakatos does not state exactly what he means by this. Clearly, mathematical definitions justified in any way are eventually involved in proofs. Therefore, the trivial idea that definitions are justified because they are involved in proofs cannot be what interested Lakatos.

To find out more, consider the Carathéodory definition of measurable sets, another proof-generated definition Lakatos discusses. The mathematician Halmos (1950, p. 44) remarks on this definition: "The greatest justification of this apparently complicated concept is, however, its possibly surprising but absolute complete success as a tool of proving the extension theorem". Lakatos (1976, p. 153) comments:

> as we learn from the second part [Halmos's remark above], this concept is a proof-generated concept in Carathéodory's theorem about the extension of measures [...]. So whether it is intuitive or not is not at all interesting: its rationale lies not in its intuitiveness, but in its proof-ancestor.

This quote and the rest of the discussion of proof-generated definitions suggests that a *proof-generated definition is a definition which is needed in order to prove a specific conjecture regarded as valuable* (Lakatos 1976, pp. 88–92, pp. 127–133, pp. 144–54; Lakatos 1978, pp. 95–97). This idea is also hinted at by Polya (1949, p. 686; and 1954, p. 148). The final theorems which involve proof-generated definitions often, but not always, result from a series of trials and revisions.

Lakatos (1976, pp. 33–50, p. 127) rightly argues that *lemma-incorporation* produces proof-generated definitions: assume that a conjecture, known not to hold for all objects of a domain, should be established. Then if conditions which are needed in order to prove the conjecture are identified, i.e., lemmas are incorporated, proof-generated definitions arise. For instance, consider the conjecture that the limit function of a convergent sequence of continuous functions is continuous. This conjecture can be proven if 'convergent' is understood as uniformly convergent but not if it is understood as the more obvious, weaker pointwise convergent; hence the definition of uniformly convergent is proof-generated (Lakatos 1976, pp. 144–146).

Lakatos (1976, pp. 90–92, p. 128, pp. 148–149, p. 153) thinks that for his examples of proof-generated definitions the justification was *reasonable* because the corresponding conjectures are valuable. Generally, if the conjecture is mathematically valuable, proof-generation is a reasonable kind of justification.[3] A proof-generated definition can be regarded as providing knowledge since it answers the question of which notion is needed to prove a specific conjecture.

Lakatos (1976, pp. 14–33, pp. 83–87) also discusses four other ways of justifying definitions. Imagine that counterexamples are presented to a conjecture of interest, and that the conjecture is defended by claiming that these are no 'real' counterexamples because a definition in the conjecture has been wrongly understood. Properly understood, it is argued, the definition excludes a class of objects which includes the alleged counterexamples, where the exclusions are made independent of any proof of the conjecture (and thus it is unknown whether the conjecture indeed holds true for the definition). Then the definition is justified via *monster-barring*. The second kind of justification is *exception-barring*. Here the definition is defended by excluding, with the extant definition, a class of objects which are, and which are regarded as, counterexamples to the conjecture; again, this is independent of any proof of the conjecture.[4] The third kind of justification is *monster-adjustment*. Here the definition is defended by reinterpreting, independent of any proof of the conjecture, the terms of the extant definition such that counterexamples to the conjecture are no longer counterexamples. The fourth and final kind of justification is *monster-including*. Here the definition is defended by extending the definition to include a new class of objects; this class of objects is defined using properties which are shared by examples for which the conjecture holds true; and again, this is independent of any proof of the conjecture.

Monster-barring, exception-barring and monster-adjustment are all ways

[3]For the proof-generated definitions discussed in Lakatos (1976) and in this chapter it is argued why the conjectures are valuable. Yet answering the question of what constitutes valuable conjectures at a general level would require further research.

[4]Contrary to exception-barring, in the case of monster-barring it is denied that the counterexamples are actual counterexamples. This is how monster-barring differs from exception-barring.

of dealing with counterexamples to conjectures. And I agree with Lakatos that for this purpose they are inferior to proof-generation because they do not take into account how the conjectures are proved; and therefore, it is even unclear whether the conjecture is true for the definition under consideration. Monster-including is a way of generalising conjectures. Yet again, since it neglects how conjectures are proved, I agree with Lakatos that for this purpose it is inferior to proof-generation. Furthermore, Lakatos thought that any of these kinds of justification were applied only because the better way of justifying definitions, namely with proof-generation, was not known (Lakatos 1976, pp. 14–42, pp. 136–140). Because of their inadequacies and since they play no role in our case study, I shall not say any more about these kinds of justification in this chapter.

Unfortunately, Lakatos (1976) never explicitly states how widely he thinks that his ideas on proof-generated definitions apply. He seems to think that mathematicians discovered the method of justifying definitions via proof-generation in the 1840s (Lakatos, 1976, p. 139). Apart from this, general claims such as

> Progress indeed replaces *naive classification* by [...] proof-generated [...]classification. [...] *Naive conjectures and naive concepts are superseded by improved conjectures (theorems) and concepts (proof-generated* [...] *concepts) growing out of the method of proofs and refutations* (Lakatos, 1976, pp. 91–92; see also p. 144, original emphasis).

suggest that mathematical definitions should be, and after mathematicians discovered the method of proof-generation, are generally proof-generated, and some have interpreted him as saying this (Brown 1999, pp. 110–111). However, as Larvor (1998) has pointed out, Lakatos stresses in his dissertation (Lakatos 1961), on which his (1976) book is based, that his account of informal mathematics does not apply to all of mathematics. What is clear is that Lakatos thought that there are many mathematical subjects with some proof-generated definitions and that there are many mathematical subjects with some definitions which should be proof-generated.[5] Maybe Lakatos

[5] Of course, the question remains what a 'mathematical subject' is; I will say more

also believed something stronger, and this would explain his strong claims such as in the above quote, namely that there are many subjects where proof-generation should be the *sole* important way in which definitions are justified; and that there are many subjects created after mathematicians discovered the method of proof-generation where proof-generation is the *sole* important way in which definitions are justified. In what follows, I will show in which ways Lakatos's ideas on justifying definitions are limited; and for this it will not matter much whether or not he endorsed the stronger claim.

Corfield (1997, pp. 111–115) argues that Lakatos did not think that his account of informal mathematics, which includes his ideas on justifying definitions, extends to established branches of mathematics of the twentieth century. Yet Corfield's claim is implausible. Lakatos (1976, p. 5, pp. 152–154) states that his ideas on informal mathematics apply to modern metamathematics and to Carathéodory's (1914) investigations on measurable sets. And substantial parts of established mathematics of the twentieth century are not any more formalised than that mathematics: e.g., ergodic theory, which will be relevant later. Thus Lakatos indeed thought that his ideas could apply to substantial parts of established branches of mathematics of the twentieth century. But I agree with Corfield's (1997) main point that Lakatos failed to see that his ideas are also relevant for highly formalised mathematics. For this reason, this chapter is not restricted to informal mathematics.

This discussion highlights that there is little work on the actual practice of how definitions are justified in articles and books. Furthermore, although Lakatos's account of proofs and refutations has been challenged (Corfield 1997, Leng 2002), his ideas on proof-generated definitions have hardly been criticised. My contribution on the guiding question and my criticism of Lakatos's ideas on justifying definitions will be based on a case study of notions of unpredictability in ergodic theory. Let me now introduce this case study.

---

about this later (see subsection 3.4.4).

3.3 Case study: notions of unpredictability in ergodic theory

My case study is on *notions of unpredictability in ergodic theory.* Ergodic theory originated from work in statistical mechanics, in particular Boltzmann's kinetic theory of gases. Some of Boltzmann's work relied on the assumption that the time-average of a function equals its space average, but no acceptable argument was provided for this (cf. Uffink 2007). Generally, the possible unpredictable motion of classical systems was a constant theme in statistical mechanics. Ergodic theory arose in the early 1930s when Birkhoff (1931) and von Neumann (1932*a*) proved the famous mean and pointwise ergodic theorems, respectively. Among other things, they found that ergodicity (cf. Definition 2.5) was the sought-after concept guaranteeing the equality of time and space averages for almost all states of the system. Motivated by these results, an investigation into the unpredictable behaviour of classical systems began. Of particular importance here was the study of unpredictability by a group of mathematicians around Kolmogorov in Russia. From the 1960s onwards, ergodic theory became prominent, and was further developed, as a mathematical framework for studying chaotic behaviour. Overall, ergodic theory had less impact on statistical mechanics than expected, partly because of the doubts, and the difficulty of proving, that the relevant systems are ergodic. But it developed into a discipline with its own internal problems and had, and continues to have, considerable impact on probability theory and chaos research (Aubin & Dahan-Dalmedico 2002; Dahan-Dalmedico 2004; Mackey 1974).

Why do notions of unpredictability in ergodic theory constitute a valuable case study? First, several of Lakatos's assertions, e.g., that mathematics is driven by counterexamples, have been criticised in the following way: while they may be correct for older mathematics, they do not hold true for twentieth century mathematics (Leng 2002, p. 10). As also Lakatos (1976, pp. 136–140) suggests, how definitions are justified may depend on when they were formulated because reasoning changes with the advancement of mathematics. To ensure that claims on the justification of definitions escape the criticism of not applying to twentieth century mathematics, I choose a branch of mathematics, viz. ergodic theory, which was created in the twentieth century. Second, concerning the justification of definitions, the picture for notions of unpredictability in ergodic theory appears different to that proposed by Lakatos, and this picture seems prevalent in mathematics.

As widely acknowledged, the main notions of unpredictability in ergodic theory are (cf. Berkovitz, Frigg & Kronz 2006; Sinai 2000, p. 21, pp. 41–46; Walters 1982, pp. 39–41, pp. 86–87, pp. 105–107):

weak mixing (three versions), strong mixing (two versions), Kolmogorov-mixing, Kolmogorov-system, *Bernoulli system (two versions)*, *Kolmogorov-Sinai entropy.*[6]

In the remaining sections of this chapter, I will present the insights on the justification of definitions which derive from this case study. I will discuss the way the discrete-time and continuous-time versions of the definitions of the above list which are italicised are justified as notions of unpredictability in the literature and whether they are reasonably justified. I will also examine the way these definitions have been initially justified.[7] A detailed investigation of them will suffice to illustrate these insights. Hence, for the remaining listed definitions, I will just state how they are justified. Let me now discuss the kinds of justification which occur in this case study. They illustrate that not only proof-generation is important.

[6]The definitions of weak mixing, strong mixing, being a Kolmogorov system and being a Bernoulli system are also sometimes referred to as the *ergodic hierarchy.*

[7]I will not investigate the use of these definitions elsewhere in mathematics. The main reason for such an investigation would be to understand how the justification of definitions varies in different contexts. Yet I think that one can also find out about this by considering only how definitions were initially justified and later justified as notions of unpredictability. Going further would required an enormous amount of work without considerable gain.

3.4 Kinds of justification of definitions

3.4.1 Natural-world justification

I claim, first, that *definitions in my case study are frequently justified because they capture a preformal idea regarded as valuable for describing or understanding the natural world.* Here I will speak of *natural-world-justified* definitions. Natural-world-justified definitions are a special case of the general idea discussed in the literature that mathematical definitions should capture a valuable preformal idea (cf. Brown 1999, p. 109).

If the preformal idea is valuable for describing or understanding the natural world, natural-world-justification is reasonable. It is important to realise that natural-world-justification does not mean that there is a 'best' definition of a vague idea. There can be several different definitions expressing a vague idea without a clearly 'best' one. Natural-world-justified definitions can be regarded as providing knowledge in the following sense: they are a possible formalisation of a preformal idea which is valuable.

Many definitions in the list of notions of unpredictability (cf. section 3.3) are natural-world-justified: I will now discuss one version of weak mixing (for discrete and continuous time), one version of a Bernoulli system (for discrete time) and the Kolmogorov-Sinai entropy (for discrete and continuous time) in detail. For illustrating natural-world-justification, it would suffice to consider the Kolmogorov-Sinai entropy. But the discussion of the remaining two definitions is crucial in order to provide the necessary background for the next sections. Moreover, all versions of strong mixing (Berkovitz, Frigg & Kronz 2006, p. 676; Hopf 1932*a*, p. 205) and Kolmogorov-mixing (Sinai 1963, p. 66) are natural-world-justified.

Weak mixing

Definition 16 *The discrete measure-preserving deterministic system* (M, Σ_M, μ, T) *is* weakly mixing *if, and only if, for all* $A, B \in \Sigma_M$ *there is a* $P \subseteq \mathbb{N}$ *of density zero such that*

$$\lim_{t\to\infty,\ t\notin P} \mu(T^t(A) \cap B) = \mu(A)\mu(B),$$

where $P \subseteq \mathbb{N}$ *is of density zero if, and only if,* $\lim_{t\to\infty,\, t\in\mathbb{N}} \#(P \cap \{i \,|\, i \leq t,\ i \in \mathbb{N}\})/t = 0$.

Definition 17 *The continuous measure-preserving deterministic system* (M, Σ_M, μ, T_t) *is* weakly mixing *if, and only if, for all* $A, B \in \Sigma_M$ *there is a* $P \subseteq \mathbb{R}^+$ *of density zero such that*

$$\lim_{t\to\infty,\ t\notin P} \mu(T_t(A) \cap B) = \mu(A)\mu(B),$$

where $P \subseteq \mathbb{R}^+$ *is of density zero if, and only if,* $\lim_{t\to\infty,\, t\in\mathbb{R}^+} \lambda(P \cap (0,\, t])/t = 0$*, where* λ *is the Lebesgue measure on* $\mathbb{R}$.

For a discrete measure-preserving deterministic system (M, Σ_M, μ, T) or a continuous measure-preserving deterministic system (M, Σ_M, μ, T_t) and a set $A \in \Sigma_M$, define A^t as the *event* that the state of the deterministic system is in A at time t. For instance, for the baker's system (Example 1) you could be interested in the event that the state of the deterministic system at time t is on the left side of the unit square, i.e., you could be interested in the event A^t where $A = [0, 1/2] \times [0, 1] \setminus D$.

Because the exact state of the deterministic system may be unknown, I introduce $p(A^t)$, the probability of the event A^t. Assume that the measure can be interpreted as time-independent probability. As explained in section 2.1, this is quite natural under certain interpretations. Then:

$$\text{For all } t \text{ and for all } A \in \Sigma_M:\ p(A^t) = \mu(A). \tag{3.1}$$

This idea can be generalised to joint simultaneous events as follows:

$$\text{For all } t \text{ and for all } A, B \in \Sigma_M:\ p(A^t \& B^t) = \mu(A \cap B). \tag{3.2}$$

This immediately implies:

$$\text{For all } t, t' \text{ and all } A, B \in \Sigma_M:\ p(A^t \& B^{t'}) = \mu(T^{t'-t}(A) \cap B) \tag{3.3}$$

since $T^{t'-t}(A)$ is the evolution of the set A from t to t'.[8]

[8] I can infer (3.3) from (3.2) as follows: $T^{t'-t}(A)$ contains exactly those points that are in A at time t. Consequently, $T^{t'-t}(A) \cap B$ consists of exactly those points which pass B at time t' and go through A at time t, i.e., for which $A^t \& B^{t'}$ is true. Thus from (3.2) it follows that $p(A^t \& B^{t'}) = \mu(T^{t'-t}(A) \cap B)$.

Definitions 16 and Definition 17 expresses that for any $A, B \in \Sigma_M$ and any $\varepsilon > 0$ there is a $t' \in \mathbb{N}$ or $t' \in \mathbb{R}^+$ and a set P of density zero with $|\mu(T^t(A) \cap B) - \mu(A)\mu(B)| < \varepsilon$ for all $t \geq t'$, $t \notin P$. Now assume, without loss of generality, that the event you want to predict occurs at time 0. Then from equation (3.3) it follows that Definition 16 and Definition 17 capture the following idea of unpredictability: for any event B^0, $B \in \Sigma_M$, any $A \in \Sigma_M$ and any $\varepsilon > 0$ there is a $t' \in \mathbb{N}$ or $\mathbb{R}^+$ and a set P of density zero with $|p(B^0 \,\&\, A^{-t}) - p(B^0)p(A^{-t})| < \varepsilon$ for all $t \geq t'$, $t \notin P$. That is, given an arbitrary level of precision $\varepsilon > 0$ any event is approximately probabilistically independent of almost any event that is sufficiently past. Independence is understood here as in probability theory. This unpredictability might apply, for instance, to systems in meteorology and make it hard to predict them.

Von Neumann (1932*b*, p. 591, p. 594) lists the main statistical properties of classical deterministic systems discussed in ergodic theory at that time. In this context he remarks that Definition 16 captures the preformal idea of approximate independence of almost all events explained above. Thus he argues that it is natural-world justified. This justification grew in importance with the rise of chaos research in the 1960s (see, e.g., Berkovitz, Frigg & Kronz 2006, p. 688). This justification also appears in a few standard books on ergodic theory (e.g., Walters 1982, p. 45), although in books often no justification is provided for weak mixing (e.g., Arnold & Avez 1968, pp. 21–22; Cornfeld et al. 1982, pp. 22–23; Sinai 2000, p. 21).

Especially before the rise of chaos research weak mixing appears to be mostly not naturally-world justified. This will be shown in subsection 3.4.2, where I will also discuss the key contexts in which weak mixing was introduced.

The next definition relates to the important topic of equivalence of measure-preserving deterministic systems.

Discrete Bernoulli system

The idea of an infinite sequence of probabilistically independent trials of an N-sided die is a very old one. Kolmogorov (1933) gave the modern measure-theoretic formulation of probability theory and laid the foundations for the modern theory of stochastic processes (as introduced in section 2.2) (von Plato 1994, pp. 230–233). Recall that in this modern framework a doubly-infinite sequence of independent rolls of an N-sided die where the possible outcomes are $\bar{M} = \{s_1, \ldots, s_N\}$ and the probability of obtaining outcome s_k is p_k, $1 \leq k \leq N$, $\sum_{k=1}^{N} p_k = 1$, is called a Bernoulli process; also, recall that a Bernoulli process can be represented as follows (see Example 4 in section 2.2): Ω is the set of realisations of the stochastic process, Σ_Ω is the σ-algebra generated by the semi-algebra of cylinder-sets, ν is the extension of the pre-measure defined by the independence property on the cylinder sets and $Z_t : \Omega \to \bar{M}$, $Z_t(\omega) = \omega_t$ (the t-th coordinate of ω). Then $\{Z_t;\ t \in \mathbb{Z}\}$ is a representation of the Bernoulli process.

Now I define a measure-preserving deterministic system: consider the following function, called a shift

$$T : \Omega \to \Omega \quad T((\ldots \omega_i \ldots)) = (\ldots \omega_{i+1} \ldots). \tag{3.4}$$

The shift is easily seen to be measurable and measure-preserving.

Definition 18 *The measure-preserving deterministic system* $(\Omega, \Sigma_\Omega, \nu, T)$ *as constructed above is called a* Bernoulli shift *with probabilities* $(p_1, \ldots, p_N)$.

The meaning of a Bernoulli shift is that it represents a Bernoulli process. For assume that one sees only the 0-th coordinate of the sequence ω, i.e., one applies the observation function $\Phi_0 : \Omega \to \bar{M}$, $\Phi_0(\omega) = \omega_0$ to the Bernoulli shift $(\Omega, \Sigma_\Omega, \nu, T)$. Then the possible outcomes of the Bernoulli process are the possible observed values of the Bernoulli shift $(\Omega, \Sigma_\Omega, \nu, T)$. It is clear that any realisation of the Bernoulli process r_ω, where r_ω generally denotes a realisation of a stochastic process (cf. section 2.2), is contained in the phase space Ω. And observing the solution s_{r_ω} of $(\Omega, \Sigma_\Omega, \nu, T)$ with Φ_0 exactly gives r_ω. Furthermore, the measure ν is defined by the probabilities which are assigned by the Bernoulli process to each cylinder set. Hence the probability distribution over the realisations of the Bernoulli process is the same as the one over the sequences of observed values of $(\Omega, \Sigma_\Omega, \nu, T)$. Thus a Bernoulli shift is a deterministic representation of a Bernoulli process.

In one of the first papers on ergodic theory, von Neumann (1932*b*) introduced the fundamental idea that measure-preserving deterministic systems are probabilistically equivalent, i.e., that their states can be put into one-to-one correspondence such that the corresponding solutions have the same probability distributions. He developed the definition of *isomorphic* deterministic systems to capture this idea (Sinai 1989, p. 833), and he called for a classification of measure-preserving deterministic systems up to isomorphism.

Definition 19 *The discrete measure-preserving deterministic systems* $(M_1, \Sigma_{M_1}, \mu_1, T_1)$ *and* $(M_2, \Sigma_{M_2}, \mu_2, T_2)$ *are* isomorphic *if, and only if, there are measurable sets* $\hat{M}_i \subseteq M_i$ *with* $\mu_i(M_i \setminus \hat{M}_i) = 0$ *and* $T_i\hat{M}_i \subseteq \hat{M}_i$ $(i = 1, 2)$*, and there is a bijection* $\phi : \hat{M}_1 \to \hat{M}_2$ *such that (i)* $\phi(A) \in \Sigma_{M_2}$ *for all* $A \in \Sigma_{M_1}, A \subseteq \hat{M}_1$*, and* $\phi^{-1}(B) \in \Sigma_{M_1}$ *for all* $B \in \Sigma_{M_2}, B \subseteq \hat{M}_2$*; (ii)* $\mu_2(\phi(A)) = \mu_1(A)$ *for all* $A \in \Sigma_{M_1}$*,* $A \subseteq \hat{M}_1$*; (iii)* $\phi(T_1(m)) = T_2(\phi(m))$ *for all* $m \in \hat{M}_1$*. For continuous measure-preserving deterministic systems* $(M_1, \Sigma_{M_1}, \mu_1, T_t^1)$ *and* $(M_2, \Sigma_{M_2}, \mu_2, T_t^2)$ *the definition of being* isomorphic *is the same except that condition (iii) is* $\phi(T_t^1(m)) = T_t^2(\phi(m))$ *for all* $m \in \hat{M}_1$ *and all* $t \in \mathbb{R}$ *(cf. Petersen 1983, p. 4).*

One easily sees that 'being isomorphic' is an equivalence relation.

Consequently, we see that the following definition captures the idea of a deterministic system which is probabilistically equivalent to a deterministic system representing a Bernoulli process, e.g., throwing a die:

Definition 20 (M, Σ_M, μ, T) *is a* discrete Bernoulli system *if, and only if, it is isomorphic to a Bernoulli shift.*

In many articles Definition 20 is natural-world-justified as capturing the idea that a deterministic system is probabilistically equivalent to a deterministic representation of a Bernoulli process (Ornstein 1989, p. 4; Rohlin 1960, p. 5). Walter's (1982, p. 107; see also Ornstein 1974, p. 4) comment

> Since a Bernoulli shift is really an independent identically distributed stochastic process indexed by the integers, we can think of a {discrete Bernoulli system} as an abstraction of such a stochastic process.[9]

[9]Square brackets indicate that the original notation has been replaced by the notation used in this dissertation. I will use this convention throughout.

shows that this justification is also found in standard books on ergodic theory. Yet some books do not provide any justification for Definition 20 (e.g., Shields 1973, p. 5).

Clearly, the Bernoulli shifts given by choices of N and, for each N, the choices of $p_1, \ldots p_N$ are discrete Bernoulli systems. In the next paragraph about the Kolmogorov-Sinai entropy we will say more about when Bernoulli shifts are isomorphic.

The next definition illustrates that a definition can be both natural-world-justified and proof-generated.

Kolmogorov-Sinai entropy

Assume that a probability distribution $P = (p_1, \ldots, p_n)$ is given over a set of possible symbols $(x_1, \ldots, x_n)$, $n \in \mathbb{N}$ (that is, $p_i \geq 0$ for all i and $\sum_{i=1}^{n} p_i = 1$). In information theory the amount of information gained when a symbol is received is understood to equal the amount of uncertainty reduced when a symbol is received. The *Shannon information* $S(P) = -\sum_{i=1}^{n} p_i \log(p_i)$ measures the average amount of uncertainty reduced when a symbol is received or, equivalently, the average amount of information gained when a symbol is received (see Cover & Thomas 2006; Frigg & Werndl 2010; Klir 2006, section 2.2.3).[10]

Ergodic theory and information theory can be connected as follows: first, recall Definition 7 of a partition α. Given a discrete measure-preserving deterministic system (M, Σ_M, μ, T) each $m \in M$ produces, relative to a partition $\alpha = \{\alpha_1, \ldots, \alpha_k\}$, a bi-infinite string of symbols $\ldots x_{-2}x_{-1}x_0x_1x_2 \ldots$ in an alphabet of k letters via the coding $x_j = \alpha_i$ if, and only if, $T^j(m) \in \alpha_i$, $j \in \mathbb{Z}$. Interpreting the measure-preserving deterministic system (M, Σ_M, μ, T) as the source, the output of the source are these strings $\ldots x_{-2}x_{-1}x_0x_1x_2 \ldots$. If the measure is interpreted as probability density, one has a probability distribution over these strings. Hence the whole apparatus of information theory can be applied to these strings.

In particular, given a partition $\alpha = \{\alpha_1, \ldots, \alpha_k\}$ of (M, Σ_M, μ), $H(\alpha) =$

[10] Throughout the dissertation 'log' stands for the logarithm to the basis of two. Also, $0 \log(0)$ is defined to be 0.

$-\sum_{i=1}^{k}\mu(\alpha_i)\log(\mu(\alpha_i))$ is the Shannon information of $P=(\mu(\alpha_1),\ldots,\mu(\alpha_k))$ and measures the average information of the symbol α_i. Let us regard strings of length n, $n\in\mathbb{N}$, produced by the deterministic system relative to a coding α as messages. The probability distribution of these possible strings of length n relative to α is $\mu(\beta_i)$, $1\leq i\leq h$, $\beta=\{\beta_1,\ldots,\beta_h\}=(\alpha\vee T^{-1}\alpha\vee\ldots\vee T^{-n+1}\alpha)$. Hence

$$H_n(\alpha,T)=\frac{1}{n}H(\alpha\vee T^{-1}\alpha\vee\ldots\vee T^{-n+1}\alpha) \tag{3.5}$$

measures the average amount of information which the measure-preserving deterministic system produces *per step* over the first n steps relative to the coding α. And the limit

$$H(\alpha,T)=\lim_{n\to\infty}H_n(\alpha,T), \tag{3.6}$$

which can be proven to exist, measures the average information which the measure-preserving deterministic system produces per step relative to α as time goes to infinity (Petersen 1983, pp. 233–240).

Now:

Definition 21 $E_{KS}(M,\Sigma_M,\mu,T)=\sup_\alpha\{H(\alpha,T)\}$ *is the* Kolmogorov-Sinai entropy *of the discrete measure-preserving deterministic system* (M,Σ_M,μ,T).

It is clear that it measures the highest average amount of information that the deterministic system can produce per step relative to a coding, or, equivalently, the average amount of uncertainty reduced per step relative to a coding. The Shannon information measures uncertainty, and this uncertainty can be regarded as a form of unpredictability (cf. Frigg 2004, Frigg 2006). Hence a positive Kolmogorov-Sinai entropy means that relative to some codings the behaviour of the system is unpredictable.

For a continuous measure-preserving deterministic system (M,Σ_M,μ,T_t) it can be shown that for any t_0, $-\infty<t_0<\infty$, (Sinai 2007):

$$E_{KS}(M,\Sigma_M,\mu,T_{t_0})=|t_0|E_{KS}(M,\Sigma_M,\mu,T_1), \tag{3.7}$$

where $E_{KS}(M,\Sigma_M,\mu,T_{t_0})$ denotes the Kolmogorov-Sinai entropy of the discrete measure-preserving deterministic system (M,Σ_M,μ,T_{t_0}) and $E_{KS}(M,\Sigma_M,\mu,T_1)$ is the Kolmogorov-Sinai entropy of the discrete measure-preserving deterministic system (M,Σ_M,μ,T_1). Consequently:

Definition 22 *The* Kolmogorov-Sinai entropy *of a continuous measure-preserving deterministic system* (M, Σ_M, μ, T_t) *is defined as* $E_{KS}(M, \Sigma_M, \mu, T_1)$.

And it measures the average amount of information or uncertainty produced by the continuous deterministic system over one time unit.

Having worked for several years on information theory, Kolmogorov (1958) was the first to apply information-theoretic ideas to ergodic theory. He introduced a definition of entropy only for what are nowadays called Kolmogorov-systems. Based on Kolmogorov's work, Sinai (1959) introduced a different notion of entropy which applies to all measure-preserving deterministic systems, the now canonical Definition 21 and Definition 22. Sinai also proved—a big surprise at that time—that automorphisms on the torus have positive Kolmogorov-Sinai entropy and thus are unpredictable because they produce information. Kolmogorov and Sinai were motivated by finding a concept which characterises the amount of randomness or unpredictability of a system (Frigg & Werndl 2010, Shiryaev 1989, Sinai 2007, Werndl 2009*c*). More specifically, as Halmos (1961, p. 76) explains: "Intuitively speaking, the entropy $\{E_{KS}\}$ is the greatest quantity of information obtainable about the universe per day [i.e., step] by repeated performances of experiments with a finite [...] number of possible outcomes". Hence Definition 21 is natural-world-justified by capturing the idea of the average amount of information produced per step explained above.

Also in some standard books on ergodic theory Definition 21 and Definition 22 are natural-world-justified in this way (Billingsley 1965, p. 63; Petersen 1983, pp. 233–240). It should, however, be mentioned that in many books Definition 21 and Definition 22 are not justified at all (e.g., Arnold & Avez 1968, pp. 35–50; Cornfeld et al. 1982, pp. 246–257; Sinai 2000, pp. 40–43).

Interestingly, Definition 21 of the Kolmogorov-Sinai entropy is *also proof-generated*. And, so far as I can see, it is the only notion of unpredictability in ergodic theory (cf. section 3.3) which is proof-generated. The central internal problem of ergodic theory is the following: which measure-preserving deterministic systems are isomorphic (cf. Definition 19)? Given a measure space (M, Σ_M, μ) consider $L^2(M, \Sigma_M, \mu)$, the Hilbert space of real-valued square integrable functions on (M, Σ_M, μ) where two functions which differ by a set of measure zero are identified and the inner product is $<f, g> = \int_M fg\, d\mu$ for any elements f, g of $L^2(M, \Sigma_M, \mu)$. Now suppose that a discrete measure-preserving deterministic system (M, Σ_M, μ, T) is given. Then $U_T : L^2(M, \Sigma_M, \mu) \to L^2(M, \Sigma_M, \mu)$, $U_T(f) = f(T(m))$, is a linear operator. Likewise, given a continuous measure-preserving deterministic system (M, Σ_M, μ, T_t) and any $t \in \mathbb{R}$, the map $U_{T_t} : L^2(M, \Sigma_M, \mu) \to L^2(M, \Sigma_M, \mu)$, $U_{T_t}(f) = f(T_t(m))$, is a linear operator. In fact, U_T and U_{T_t} are unitary operators. An operator V on a Hilbert space is called *unitary* if, and only if, (i) V is linear, (ii) V is invertible and (iii) $< Vf, Vg >=< f, g >$ for all elements f, g of the Hilbert space.[11] This was first discovered by Koopman (1931), and the investigation of measure-preserving deterministic systems by these operators is referred to as the spectral theory of deterministic systems (cf. Petersen 1983, section 2).

Measure-preserving deterministic systems which are equivalent from this viewpoint are said to be spectrally isomorphic. Formally, the discrete measure-preserving deterministic systems $(M_1, \Sigma_{M_1}, \mu_1, T_1)$ and $(M_2, \Sigma_{M_2}, \mu_2, T_2)$ are *spectrally isomorphic* if, and only if, there exists an unitary operator V on $L^2(M_1, \Sigma_{M_1}, \mu_1)$ such that $V^* U_{T_1} V = U_{T_2}$, where V^* is the adjoint of V. And the continuous measure-preserving deterministic systems $(M_1, \Sigma_{M_1}, \mu_1, T_t^1)$ and $(M_2, \Sigma_{M_2}, \mu_2, T_t^2)$ are *spectrally isomorphic* if, and only if, there exists an unitary operator V on $L^2(M_1, \Sigma_{M_1}, \mu_1)$ such that $V^* U_{T_t^1} V = U_{T_t^2}$ for all $t \in \mathbb{R}$.

In the 1950s it was known that deterministic systems with discrete spec-

[11] Clearly, U_T and U_{T_t} are linear. And it is clear that U_T is invertible and that $U_T^{-1}(f) = f(T^{-1}(m))$, and that U_{T_t} is invertible for all $t \in \mathbb{R}$ and that $U_{T_t}^{-1}(f) = f(T^{-t}(m))$. Finally, the fact that (M, Σ_M, μ, T) and (M, Σ_M, μ, T_t) are measure-preserving implies that (cf. Petersen 1983, section 2):

$$<U_T(f), U_T(g)> = \int_M U_T(f) U_T(g) d\mu = \int_M f(T(m)) g(T(m)) d\mu = \int_M f(m) g(m) d\mu = <f, g> \tag{3.8}$$

and

$$<U_{T_t}(f), U_{T_t}(g)> = \int_M U_{T_t}(f) U_{T_t}(g) d\mu = \int_M f(T_t(m)) g(T_t(m)) d\mu = \int_M f(m) g(m) d\mu = <f, g> \tag{3.9}$$

is true for all characteristic functions, all combinations of characteristic functions and hence, by approximation, also for all $f, g \in L^2(M, \Sigma_M, \mu)$.

trum are isomorphic if, and only if, they are spectrally isomorphic and that this is not so for deterministic systems with mixed spectrum. Most importantly, however, is the case of a continuous spectrum since measure-preserving deterministic systems typically have this property (Arnold & Avez 1968, pp. 27–32). Measure-preserving deterministic systems have *continuous spectrum* if, and only if, their only eigenfunctions are the constant functions. That is, for discrete time if, and only if, the only functions $f \in L^2(M, \Sigma_M, \mu)$ satisfying $U_T(f) = \lambda f$, where $\lambda \in \mathbb{R}$ arbitrary, are the constant functions; and for continuous time if, and only if, the only functions $f \in L^2(M, \Sigma_M, \mu)$ satisfying $U_{T_t}(f) = \lambda f$ for all $t \in \mathbb{R}$, where $\lambda \in \mathbb{R}$ arbitrary, are the constant functions. For measure-preserving deterministic systems with continuous spectrum, e.g., discrete Bernoulli systems, the conjecture emerged that spectrally isomorphic systems are not always isomorphic, but the problem resisted solution.

Kolmogorov (1958) and Sinai (1959) were motivated by making progress about this conjecture (Shiryaev 1989, pp. 914–915; Sinai 1989, pp. 834–836). And Kolmogorov's (1958) main result is that this conjecture is true. As hinted at by Rohlin (1960, pp. 1–2, p. 8), the Kolmogorov-Sinai entropy can be justified as being precisely the definition which is needed to prove that conjecture, i.e., it is proof-generated. The argument, which goes back to Kolmogorov's work, is as follows: isomorphic measure-preserving deterministic system have the same Kolmogorov-Sinai entropy. Now look at Bernoulli shifts, whose Kolmogorov-Sinai entropy is $\sum_i p_i \log(p_i)$ and hence takes a continuum of different values. Since all Bernoulli shifts are spectrally isomorphic, there is a continuum of measure-preserving deterministic systems being spectrally isomorphic but not isomorphic.

Billingsley's (1965, p. 65) comment

> It is essential to understand the difference between $H(\alpha, T)$ and $\{E_{KS}(M, \Sigma_M, \mu, T)\}$ and why the latter is introduced. If the entropy of T were taken to be $H(\alpha, T)$ for some "naturally" selected α [...], then it would be useless for the isomorphism problem.

shows that the justification of Definition 21 as being proof-generated made it into standard books on ergodic theory too (see also Petersen 1983, p. 227,

p. 246).

Let us turn to the second kind of justification I have identified.

3.4.2 Condition justification

I claim that another kind of justification abounds in my case study: *a definition is justified by the fact that it is equivalent in an allegedly natural way to a previously specified condition which is regarded as mathematically valuable.* I speak here of *condition-justified* definitions.

If the previously specified condition is valuable and the kind of equivalence is natural, condition justification is a reasonable kind of justification.[12] A condition-justified definition can be regarded as providing knowledge because it answers the question of which definition corresponds naturally to a previously specified condition.

The following notions of unpredictability in ergodic theory (cf. section 3.3) are condition-justified: all versions of weak mixing (for discrete and continuous time) and one version of being a discrete Bernoulli system (for discrete time). Let us discuss them now.

Weak mixing

Recall Definition 16 and Definition 17 of weak mixing. Two alternative equivalent definitions for discrete and continuous time are (Cornfeld et al. 1982, pp. 22–23; Petersen 1983, pp. 65–67):

Definition 23 *A discrete measure-preserving deterministic system* (M, Σ_M, μ, T) *is* weakly mixing *if, and only if, for all* $A, B \in \Sigma_M$

$$\lim_{t\to\infty} \frac{1}{t} \sum_{i=0}^{t-1} |\mu(T^i(A) \cap B) - \mu(A)\mu(B)| = 0.$$

[12] For the condition-justified definitions of my case study we will see why the conditions are valuable and the equivalences are natural. Yet characterising what constitutes valuable conditions or natural kinds of equivalence at a general level would require further research.

Definition 24 *A continuous measure-preserving deterministic system* (M, Σ_M, μ, T_t) *is* weakly mixing *if, and only if, for all* $A, B \in \Sigma_M$

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t |\mu(T_\tau(A) \cap B) - \mu(A)\mu(B)| d\tau = 0,$$

where the measure on the time axis $\tau \in \mathbb{R}_0^+$ is the Lebesgue measure.

Definition 25 *The discrete measure-preserving deterministic system* (M, Σ_M, μ, T) *is* weakly mixing *if, and only if, for all* $f, g \in L^2(M, \Sigma_M, \mu)$

$$\lim_{t\to\infty} \frac{1}{t} \sum_{i=0}^{t-1} | \int f(T^i(m))g(m)d\mu - \int f(m)d\mu \int g(m)d\mu |= 0,$$

where $L^2(M, \Sigma_M, \mu)$ is the Hilbert space of real-valued square integrable functions on (M, Σ_M, μ) where two functions which differ by a set of measure zero are identified.

Definition 26 *The continuous measure-preserving deterministic system* (M, Σ_M, μ, T_t) *is* weakly mixing *if, and only if, for all* $f, g \in L^2(M, \Sigma_M, \mu)$

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t | \int f(T_\tau(m))g(m)d\mu - \int f(m)d\mu \int g(m)d\mu \mid d\tau = 0,$$

where $L^2(M, \Sigma_M, \mu)$ is the Hilbert space of real-valued square integrable functions on (M, Σ_M, μ) where two functions which differ by a set of measure zero are identified, and the measure on the time axis $\tau \in \mathbb{R}_0^+$ is the Lebesgue measure.

I already argued that Definition 16 and Definition 17 of weak mixing can be natural-world-justified. The first three papers discussing weak mixing seem to be Hopf (1932*a*), Hopf (1932*b*), and Koopman & von Neumann (1932), which all discuss weak mixing for continuous deterministic sytems. These papers show that there is more to say; for three reasons.

First, Hopf (1932*a*) starts by emphasising the importance of ergodicity for statistical mechanics (cf. Definition 2.5). He then considers a statistical property discussed by Poincaré: when initially a certain part of a fluid is coloured, experience shows that after a long time the colour uniformly dissolves in the fluid. Mathematically, Hopf expresses this by strong mixing.

Definition 27 *A discrete measure-preserving deterministic system* (M, Σ_M, μ, T) *is* strongly mixing *if, and only if, for all* $A, B \in \Sigma_M$

$$\lim_{t\to\infty} \mu(T^t(A) \cap B) = \mu(A)\mu(B).$$

Definition 28 *A continuous measure-preserving deterministic system* (M, Σ_M, μ, T_t) *is* strongly mixing *if, and only if, for all* $A, B \in \Sigma_M$

$$\lim_{t\to\infty} \mu(T_t(A) \cap B) = \mu(A)\mu(B).$$

By looking at Definition 16 and Definition 17, we immediately see that any strongly mixing measure-preserving deterministic system is also weakly mixing. Interested in the interrelationship between strong mixing and ergodicity, Hopf indeed conjectures that a continuous measure-preserving deterministic system (M, Σ_M, μ, T_t) is strongly mixing if, and only if, for all $t_0 \in \mathbb{R}^+$ the discrete measure-preserving deterministic system $(M, \Sigma_M, \mu, T_{t_0})$ is ergodic. Yet he is unable to prove this (it was later shown to be false, see Lind 1975). As a result, Hopf attends to the question of which weaker statistical property is equivalent to the condition that for all $t_0 \in \mathbb{R}^+$ the discrete measure-preserving deterministic system $(M, \Sigma_M, \mu, T_{t_0})$ is ergodic. The answer he arrives at is Definition 26. Therefore, Definition 26 of weak mixing is condition-justified because its justification stems from it being equivalent in a natural way to a condition regarded as valuable. This justification only works for continuous deterministic systems and not for discrete deterministic system because it is not true that a discrete measure-preserving deterministic system (M, Σ_M, μ, T) is weakly mixing if, and only if, for all $t_0 \in \mathbb{N}$ the discrete deterministic system $(M, \Sigma_M, \mu, T^{t_0})$ is ergodic.[13]

Second, Hopf (1932*b*) is concerned with Gibbs' fundamental hypothesis that any initial distribution tends toward statistical equilibrium, and he derives several conditions under which this hypothesis holds true. Within this context, the question arises how properties of a discrete measure-preserving deterministic system (M, Σ_M, μ, T) or a continuous measure-preserving deterministic system (M, Σ_M, μ, T_t) relate to the composite system $(M \times M, \Sigma_M \otimes$

[13] The irrational rotation on the circle, which I will discuss in subsection 5.5.2, is a counterexample (Petersen 1983, p. 8).

$\Sigma_M, \mu\times\mu, T\times T)$ or $(M\times M, \Sigma_M\otimes\Sigma_M, \mu\times\mu, T_t\times T_t)$ comprising two copies of the single system.[14] Because of the importance of ergodicity, it is natural to ask: which property of the single system is equivalent to the composite system being ergodic? Hopf (1932*b*) provides the answer for continuous deterministic systems, namely Definition 26 of weak mixing. The same answer, namely weak mixing, is also true for discrete measure-preserving deterministic systems (Halmos 1949, pp. 1021–1022). Hence weak mixing is condition-justified as Halmos (1949, p. 1022) stresses by referring to Definition 16 and Definition 23: an "indication that weak mixing is more than an analytic artificiality is in the assertion that T is weakly mixing if, and only if, its direct product with itself is indecomposable [ergodic]".

Third, when discussing Definition 21 of the Kolmogorov-Sinai entropy, we encountered the property of a continuous spectrum which arises in spectral theory. Koopman & von Neumann (1932) emphasise the naturalness of, and devote their paper to, this property. From the beginning of ergodic theory the correspondence of concepts from spectral theory and set-theoretic and integral-theoretic concepts from ergodic theory has been a core theme. Hence it was natural to address the question, as Koopman & von Neumann did, which set-theoretic or integral-theoretic definition is equivalent to having a continuous spectrum. The answer they arrived at for continuous deterministic systems is Definition 17 of weak mixing, and the same answer, namely weak mixing, is also true for discrete deterministic systems (Petersen 1983, p. 64). Thus, again, Definition 16 and Definition 17 of weak mixing are condition-justified.

I have found no book motivating the continuous version of weak mixing by the condition that for all t_0, $t_0\neq 0$, the discrete deterministic system $(M, \Sigma_M, \mu, T_{t_0})$ is ergodic. This might be because that characterisation does not hold for discrete systems. The other two interpretations of weak mixing as condition-justified appear in standard books on ergodic theory, e.g., Halmos (1956, p. 39) and Petersen (1983, p. 64). The latter comments:

[14] Here $M\times M$ is the Cartesian product of M with M; $\Sigma_M\otimes\Sigma_M$ is the product σ-algebra, that is, the σ-algebra generated by sets of the form $A\times B$, where $A, B\in\Sigma_M$; $\mu\times\mu$ is the product measure, that is, the unique measure satisfying the property $\mu\times\mu(A\times B) = \mu(A)\mu(B)$; $T\times T(m,q) = (T(m), T(q))$ and $T_t\times T_t(m,q) = (T_t(m), T_t(q))$.

> That the concept of weak mixing is natural and important can be seen from the following theorem, according to which a transformation is weakly mixing if, and only if, its only measurable eigenfunctions are the constants.

To summarise, all versions of weak mixing are condition-justified because their justification stems from their being equivalent in a natural way to a condition regarded as valuable. The next definition illustrates the danger of not appreciating that a definition is condition-justified.

Discrete Bernoulli system

Recall Definition 20 of a discrete Bernoulli system. The appeal to isomorphisms makes this definition indirect. Furthermore, most states of the deterministic systems encountered in the sciences, e.g., states of Newtonian systems, are not infinite sequences. Thus it is often easier to work without notions referring to infinite sequences. In investigating simple systems isomorphic to Bernoulli shifts, it became clear that proving an isomorphism amounts to finding a partition which can be used to code the dynamics. Hence it was natural to ask which condition that does not appeal to isomorphisms and infinite sequences, but to partitions, is equivalent to a discrete Bernoulli system.

Definition 29 *The discrete measure-preserving deterministic system* (M, Σ_M, μ, T) *is a* discrete Bernoulli system *if, and only if, there is a partition* α *such that*
(i) $T^i\alpha$ *is an independent sequence, i.e., for any distinct* $i_1, \ldots, i_r \in \mathbb{Z}$*, and not necessarily distinct* $\alpha_j \in \alpha$*,* $j = 1, \ldots, r$ $(r \geq 1)$*:*

$$\mu(T^{i_1}\alpha_1 \cap \ldots \cap T^{i_r}\alpha_r) = \mu(\alpha_1) \ldots \mu(\alpha_r).$$

(ii) Σ_M *is generated by* $\{T^i\alpha \,|i \in \mathbb{Z}\}$*.*

Hence Definition 29 can be justified by the fact that it gives an answer to the above question, i.e., it is condition-justified. Standard books on ergodic theory also hint at this justification (Shields 1973, p. 8, p. 11; Sinai 2000, p. 47).

There have been attempts to justify Definition 29 as capturing a preformal idea of randomness or unpredictability. Interpreting the measure as time-independent probability, condition (i) captures the idea that any finite number of events of a specific partition at different times are probabilistically independent. Berkovitz et al. (2006) argue that because condition (i) can be thus interpreted, discrete Bernoulli systems capture unpredictability;[15] they do not say anything about condition (ii). Yet since (i) is only one part of this definition, this justification of Definition 29 fails.[16] Generally, if a definition does not capture the idea it is said to capture, the justification fails because it is unclear why this definition is chosen.

Batterman's (1991) and Sklar's (1993, pp. 238–239) motivation for Definition 29 is also that it captures a preformal idea of randomness or unpredictability. Their argument as expressed by Batterman (1991, pp. 249–250) is:

> Now let us see just how random a Bernoulli system is. [...] The Bernoulli systems are those in which knowing the entire past history of box-occupations even relative to a partition (measurement) which is generating in the above sense, is insufficient (in the sense of being probabilistically independent) for improving the odds that the system will next be found in a given box.

[15] Actually, a slip occurred in Berkovitz et al.'s (2006, p. 667) interpretation of condition (i); (i) holds only for any finite number of events of a *specific* partition at different times, not for *any* events.

[16] For instance, the following measure-preserving deterministic system fulfills (i) but not (ii): let $M = ([0,1] \times [0,1] \times [0,1]) \setminus (D \times [0,1])$ where D is defined as for the baker's system (cf. Example 1). Let Σ_M bet the Lebesgue σ-algebra on M and μ be the Lebesgue measure. Let

$$T(m,y,z) = (2m, \frac{y}{2}, z) \text{ if } 0 \leq m < \frac{1}{2},\ (2m-1, \frac{y+1}{2}, z) \text{ if } \frac{1}{2} \leq m \leq 1.$$

Obviously, for (M, Σ_M, μ, T) condition (i) of Definition 29 holds for $\alpha = \{\{m \in M \mid 0 \leq m < \frac{1}{2}\}, \{m \in M \mid \frac{1}{2} \leq m \leq 1\}\}$. But (M, Σ_M, μ, T) is not a discrete Bernoulli system since it is not even ergodic.

As an interpretation of randomness or unpredictability this is puzzling. Even if it exactly corresponded to Definition 29,[17] it is unclear, *from the viewpoint of capturing a preformal idea of randomness or unpredictability*, why independence is required relative to *generating* partitions; and I found no convincing justification for this.

It seems that the difficulty stems from the fact that Definition 29 is really condition-justified. As we have seen for weak mixing, condition-justified definitions may in other contexts also capture a preformal idea valuable in some sense. However, often—and this is true for Definition 29 as discussed—this will not be the case. Then there is the danger of not appreciating that a definition is condition-justified and claiming that it captures a valuable preformal idea, when it does not. It seems that in interpreting Definition 29 Batterman and Sklar fell into this trap. This danger is similar to the one identified by Lakatos (1976, p. 153), viz. claiming that a proof-generated definition captures a valuable preformal idea when it does not.

Let us now turn to the final kind of justification I have identified.

3.4.3 Redundancy justification

I call *a definition which is justified because it eliminates as redundant at least one condition in an already accepted definition redundancy-justified.* A redundancy-justified definition can be regarded as providing knowledge since it shows that specific conditions in an accepted definition are redundant.

It is obviously desirable in mathematics to find out whether there are any redundant conditions in an already accepted definition. Typically, both the original definition, and the one in which the redundant conditions are

[17]It does not. First, their interpretation does not make clear that the matter of concern is the *existence* of a partition satisfying (i) and (ii). Even if this is disregarded, their interpretation applies to more systems than discrete Bernoulli systems. This is so because it applies to every discrete measure-preserving deterministic system where there is a generating partition where any events constituting the entire-history of box-occupation are of probability zero, and some of these deterministic systems are not Bernoulli (Ornstein 1974, pp. 93–95). The correct thing to say is: any finite number of events of a specific partition at different times are probabilistically independent, even though the partition is generating.

eliminated, each have their own advantages. It depends on the definitions, but the former might be easier to understand or might allow for a more fine-grained analysis; the latter is simpler (in the sense of being more concise), and it might be that only the latter is easier to use in proofs, allows for natural generalizations, or suggests important analogies.

So when is it better to propound the original definition? And when is it better to introduce instead the new definition without the redundant conditions, i.e., when is redundancy justification a reasonable kind of justification? I think the answer depends on the definition and the context in which the definition is considered. For the purpose of an introductory textbook it might be better to propound the original definition because it is easier to understand. Conversely, for the purpose of a research article it might be better instead to use the new, concise definition, since it is easier to use in some proofs. Furthermore, in many cases it does not seem to matter much whether the original definition or the definition without the redundant conditions is introduced, so long as the origin of the definition and the redundant conditions are clearly pointed out.

As in the case of proof-generated and condition-justified definitions, there is the danger of not understanding that a definition is redundancy-justified and claiming that it captures a valuable preformal idea, when it does not.

Two definitions in the list of notions of unpredictability in ergodic theory (cf. section 3.3) are redundancy-justified: the continuous version of a Bernoulli system, which I will discuss for illustration, and a Kolmogorov-system (Sinai 1963, pp. 64–65; Uffink 2007, pp. 94–96).

Continuous Bernoulli system

We have seen that Kolmogorov (1958) and Sinai (1959) established that isomorphic discrete Bernoulli systems have the same Kolmogorov-Sinai entropy (cf. subsection 3.4.1). A decade later Ornstein (1970*a*, 1971) proved the converse, i.e., that discrete Bernoulli systems with equal entropy are isomorphic.

Having established that celebrated result, Ornstein became interested in finding an analogous definition of a Bernoulli system for continuous time, and he asked whether the Kolmogorov-Sinai entropy could be used to classify them too. The most obvious definition of a continuous measure-preserving deterministic system (M, Σ_M, μ, T_t) describing an independent process is that for all $t_0 \in \mathbb{R}$, $t_0 \neq 0$, the discrete measure-preserving deterministic system $(M, \Sigma_M, \mu, T_{t_0})$ is a discrete Bernoulli system. Ornstein (1973*a*) first introduces this definition of a continuous Bernoulli system, and then he shows that there are redundant conditions in this definition because it is equivalent to the following definition:

Definition 30 *The continuous measure-preserving deterministic system* (M, Σ_M, μ, T_t) *is a* continuous Bernoulli system *if, and only if, the discrete measure-preserving deterministic system* (M, Σ_M, μ, T_1) *is a discrete Bernoulli system.*

Hence Definition 30 is redundancy-justified because it eliminates redundant conditions. In this way it seems to be justified in Ornstein's (1974, p. 56) book too.[18]

Ornstein (1973*b*) indeed showed that two continuous Bernoulli systems are isomorphic if, and only if, they have the same Kolmogorov-Sinai entropy. From Ornstein's result immediately follows that even more holds, namely that up to a scaling of the time t any two continuous Bernoulli systems are isomorphic. Let me explain this. For any continuous measure-preserving deterministic system (M, Σ_M, μ, T_t) the Kolmogorov-Sinai entropy of the discrete deterministic system $(M, \Sigma_M, \mu, T_{t_0})$, $t_0 \in \mathbb{R}$ arbitrary, $t_0 \neq 0$, is $|t_0|$ times the Kolmogorov-Sinai entropy of the discrete deterministic system (M, Σ_M, μ, T_1) (cf. equation (3.7)). So assume that two continuous Bernoulli systems (M, Σ_M, μ, T_t) and $(M_2, \Sigma_{M_2}, \mu_2, T_t^2)$ with Kolmogorov-Sinai entropy $E_{KS}(M, \Sigma_M, \mu, T_t)$ and $E_{KS}(M_2, \Sigma_{M_2}, \mu_2, T_t^2)$, respectively, are given. Now make the transformation $t' = ct$, for $c = \frac{E_{KS}(M_2, \Sigma_{M_2}, \mu_2, T_t^2)}{E_{KS}(M, \Sigma_M, \mu, T_t)}$. Then we obtain that (M, Σ_M, μ, T_t) is isomorphic to $(M_2, \Sigma_{M_2}, \mu_2, T_{t'}^2)$ since the Kolmogorov-Sinai entropy of the continuous measure-preserving deterministic system $(M_2, \Sigma_{M_2}, \mu_2, T_{t'}^2)$ is the Kolmogorov-Sinai entropy of the

[18] Ornstein (1974, p. 56) expresses this indirectly by introducing continuous Bernoulli systems as follows; "We will call a flow $\{(M, \Sigma_M, \mu, T_t)\}$ a {continuous Bernoulli system} if $\{(M, \Sigma_M, \mu, T_1)\}$ is a {discrete Bernoulli system}. (We will prove later that if $\{(M, \Sigma_M, \mu, T_1)\}$ is a {continuous Bernoulli system}, then $\{(M, \Sigma_M, \mu, T_{t_0})\}$ for each fixed t_0 is a {discrete Bernoulli system})."

discrete measure-preserving deterministic system $(M_2, \Sigma_{M_2}, \mu_2, T^2_{\frac{1}{c}})$, which is $\frac{1}{c}E_{KS}(M_2, \Sigma_{M_2}, \mu_2, T^2_t) = E_{KS}(M, \Sigma_M, \mu, T_t)$.

3.4.4 Occurrence of the kinds of justification

To sum up: in addition to Lakatos's proof-generated definitions, I have identified three kinds of justification of definitions. To my knowledge, condition justification and redundancy justification have not been identified before. I do not claim that the kinds of justification I have discussed are the only ones at work in mathematics. Further studies might unveil yet other ones.

Two more general comments about justifying definitions should be added here. First, for any kind of justification there are three possibilities: (i) a definition is reasonably justified in this way; (ii) it is justified but not reasonably justified in this way; (iii) it is not justified in this way. As regards (ii), for instance, if the idea of being equivalent in a measure-theoretic sense to an independent process like throwing a die was not valuable, Definition 20 would be natural-world-justified but not reasonably justified. Second, an already justified definition has sometimes additional good features which support this definition but which do not by themselves constitute a sufficient justification. These features may also be important in deciding between different definitions. For instance, it is often said that a merit of the Kolmogorov-Sinai entropy is its neat connection to other notions of unpredictability such as being a Kolmogorov-system. These are good features but not sufficient justifications; since if there were no further reasons for studying the definition, there would still remain the question why we should regard it as worth considering (cf. Smith 1998, pp. 174–175).

How widely do the kinds of justification I have discussed occur? To answer this, I first comment on the notion of a mathematical subject. I think that regardless of which plausible understanding of 'subject' is adopted, my claims are true. But a possible way to operationalise this idea is the following: with the subjects identified by the *Mathematical Subject Classification*[19] it would

[19] This is a five digit classification scheme of subjects formulated by the American Mathematical Society; see www.ams.org/msc. For our purposes subjects concerned with education, history or experimental studies have to be excluded.

be possible to create a list of subjects within mathematics from the nineteenth century up to today. Then the definitions of my case study (notions of unpredictability in ergodic theory) belong to the mathematical subject 'strange attractors, chaotic dynamics'.

Based on my knowledge of mathematics, I endorse the following claims about mathematics produced in the twentieth century and up to the present day:[20] *all the kinds of justifications I have discussed are widespread.* More specifically, proof-generated, condition-justified, and redundancy-justified definitions are all found in the majority of mathematical subjects with explicit definitions. Also, for nearly all mathematical subjects with explicit definitions which (among other things) aim at describing or understanding the natural world, natural-world-justified definitions are found. This includes subjects not only from what is called applied mathematics but also from pure mathematics, e.g., measure theory. Furthermore, as in my case study, *for nearly all mathematical subjects with explicit definitions many different ways of justifying definitions are found and are reasonable.* Indeed, I would be surprised if one subject could be found where only one kind of justification is important. Clearly, my case study shows that for the subject 'strange attractors, chaotic dynamics' these claims hold true.

For my case study the argumentation involved in justifying definitions is typically not explicitly stated but is merely hinted at or merely implicit in the mathematics. Because of the conventional style of mathematical writing, this appears to be generally the case in mathematics, as also Lakatos (1976, pp. 142–144) claimed. Also, it should be mentioned that detailed knowledge of parts of ergodic theory is necessary to assess how definitions are justified in my case study. This confirms Tappenden's claim that judgments about definitions require detailed knowledge of the relevant mathematics (cf. section 3.2).

Let us reflect on the interrelationships between the kinds of justification, an issue which seems not discussed in the literature.

[20]Starting with the twentieth century is somewhat arbitrary. All the here-discussed kinds of justification appear also important in nineteenth century mathematics. Yet older mathematics may be significantly different. Hence a close investigation would be necessary to identify the role the kinds justification play in older mathematics.

3.5 Interrelationships between the kinds of justification

In what follows when I speak of an argument for a definition I mean that a reason is provided for a definition which cannot be split into two separate reasons for this definition. Now I first ask about the *interrelationships in one argument*: assume that a specific argument establishes that a definition is justified according to one kind of justification. Can it be that this argument implies that the definition is at the same time also justified according to another kind of justification? Intuitively, one might think that in an argument a definition can only be justified according to one kind of justification. Yet, as we will see, the matter is more complicated. Second, I ask about the *interrelationships between the kinds of justification in different arguments*: if different arguments justify the same definition, what combination of kinds of justification do we find? I will discuss these two cases in the next two subsections.

3.5.1 One argument

Clearly, there are arguments where a definition is only proof-justified, natural-world-justified, condition-justified or redundancy-justified. For example, uniform convergence as discussed by Lakatos (1976, pp. 131–133) is only proof-justified, Definition 20 of a discrete Bernoulli system as capturing the idea of a measure-preserving system being equivalent to an independent process is only natural-world-justified, weak mixing as corresponding to ergodicity of the composite system is only condition-justified, and Definition 30 of a continuous Bernoulli system as eliminating redundant conditions is only redundancy-justified.

By going back to the characterisation of the kinds of justification, we see that the intuition that in an argument a definition can only be (reasonably) justified according to one kind of justification is correct except for one case. Namely, in rare cases condition-justified definitions are at the same time proof-generated in an argument. This is so if, and only if, the kind of equivalence is regarded as natural because it occurs in the formulation of a conjecture that should be established. For example, assume the following conjecture is regarded as valuable: each function in a convergent sequence of functions is continuous if, and only if, the limit function of the convergent sequence is continuous. Further, assume that sequences of pointwise convergent continuous functions without continuous limit functions are known. Then mathematicians might ask: how has the notion of convergence to be changed such that if, and only if, the limit function is continuous the sequence of continuous functions is convergent? The definition answering this question would be clearly condition-justified. But it would also be proof-generated since it is needed in order to prove the above conjecture.

Let us now turn to the interrelationships in different arguments.

3.5.2 Different arguments

In our case study different arguments establish that weak mixing is condition-justified: weak mixing corresponds to ergodicity of the composite deterministic system, to the set-theoretic or integral-theoretic condition equivalent to having a continuous spectrum, and for continuous measure-preserving deterministic systems to the condition that for all $t_0 \in \mathbb{R}^+$ the discrete measure-preserving deterministic system $(M, \Sigma_M, \mu, T_{t_0})$ is ergodic. Generally, *one and the same definition can be (reasonably) justified in the same way in different arguments by referring to different conjectures, preformal ideas etc.* For proof-generated definitions Lakatos (1976, pp. 127–128) also recognises this pattern.

What is more, we have seen that in different arguments Definition 16 and Definition 17 of weak mixing are justified in different ways: as mentioned above, these definitions are condition-justified but also natural-world-justified, expressing the idea that given an arbitrary level of precision $\varepsilon > 0$ any event is approximately independent of almost any event that is sufficiently past. Likewise, the discrete version of the Kolmogorov-Sinai entropy is natural-world-justified, expressing the idea of the highest average amount of information produced per step relative to a coding; but it is also proof-generated concerning the conjecture that spectrally isomorphic systems are not always isomorphic. Generally, *one and the same definition can in differ-*

ent arguments be (reasonably) justified in different ways.

Finally, a definition which is justified in any way can be used to (reasonably) justify a definition in an arbitrary way. In this sense the different kinds of justification are closely connected. For example, the natural-world-justified Definition 20 of a discrete Bernoulli system is used to justify the condition-justified Definition 29 of a Bernoulli system.

A special case of this is when for proof-generated definitions preformal ideas shine through (which can be, but does not have to be the case). For instance, consider definitions of polyhedron as discussed by Lakatos (1976). Early definitions of polyhedron, which seem to be justified because they capture the preformal idea of a solid with plane faces and straight edges, were eventually replaced by definitions which are needed to prove the Euler conjecture. For these proof-generated definitions, to some extent, the preformal idea of the old definitions still shine through. Hence Lakatos's (1976, p. 90) claim "In the different proof-generated theorems we have nothing of the naive concept" is an unfortunate exaggeration.

I now return to Lakatos's ideas on justifying definitions.

3.6 Assessment of Lakatos's ideas on proof-generated definitions

First, in focusing on proof-generated definitions, *Lakatos did not recognise the interplay between the different kinds of justification of definitions*, which I discussed in section 3.5. In particular, Lakatos never indicates that in different arguments the same definition can be justified in different ways.

Second, *Lakatos did not show*, as I did for notions of unpredictability in ergodic theory, *that often various kinds of justification are important and that a variety of kinds of justification can be reasonable.* I argued that Lakatos may have believed the following (cf. section 3.2): there are many mathematical subjects where proof-generation should be the sole important way that definitions are justified; and there are many subjects created after mathematicians discovered the method of proof-generation where proof-generation is the sole important way that definitions are justified. From our claim that for nearly all mathematical subjects many different ways of justifying definitions are found and are reasonable, it follows that this must be wrong (cf. subsection 3.4.4). That is, subjects which were created after mathematicians discovered the method of proof-generation where solely proof-generated definitions are found and are reasonable appear to be exceptional.

Indeed, Lakatos could have shown with his case studies that often various kinds of justification are found and that various kinds of justification can be reasonable. To demonstrate this, I will now show that even for the subjects discussed by Lakatos (1976), not only proof-generation but also other kinds of justification are important. To avoid getting the discussion lengthy, I show this here only for the subjects to which the definition of uniform convergence and the Carathéodory definition of measurable sets belong. But this hypothesis can easily seen to be also true for the subjects to which the other proof-justified definitions Lakatos discusses (namely the definitions of polyhedron, bounded variation and the Riemann integral) belong.

Lakatos (1976, pp. 144–146) argues that *uniform convergence* is proof-generated, also by referring to textbooks. This definition falls under the subject of the Mathematical Subject Classification 'convergence and divergence of series and sequences of functions'. A definition discussed in this subject is the radius of convergence of a power series. A power series is of the form $\sum_{k=0}^{\infty} a_k(x - x_0)^k$, where a_k, x_0 and $x \in \mathbb{R}$.

Definition 31 *Its* radius of convergence *is the unique number $R \in [0, \infty]$ such that the series converges absolutely if $|x - x_0| < R$ and diverges if $|x - x_0| > R$.*

The radius of convergence is often defined differently as follows. The *root test* is a powerful criterion for the convergence of infinite series. Hence the question arises whether there is a definition which is equivalent to the radius of convergence as defined above but which gives an explicit way to calculate this radius by referring to the root test. The answer is yes, namely:

Definition 32 *For a power series the* radius of convergence *is*

$$R = 1/\limsup_{k\to\infty} \sqrt[k]{|a_k|}.$$

Thus Definition 32 is condition-justified, as, for example, hinted at in Marsden and Hoffman's (1974, pp. 289–290) standard analysis textbook: "The reason for the terminology in {Definition 32} is brought out by the following result [that by applying the root test, Definition 32 is equivalent to Definition 31]."

Lakatos (1976, pp. 152–154), mainly by referring to Halmos's (1950) book, argues that the *Carathéodory definition of measurable sets* is proof-generated. This definition falls under the subject of the Mathematical Subject Classification 'classes of sets, measurable sets, Suslin sets'. The definition of a σ-algebra clearly belongs to this subject. The basic idea of a σ-algebra is to have a collection of subsets of X including X which is closed under countable set-theoretic operations. Thus a usual definition is (Cohn 1980, pp. 1–2):

Definition 33 *A set Σ of subsets of X is a σ-*algebra *if, and only if,*
(i) $X \in \Sigma$,
(ii) for all $A \subseteq X$ if $A \in \Sigma$, then $X \setminus A \in \Sigma$,
(iii) for all sequences $(A_k)_{k\geq 0}$ if $A_k \in \Sigma$ for all $k \geq 0$, then $\bigcup_{i=0}^{\infty} A_k \in \Sigma$,
(iv) for all sequences $(A_k)_{k\geq 0}$ if $A_k \in \Sigma$ for all $k \geq 0$, then $\bigcap_{i=0}^{\infty} A_k \in \Sigma$.

Now one can easily see that the conditions (i), (ii) and (iii) imply (iv). Consequently, many use the following definition because it eliminates a redundant condition.

Definition 34 *A set Σ of subsets of a set X is a σ-*algebra *if, and only if, (i), (ii) and (iii) hold.*

Clearly, this definition is redundancy-justified as, for instance, in Ash's (1972, p. 4) standard book on measure theory.

To conclude, even for the subjects discussed by Lakatos various kinds of justification are found and are reasonable.

3.7 Conclusion

Mathematical practice suggests that there have to be good reasons for definitions to be worth studying, i.e., mathematical practice suggests that mathematical definitions are justified. And this chapter has addressed the actual practice of how definitions in mathematics are justified in articles and books and whether the justification is reasonable.

After some introductory remarks, in section 3.2 I discussed the main account of these issues, namely Lakatos's ideas on proof-generated definitions. Lakatos claims that in many subjects mathematical definitions are and should be 'proof-generated', by which he means that the definition is needed to prove a specific conjecture regarded as valuable. While important, this chapter has shown how Lakatos's ideas are limited. My assessment of Lakatos and my thoughts on justifying definitions are based on a case study of notions of unpredictability in ergodic theory, which was introduced in section 3.3. In section 4.3 I identified three other important and common ways of justifying definitions: natural-world-justification, condition justification and redundancy justification. A condition-justified definition is a definition which is justified because it is equivalent in a natural way to a previously specified condition regarded as valuable. A redundancy-justified definition is a definition which is justified because it eliminates redundant conditions. To my knowledge, condition justification and redundancy justification have not been discussed so far. Also, I showed that awareness of the ways definitions are justified is important for mathematical understanding and for avoiding mistakes. Then in section 3.5 I discussed the interrelationships between the different kinds of justification of definitions, an issue which has not been addressed before. In particular, I argued that in different arguments the same definition can be justified in different ways. Finally, in section 3.6 I pointed out how Lakatos's ideas are limited. Lakatos did not recognise the interplay between the different kinds of justification. Furthermore, his ideas fail to show that often various kinds of justification are found and that a variety of kinds of justification can be reasonable. I substantiated this claim by showing that even for the subjects Lakatos discusses proof-generation is not the only important kind of justification.

With this background on notions of unpredictability in ergodic theory, we are now ready to tackle one of the key questions about chaos and unpredictability, namely the question of what is the unpredictability which is specific to chaotic behaviour.

Chapter 4

The unpredictability specific to chaos

4.1 Introduction

Since the beginnings of systematically investigating chaos until today, the unpredictability of chaotic systems has been at the centre of interest. There is widespread belief in the philosophy, mathematics and physics communities (and it has been claimed in various articles and books) that *there is a kind of unpredictability specific to chaotic systems, meaning that chaotic systems are unpredictable in a way other deterministic systems are not.* More specifically, what is usually believed is that there is *at least one* kind of unpredictability specific to chaotic systems that is shown by *all* chaotic systems.

The physicist James Lighthill, commenting on the impact of chaos on unpredictability, expresses this point as follows:

> We are all deeply conscious today that the enthusiasm of our forebears for the marvellous achievements of Newtonian mechanics led them to make generalizations in this area of predictability which, indeed, we may have generally tended to believe before 1960, but which we now recognize were false (Lighthill 1986, p. 38).
> These features connected with predictability that I shall describe from now on, then, are characteristic of absolutely all chaotic systems (*Ibid.*, p. 42).

Similarly, Weingartner (1996, p. 50) says that "the new discovery now was that [...] a dynamical system obeying Newton's laws [...] can become chaotic in its behaviour and practically unpredictable".

Thus the question *'What is the unpredictability specific to chaos?'* appears natural, and one might well suppose that it has already been satisfactorily answered. However, this is not the case. On the contrary, there is a lot of confusion about what exactly the unpredictability specific to chaotic behaviour ist. Several answers have been proposed, but, as we will see, none of them fits the bill.

Fundamental questions about the limits of predictability have always been of concern to philosophy. So the widespread belief and the various flawed accounts about the unpredictability specific to chaotic systems demand clarification. The aim of this chapter is to critically discuss existing accounts and to propose a novel and more satisfactory answer.

My answer will be based on two insights. First, I will show that chaos can be defined in terms of strong mixing. Although strong mixing is occasionally mentioned in connection with chaos, I have not found a publication in print arguing that chaos can be thus defined. Second, I will argue that strong mixing has a natural interpretation as a particular form of approximate probabilistic irrelevance which is a form of unpredictability. On this basis, I will propose a general novel answer: a kind of unpredictability specific to chaotic systems is that for predicting any event at any level of precision, all sufficiently past events are approximately probabilistically irrelevant.

The structure of the chapter is as follows. In section 4.2 I will discuss the concepts of unpredictability relevant for this chapter. Section 4.3 will be about chaotic behaviour. Here I will show that chaotic behaviour can be defined in terms of strong mixing. After that, in section 4.4 I will examine the existing answers to the question of what is the unpredictability specific to chaotic systems, and I will dismiss them as mistaken. In section 4.5 I propose a general answer that does not suffer from the shortcomings of the other answers.

4.2 Unpredictability

There are different conceptual accounts of unpredictability for deterministic systems. I will introduce two concepts of unpredictability which will be needed in this chapter.

According to the first concept of unpredictability, a deterministic system is unpredictable when any bundle of initial conditions spreads out more than a specific diameter representing the prediction accuracy of interest (usually of larger diameter than the one of the bundle of initial conditions). When this happens, the deterministic system is unpredictable in the sense that the prediction based on any bundle of initial conditions is so imprecise that it is impossible to determine the outcome of the deterministic system with the desired prediction accuracy.[1] A well-known example is a deterministic system in which, due to exponential divergence of solutions, any bundle of initial conditions of at least a specific diameter spreads out over short time periods more than a diameter of interest.

The second concept of unpredictability is probabilistic. It says that for practical purposes any bundle of initial conditions is irrelevant, i.e., makes it neither more nor less likely that the state is in a region of phase space of interest. According to this concept, it is not only impossible to predict with certainty in which region the deterministic system will be, but in addition, for practical purposes knowledge of the possible initial conditions neither heightens, nor lowers, the probability that the state is in a given region of phase space. An example is that knowledge of any bundle of sufficiently past initial conditions is practically irrelevant for predicting that the state of the deterministic system is in a region of phase space. Eagle (2005, p. 775) defines randomness as a strong form of unpredictability: an event is random if, and only if, the probability of the event conditional on evidence equals the prior probability of the event. This idea relativised to practical purposes is at the heart of our second concept. Consequently, this second concept can also be regarded as a form of randomness.

Clearly, the first and second concepts of unpredictability are different and cannot be expressed in terms of each other since the notions of 'diameter'

[1] Schurz (1996, pp. 133–139) discusses several variants of this form of unpredictability.

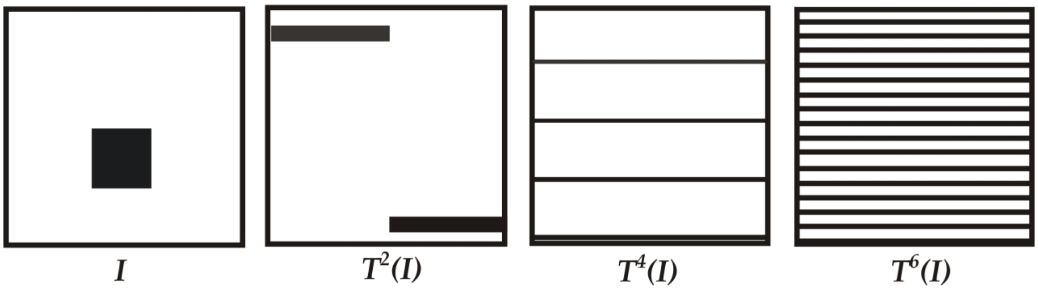

Figure 4.1: evolution of a small bundle of initial conditions I under the baker's system

and 'probability' are not expressible in terms of each other.

4.3 Chaos

4.3.1 Defining chaos

I base the discussion of defining chaos on the following assumption, which is widely accepted in the literature (e.g., Brin & Stuck 2002, p. 23; Devaney 1986, p. 51). A *formal definition of chaos is adequate* if, and only if,

(i) it captures the main *pretheoretic intuitions* about chaos, and

(ii) it is *extensionally correct* (i.e., correctly classifies essentially all systems which, according to the pretheoretic understanding, are uncontroversially chaotic or non-chaotic).

Let us first direct our attention to (i). Roughly, chaotic systems are deterministic systems showing irregular behaviour and sensitive dependence to initial conditions, or even random behaviour. *Sensitive dependence to initial conditions* (SDIC) means that small errors in initial conditions lead to totally different solutions.

Recall the baker's system, our example of a discrete measure-preserving deterministic system (Example 1), and recall a billiard system with convex obstacles, one of our main examples of a continuous measure-preserving deterministic system (Example 2). Figure 4.1 shows the second, forth and sixth iterates of a small bundle of initial conditions I of the baker's system and suggests that any bundle of initial conditions spreads out in phase space. Likewise, Figure 1.1(a) suggests that any bundle of initial conditions of a billiard system with convex obstacles spreads out in phase space (cf. Chapter 1). Thus these deterministic systems appear to exhibit SDIC. Moreover, Figure 4.1 suggests that for the baker's system, and Figure 1.1(b) suggests that for billiard systems with convex obstacles, the motion exhibits irregular behaviour in the following sense: any bundle of initial conditions eventually intersects with any other region in phase space, a property called *denseness*. It is widely agreed that SDIC and denseness are necessary conditions for chaos (Nillsen 1999, pp. 14–15; Peitgen, Jürgens & Saupe 1992, pp. 509–521; Smith 1998, pp. 167–169). This motivates the following criterion: a *definition captures the main pretheoretic intuitions about chaos if, and only if, it implies SDIC and denseness.*

Let us now discuss (ii), the requirement of extensional correctness. Imagine we are concerned with a pretheoretic property P. Further, assume that we are faced with a class of objects some of which uncontroversially have property P, others uncontroversially fail to have property P, and yet others are borderline cases or controversial in some sense. The task is to find an unambiguous definition of P. Then it is natural to say that an unambiguous definition of the property P is extensionally correct if, and only if, it classifies all objects correctly which uncontroversially have or do not have property P. For the borderline objects it is unimportant how they are classified, and I defer to the definition.

Being chaotic is such a property because the pretheoretic idea of chaos is somewhat vague. Among the deterministic systems whose behaviour is mathematically well understood, there is a broad class of uncontroversially chaotic systems and a broad class of uncontroversially non-chaotic systems. Moreover, there are a few borderline cases, for example the system discussed by Martinelli, Dang & Seph (1998, p. 199), where it is not clear whether they are chaotic (Brin & Stuck 2002, p. 23; Robinson 1995, pp. 81–85; Zaslavsky 2005, pp. 53–54). Consequently, I say that a *formal definition of chaos is extensionally correct if, and only if, it correctly classifies essentially all mathematically well understood uncontroversially chaotic and non-chaotic behaviour.*

Several definitions of chaos have been proposed (cf. Lichtenberg & Lieberman 1992, pp. 302–309; Robinson 1995, pp. 81–86). While these definitions are very similar, they are all inequivalent. For want of space I cannot discuss all these definitions here and instead focus on a definition of chaos in terms of strong mixing, which will be crucial later on.

4.3.2 Defining chaos via strong mixing

Recall Definition 27 and Definition 28 of strong mixing (see subsection 3.4.2). Intuitively speaking, the fact that a deterministic system is strongly mixing means that any bundle of solutions spreads out in phase space like a drop of ink in a glass of water.

Strong mixing is occasionally mentioned in connection with chaos, usually only in the context of volume-preserving deterministic systems (e.g., Lichtenberg & Lieberman 1992, pp. 302–303; Schuster & Just 2005, p. 177). Yet, to the best of my knowledge, I have found no publication arguing that chaos can defined in terms of strong mixing. I will argue for this and propose that a possible definition of chaos is in terms of strong mixing: a deterministic system is chaotic if, and only if, it is strongly mixing.

Since strong mixing was introduced before the 1960s, the beginning of the systematic investigation of chaos, it might seem puzzling that chaos can be adequately defined via strong mixing. However, many formal definitions and measures of chaos were invented before the 1960s (Dahan-Dalmedico 2004, p. 70), but rather few deterministic systems were known to which these notions apply. Novel from the 1960s onwards was that many different interesting deterministic systems, surprisingly also very simple systems, were found to which these concepts apply.

Let us first discuss whether strong mixing captures the pretheoretic intuitions. Strong mixing implies denseness: first, strongly mixing discrete measure-preserving deterministic systems are ergodic (Cornfeld et al. 1982, p. 25). By looking at Definition 2.5 of ergodicity, one sees that from this follows that any region, naturally interpreted as a set of positive measure, eventually visits every region in phase space. Second, it is clear that strongly mixing continuous measure-preserving deterministic systems are weakly mixing. And as we have seen in subsection 3.4.2, if a continuous deterministic system (M, Σ_M, μ, T_t) is weakly mixing, then for all $t_0 \in \mathbb{R}^+$, the discrete measure-preserving deterministic system $(M, \Sigma_M, \mu, T_{t_0})$ is ergodic. Hence, again, by looking at Definition 2.5, one sees that also for continuous deterministic systems any region, naturally interpreted as a set of positive measure, eventually visits every region in phase space.

Strong mixing also implies SDIC. This can be seen as follows. Strong mixing implies that any bundle of initial conditions spreads out uniformly over the phase space. Therefore, any bundle eventually spreads out considerably, thus exhibiting SDIC. Formally, assume that a strongly mixing discrete measure-preserving deterministic system (M, Σ_M, μ, T) is given where a metric d is defined on M and Σ_M contains every open set of (M, d). Further, assume that every open set has positive measure.[2] Consider two open sets O_1 and O_2 with $0 < \varepsilon = \inf_{m \in O_1, y \in O_2}\{d(m, y)\}$. Strong mixing implies that for any open set $O \subseteq M$ there is a $t \in \mathbb{N}_0$ such that $T^t(O) \cap O_1 \neq \emptyset$ and $T^t(O) \cap O_2 \neq \emptyset$. But this means that $\varepsilon \leq \sup_{m,y \in T^t(O)}\{d(m, y)\}$. Hence the following condition holds, which in definitions like *Devaney chaos* is taken to be the SDIC implied by discrete chaotic motio (see Devaney 1986, p. 51; Werndl 2009*d*):

$$\begin{gathered}\text{There is an } \varepsilon > 0 \text{ such that for all } m \in M \text{ and for all } \delta > 0 \\ \text{there is a } y \in M \text{ and a } t \in \mathbb{N}_0 \text{ with } d(m, y) < \delta \text{ and } d(T^t(m), T^t(y)) \geq \varepsilon.\end{gathered} \tag{4.1}$$

Likewise, assume that a strongly mixing continuous measure-preserving deterministic system (M, Σ_M, μ, T_t) is given where a metric d is defined on M, Σ_M contains every open set of (M, d) and every open set has positive measure. Again, consider two open sets O_1 and O_2 with $0 < \varepsilon = \inf_{m \in O_1, y \in O_2}\{d(m, y)\}$. Strong mixing implies that for an arbitrary open set $O \subseteq M$ there is a $t \in \mathbb{R}_0^+$ such that $T_t(O) \cap O_1 \neq \emptyset$ and $T_t(O) \cap O_2 \neq \emptyset$. Consequently, $\varepsilon \leq \sup_{m,y \in T_t(O)}\{d(m, y)\}$. Therefore, the following condition holds which is often taken to indicate the SDIC of continuous chaotic motion:

$$\begin{gathered}\text{There is an } \varepsilon > 0 \text{ such that for all } m \in M \text{ and for all } \delta > 0 \\ \text{there is a } y \in M \text{ and a } t \in \mathbb{R}_0^+ \text{ with } d(m, y) < \delta \text{ and } d(T_t(m), T_t(y)) \geq \varepsilon.\end{gathered} \tag{4.2}$$

[2] This is standardly assumed and, to the best of my knowledge, applies to all paradigmatic chaotic systems.

As SDIC is often linked to positive Lyapunov exponents, let us now turn to a discussion of this issue. For a discrete measure-preserving deterministic system (M, Σ_M, μ, T) where $M \subseteq \mathbb{R}$ is an open set and T is continuously differentiable, the *Lyapunov exponent* of $m \in M$ is

$$\lambda(m) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \log(|T'(T^i(m))|), \tag{4.3}$$

where T' is the derivative of T (for a general definition for discrete deterministic systems and for a definition for continuous measure-preserving deterministic system see Mañé 1987, p. 263, and Oseledec 1968). For ergodic deterministic systems the Lyapunov exponent exists and is equal for all points except for a set of measure zero (Oseledec 1968; Robinson 1995, p. 86). Hence one can speak of the Lyapunov exponent of a deterministic system. Accordingly, one definition of chaos that has been suggested is that the deterministic system is ergodic and has a positive Lyapunov exponent.

From a positive Lyapunov exponent it is commonly concluded that the SDIC shown by chaos consists of the exponential spreading of inaccuracies over finite time periods (e.g., Lighthill 1986, p. 46; Ott 2002, p. 140; Smith 1998, p. 15).[3] However, this is mistaken. Positive Lyapunov exponents imply that for almost all points m in phase space the average over all $i \geq 0$ of $\log(|T'(T^i(m))|)$—the exponential growth rate of an inaccuracy at the point $T^i(m)$—is positive. Here the average is taken for the solution starting from m over an *infinite* time period. But positive *on average* exponential growth rates over an *infinite time period* do *not* imply that nearby solutions diverge exponentially or rapidly over *finite time periods.* The growth rate over finite time periods can be anything; inaccuracies can even shrink (Smith, Ziehmann & Fraedrich 1999, pp. 2861–2861).[4] Furthermore, it is not true that inaccuracies of chaotic systems spread exponentially or rapidly over finite time periods: for paradigmatic chaotic systems like the Lorenz system

[3] With the qualification that the time periods have to be small enough such that the inaccuracy does not eventually saturate at the diameter of the deterministic system.

[4] Moreover, Lyapunov exponents only measure the average growth rate of an *infinitesimal* inaccuracy around m, which is defined as the growth rate of a small ball of radius $\varepsilon > 0$ with centre m as $\varepsilon \to 0$; yet in practice the uncertainty is finite and may not behave like the infinitesimal one (cf. Bishop 2008, p. 8).

(Example 3) there are regions where inaccuracies even *shrink* over finite time periods, and numerical evidence suggests such regions for many chaotic systems (Smith et al. 1999, p. 2881; Zaslavsky 2005, p. 315; Ziehmann, Smith & Kurths 1986, pp. 10–11).

Strongly mixing deterministic systems need not have positive Lyapunov exponents, and thus inaccuracies need not grow exponentially on average as time goes to infinity. Is this a problem for strong mixing as a definition of chaos? No. First, there is no agreement in the literature whether chaotic behaviour should show this on average exponential growth. Some definitions do indeed demand it, others such as Devaney chaos do not. Second, the arguments for requiring positive Lyapunov exponents are not convincing. The standard rationale is that the SDIC shown by chaotic system has to be exponential divergence of nearby solutions over finite time periods. But as shown above, this is not implied by a positive Lyapunov exponent and also does not generally hold for chaotic systems. Another possible argument is that for chaotic behaviour inaccuracies should spread out rapidly. Yet the rate of divergence of strongly mixing deterministic systems not having positive Lyapunov exponents can be much faster for arbitrary long time periods than for systems with positive Lyapunov exponents; thus it is not clear why positive Lyapunov exponents should be required (Berkovitz et al. 2006, p. 689; Wiggins 1990, p. 615). To conclude, *strong mixing captures the pretheoretic intuitions about chaos.* It remains to show that the definition of chaos in terms of strong mixing is extensionally correct.

To do this, I have to consider the main classes of uncontroversially chaotic and non-chaotic behaviour.[5] I start with uncontroversially chaotic behaviour and first discuss volume-preserving deterministic systems. There are (i) Hamiltonian system which are chaotic on the whole hypersurface of constant energy. Three types of continuous measure-preserving deterministic systems are mainly discussed here: first, chaotic billiards, such as billiards with convex obstacles (Example 2), which are strongly mixing (Chernov & Markarian 2006; Ott 2002, p. 296); second, hard sphere systems, which describe the motion of a number of hard spheres undergoing elastic reflections

[5]Obviously, I cannot discuss every single deterministic system regarded as clearly chaotic or non-chaotic. Yet our discussion covers all main examples.

at the boundary and collisions amongst each other; e.g., the motion of N hard balls on the m torus for $N \geq 2$ and $m \geq N$; hard-sphere systems are important in statistical mechanics because they are a model of the ideal gas, and they are either proven or conjectured to be strongly mixing (Berkovitz et al. 2006, pp. 679–680; Ornstein & Weiss 1974, pp. 8–9; see also Sźasz 2000); third, geodesic flows of space with negative Gaussian curvature, i.e., frictionless motion of a particle moving with unit speed on a compact manifold with everywhere negative curvature, are strongly mixing too (Schuster & Just 2005, p. 181).

Another class are (ii) Hamiltonian systems to which the KAM-theorem applies, e.g., the Hénon-Heiles system or the standard map. This class includes simplified versions of Poincaré maps of continuous measure-preserving deterministic systems to which the KAM-theorem applies. The KAM-theorem describes what happens when integrable systems are perturbed by a nonintegrable perturbation. It says that tori with sufficiently irrational winding number survive the perturbation. Between the stable motion on surviving tori there appear to be regions of unpredictable motion. As the perturbation increases, these regions become larger and often eventually cover nearly the entire hypersurface of constant energy.

For these deterministic systems the phase space is separated into regions, each of which has its own dynamics: in some of them the motion appears unpredictable and in others it is stable. Because of this separation into regions, unpredictable behaviour can only be found in a region. Consequently, as is widely acknowledged, proper chaotic motion can only occur on a region (Ott 2002, pp. 267–295; Schuster & Just 2005, pp. 165–174). Thus I have to show that the mathematically well-understood unpredictable motion in a region is strongly mixing. Yet the conjectured chaotic motion of KAM-type systems is understood only poorly (Zaslavsky 2005, p. 139). It has only been proven that there is chaotic behaviour near hyperbolic fixed points, where the motion is indeed strongly mixing (Moser 1973, chapter 3). Apart from this, some numerical evidence suggests that the motion conjectured to be chaotic is strongly mixing (e.g., Chirikov 1979). Thus Lichtenberg & Lieberman (1992, p. 303) comment that we "expect that the stochastic orbits that we have encountered in previous sections are strongly mixing over the bounded portion of phase space for which they exist".

I should mention that numerical experiments suggest that for a few KAM-type maps there are sets on which the motion seems somewhat random, but these sets consist of $n \geq 2$ component areas, each of which is mapped successively on to another, returning to itself after n iterations. There is no agreement whether such motion, which cannot be strongly mixing, should be called 'chaotic' (e.g., Belot & Earman 1997, p. 154, vs. Ott 2002, p. 300). If it is, chaos can still be defined via strong mixing: one can say that a deterministic system is chaotic if, and only if, it is ergodic (cf. Definition 2.5) and its phase space is decomposable into $n \geq 1$ sets with disjoint interior such that the n-th iterate is strongly mixing on each of these sets. I call this the *'broad definition of chaos via strong mixing'*. There are numerical experiments which suggest that the behaviour mentioned above is chaotic according to this definition (Ott 2002, p. 303).

Next in line are (iii) chaotic volume-preserving non-Hamiltonian systems. Here the main examples discussed are discrete. First, the baker's system (Example 1) and volume-preserving Anosov diffeomorphisms such as the cat map are strongly mixing (Arnold & Avez 1968, p. 75; Lichtenberg & Lieberman 1992, p. 303). Second, paradigmatic chaotic systems are expanding piecewise maps such as the tent map, which are strongly mixing too (Bowen 1977).

I now turn to dissipative systems and first discuss strange attractors. One class are (iv) strange attractors where the attracted solutions never enter the attractor. Three main groups are treated here: first, for Smale's Solenoid and generalised Solenoid systems there is a measure on which the motion is strongly mixing (Mayer & Roepstorff 1983). Second, for the Lorenz system investigated by (Lorenz 1963) (see Example 3) and the Lorenz model, and generalised versions thereof, which have been used to model weather phenomena and waterwheels, there is a physical measure on which the motion is strongly mixing (see the end of section 2.1 for a discussion of physical measures) (Luzzatto et al. 2005). Third, for generalised Hénon systems such as the Hénon map, which has been proposed as a simple model of weather dynamics, there exists a physical measure such that the motion on the attractor is strongly mixing (Benedicks & Young 1993, Hénon 1976).

Also important is the (v) visible chaotic behaviour of generalised versions of the logistic map; the logistic map has been endorsed as a simplified model of population dynamics and climate dynamics (Lorenz 1964; Lyubich 2002; May 1976). For these measure-preserving deterministic systems for most parameter values the solutions enter an attractor with a physical measure on which the motion is either strongly mixing or chaotic according to the broad definition via strongly mixing. But for a few parameter values there is chaotic behaviour on an entire interval; in these cases there is also a physical measure on which the motion is strongly mixing (Jacobson 1981; Lyubich 2002).

Finally, another class of uncontroversially chaotic behaviour is (vi) repelling chaotic behaviour on Cantor sets. Two main kinds of discrete deterministic systems are discussed here: first, geometric horseshoe-systems such as Smale's horseshoe, which are strongly mixing (Robinson 1995, pp. 249–274). The second example is chaotic motion on Cantor sets for the logistic map with parameter greater than 4, which is also strongly mixing (Robinson 1995, p. 33).[6]

Let us now turn to uncontroversially non-chaotic motion. I again start with volume-preserving deterministic systems. A paradigmatic class are (i) integrable Hamiltonian systems, where there is periodic or quasi-periodic motion on tori, which is not strongly mixing (Arnold & Avez 1968, pp. 210–214).

Another class is the (ii) motion on clearly non-chaotic regions of KAM-type systems. Again, this class includes simplified versions of Poincaré maps of KAM-type deterministic systems. As already discussed, for KAM-type systems the phase space is separated into regions, and on some regions the motion is stable. Thus I have to show that the stable motion is not strongly mixing. And indeed, the behaviour in these regions, e.g., the motion on surviving tori or the one near specific elliptic periodic points, is not strongly mixing (Arnold & Avez 1968, pp. 86–90; Lichtenberg & Lieberman 1992, chapter 3–5).

I now turn to dissipative measure-preserving deterministic systems. Important here are (iii) non-chaotic attractors. These are attracting periodic cycles and fixed points and also quasi-periodic attractors as discussed by Ott

[6] This follows because these deterministic systems are isomorphic to a Bernoulli shift.

(2002, chapter 7), which obviously cannot be strongly mixing. Moreover, the motion approaching such attractors, e.g., the behaviour around stable nodes or stable foci, clearly cannot be strongly mixing (cf. Robinson 1995, p. 105).[7]

Finally, let us mention two further very broad classes of clearly non-chaotic deterministic systems. Since strong mixing captures SDIC, (iv) systems not exhibiting any kind of SDIC, e.g., the identity function, cannot be strongly mixing.

Moreover, since strong mixing captures denseness, (v) motion showing SDIC but where, in any sense, typical solutions do not come arbitrarily near to any region in phase space cannot be strongly mixing. Examples here are discrete-time deterministic systems where the evolution function is $T(m) = cm$ for $c > 1$ on $(0, \infty)$ or the motion around unstable nodes or unstable foci (cf. Robinson 1995, p. 105).[7]

In sum, I have first demonstrated that strong mixing captures the pretheoretic intuitions about chaos. After that I have briefly shown that a definition of chaos in terms of strong mixing is extensionally correct in the sense explained above. Consequently, *chaos can be adequately defined in terms of strong mixing.*

With this knowledge about chaos we are ready to critically discuss the answers suggested in the literature to our main question.

4.4 Criticism of answers in the literature

4.4.1 Asymptotically unpredictable?

Let us first discuss an answer based on the concept of asymptotic unpredictability. Roughly, systems whose asymptotic behaviour cannot be predicted with arbitrary accuracy for all times, even if the bundle of initial conditions is made arbitrarily small, are said to be asymptotically unpredictable. Formally, given a topological deterministic system, let ε be the desired prediction accuracy and let δ be the diameter of the bundle of initial conditions. For a discrete topological deterministic system (M, d, T) and an $m \in M$ the

[7]Here there sometimes exists no invariant measure of interest.

solution s_m is *asymptotically predictable* if, and only if,

$$\forall \varepsilon > 0\ \exists \delta > 0\ \forall y \in M\ \forall t \in \mathbb{N}_0\ \ (d(m, y) < \delta \rightarrow d(T^t(m), T^t(y)) < \varepsilon). \tag{4.4}$$

The discrete topological deterministic system (M, d, T) is *asymptotically unpredictable* if, and only if, for all $m \in M$ the solution s_m is not asymptotically predictable.[8] Likewise, for a continuous topological deterministic system (M, d, T_t) and an arbitrary $m \in M$ the solution s_m is *asymptotically predictable* if, and only if,

$$\forall \varepsilon > 0\ \exists \delta > 0\ \forall y \in M\ \forall t \in \mathbb{R}_0^+\ \ (d(m, y) < \delta \rightarrow d(T_t(m), T_t(y)) < \varepsilon). \tag{4.5}$$

The continuous topological deterministic system (M, d, T) is *asymptotically unpredictable* if, and only if, for all $m \in M$ the solution s_m is not asymptotically predictable. In terms of the distinction introduced in section 4.2, this is clearly a version of the first concept of unpredictability.

Miller (1996, pp. 106–107) and Stone (1989, p. 127) argue that the *unpredictability specific to chaotic systems is that chaotic systems are asymptotically unpredictable*. Indeed, all chaotic systems discussed in the literature are asymptotically unpredictable, and standard definitions of chaos imply asymptotic unpredictability. For instance, (4.1) and (4.2), a condition of Devaney chaos and, as we have seen, a consequence of strong mixing, clearly implies asymptotic unpredictability.

However, as Smith (1998, p. 58) has pointed out, many non-chaotic deterministic systems, e.g., one only showing SDIC as it happens for the evolution function $T(m) = cm$ for $c > 1$ on $(0, \infty)$ (class (v) of clearly non-chaotic behaviour), are asymptotically unpredictable. Hence this answer is *wrong*. But maybe the account can be strengthened in the following way: the *the unpredictability specific to chaotic systems is that they are asymptotically unpredictable and bounded*. I maintain that this is *not correct* either: there are unbounded chaotic systems (Smith 1998, pp. 168–169), a point which is reflected in usual definitions of chaos, which do not require boundedness. Furthermore, for many bounded integrable systems (part of class (i) of the

[8]Bishop (2003, pp. 174–177) also aims to formalise asymptotic unpredictability. However, he does not list the most obvious notion presented here.

clearly non-chaotic behaviour) the solutions loop around tori in such a way that they are asymptotically unpredictable (Arnold & Avez 1968, pp. 210–214). Hence there are examples of non-chaotic, bounded and asymptotically unpredictable deterministic systems.

I conclude that the sole connection between asymptotic unpredictability and chaos is this: while only some non-chaotic deterministic systems are asymptotically unpredictable, every chaotic system is asymptotically unpredictable.

4.4.2 Unpredictable due to rapid or exponential divergence of solutions?

It is widely believed and often claimed that the unpredictability specific to chaotic systems is the following: *due to rapid or exponential divergence of nearby solutions, bundles of initial conditions spread out a distance more than a diameter of interest over short time periods* (e.g., Ruelle 1997, pp. 27–28); *often it is added that this is so despite the fact that the deterministic systems are bounded* (e.g., Lighthill 1986, p. 46). In terms of the distinction introduced in section 4.2, this is a form of the first concept of unpredictability.

As many unbounded non-chaotic deterministic systems, such as a discrete deterministic systems with evolution function $T(m) = cm$, $c > 1$, on $(0, \infty)$ show (part of class (v) of clearly non-chaotic behaviour), rapid or exponentially divergence everywhere is 'nothing new' (Smith 1998, p. 15). Thus the version not requiring boundedness *cannot be true*. But also the version requiring boundedness is *wrong*: as mentioned above, there are unbounded chaotic systems. Furthermore, as argued in section 4.3, it is often *not* true that nearby solutions of chaotic systems diverge rapidly or exponentially over finite time periods as is so widely believed in the philosophy, physics and mathematics communities (e.g., Eagle 2005, p. 767; Schurz 1996, p. 140; Smith 1998, p. 15). Hence this is not the sought-after unpredictability specific to chaotic systems.

Why is it so widely believed that inaccuracies of chaotic systems spread rapidly or exponentially over finite time periods? One plausible reason is that because very simple chaotic systems such as the baker's system (Example 1) or the cat map show this property, this claim is wrongly generalized to all chaotic systems. Also, the wrong belief stems at least in part from misinterpreting Lyapunov exponents. As pointed out in section 4.3, positive on average exponential growth rates over an infinite time period are wrongly taken to imply that inaccuracies spread exponentially over relatively short finite time periods.

The only connection between the unpredictability of chaotic systems and the rapid or exponential increase of inaccuracies over finite time periods seems to be this: it is more often the case for chaotic than for non-chaotic deterministic systems that bundles of initial conditions spread out more than a diameter of interest over short time periods.

4.4.3 Macro-predictable and micro-unpredictable?

Macro-predictable yet micro-unpredictable behaviour is a broad and interesting topic in physics. For instance, in statistical mechanics deterministic systems are often macro-predictable but micro-unpredictable. Here I concentrate only on whether there is any combination of macro-predictability and micro-unpredictability in chaotic systems that other deterministic systems do not have.

To gain an understanding of this proposed answer, recall the Lorenz system (Example 3 and Figure 2.2). This system exhibits macro-predictability: the solutions are attracted by an attractor, a small region of phase space. There is also micro-unpredictability since the motion on the attractor exhibits SDIC. Smith (1998) argues that this *combination of macro-predictability and micro-unpredictability is a kind of unpredictability specific to chaotic systems*:

> *This type of combination of large-scale order with small scale disorder, of macro-predictability with the micro-unpredictability due to sensitive dependence, is one paradigm of what has come to be called chaos.* [...] So error inflation by itself is entirely old-hat. The novelty in the newfangled chaotic cases that will concern us is, to repeat, the *combination* of exponential error inflation with the tight confinement of trajectories by an attractor (Smith 1998, pp. 13–15, original emphasis).

Here macro-predictability means that the deterministic system eventually shows the behaviour corresponding to the motion on the attractor, a proper subset of phase space. Micro-unpredictability is understood as the unpredictability implied by exponential error inflation. Yet, as shown in section 4.3, solutions of chaotic systems need not diverge exponentially or rapidly over finite time periods. Therefore, micro-unpredictability has to be interpreted as a weaker notion, e.g., asymptotic unpredictability (cf. subsection 4.4.1).

As becomes clear from the Lorenz system (Example 3), strange attractors imply this combination of macro-predictability and micro-unpredictability. However, this combination is *no kind of unpredictability which is specific to chaotic systems* since there are many chaotic systems without attractors. As already pointed out, all chaotic volume-preserving deterministic systems such as chaotic Hamiltonian systems or the baker's system (classes (i), (ii) and (iii) of uncontroversially chaotic behaviour) cannot have attractors. And some chaotic dissipative systems, e.g., repelling chaotic motion on Cantor sets or the logistic map on $[0, 1]$ (class (vi) and a part of class (v) of uncontroversially chaotic behaviour), have no attractors. Hence these deterministic systems are *not* macro-predictable in the above sense, viz. that appeals to attractors.

It could be that Smith (1998) only meant to say that this combination of macro-predictability and micro-unpredictability found in strange attractors *is a novelty for deterministic systems with attractors*. But this would *not* help. Clearly, this claim would be no satisfying answer to our main question because it does not apply to essentially all chaotic systems. Furthermore, also non-chaotic deterministic systems can be macro-predictable and micro-unpredictable as discussed here. For instance, in the plane let R be the region enclosed by a circle of radius r around the origin (boundary included). Imagine that all solutions in R go in circles around the origin and that all solutions outside R are attracted by the periodic motion in R such that all solutions are continuous. Such non-chaotic attractors (part of class (iii) of clearly non-chaotic behaviour) obviously imply macro-predictability and micro-unpredictability. Thus this combination of macro-predictability and micro-unpredictability is not even a kind of unpredictability specific to deterministic systems with attractors.

Of course, there are also other concepts of macro-predictability and micro-unpredictability (e.g., Smith 1998, pp. 60–61). However, to the best of my knowledge, none of them provides a combination of macro-predictability and micro-unpredictability that is characteristic of chaotic behaviour.

To conclude, strange attractors are macro-predictable and micro-unpredictable in the above specified sense. However, it is not the case that a combination of macro-predictability and micro-unpredictability constitutes a kind of unpredictability specific to chaotic behaviour.

None of the answers examined so far have proven to be correct. There is one more answer suggested in the literature: some physicists, e.g., Ford (1989), have defined chaos by the condition that almost all solutions have positive algorithmic complexity. In other words, they have argued that the unpredictability implied by positive algorithmic complexity is specific to chaotic systems. However, Batterman & White (1996) and Smith (1998, p. 160) have made it clear that chaos cannot be defined via algorithmic complexity since many deterministic systems without SDIC (part of class (iv) of clearly non-chaotic behaviour) have positive algorithmic complexity too. Consequently, this is *not* a kind of unpredictability which is specific to chaotic behaviour, and this is all we need to know.

In sum, the answers in the literature do not fit the bill.

4.5 A kind of unpredictability specific to chaos

4.5.1 Approximate probabilistic irrelevance

The answer I propose starts from the idea that strong mixing goes along with loss of information as recently discussed by Berkovitz et al. (2006). First of all, let us introduce the approximate probabilistic irrelevance, the notion of unpredictability which will be crucial for our claim.

Recall the definition of an event and the definition of a probability of an event as introduced when discussion weak mixing in subsection 3.4.1 (see also Berkovitz et al. 2006, pp. 670–672; Werndl 2009*e*): given a discrete measure-preserving deterministic system (M, Σ_M, μ, T) or a continuous measure-preserving system (M, Σ_M, μ, T_t), A^t is defined as the event that the state of the deterministic system is in A at time t, $A \in \Sigma_M$ arbitrary,

$t \in \mathbb{Z}$ or $\mathbb{R}$. And $p(A^t)$ is the probability that the event A^t obtains. Let me introduce conditional probabilities: $p(B^{t'} \mid A^t)$, for arbitrary A, $B \in \Sigma_M$ with $\mu(A) > 0$, is the probability that P_B obtains at time t' given that P_A obtains at time t. By the usual definition, $p(B^{t'} \mid A^t) = p(B^{t'} \& A^t)/p(A^t)$. Because the measure is interpreted as probability density, the probability of events is given by the equations (3.1), (3.2) and (3.3) (see subsection 3.4.1).

Now recall the second conception of unpredictability of section 4.2. For this conception I have to say what it means that knowledge that the deterministic system is in a region A at t is practically irrelevant for predicting that it will be in region B at t'. I say that this is so if the probability of the event $B^{t'}$ given knowledge of the event A^t approximately equals the unconditionalised probability of the event $B^{t'}$. Let $\varepsilon > 0$ be the level at which probabilities differing by less than ε are considered as practically equivalent. Further, assume that $p(A^t) > 0$; I will later explain why I am justified to do so. Then formally this is captured by the following definition:[9]

$$A^t \,\text{is}\, \textit{approximately probabilistically irrelevant}\, \text{for predicting}\, B^{t'} \quad (4.6)$$
$$(t, t' \in \mathbb{Z}\,\text{or}\,\mathbb{R})\; \textit{at level}\, \varepsilon > 0\,\text{if, and only if,}\, |p(B^{t'} \mid A^t) - p(B^{t'})| < \varepsilon.$$

Or equivalently, but simpler (still assuming that $p(A^t) > 0$):

$$A^t \,\text{is}\, \textit{approximately probabilistically irrelevant}\, \text{for predicting}\, B^{t'} \quad (4.7)$$
$$(t, t' \in \mathbb{Z}\,\text{or}\,\mathbb{R})\; \textit{at level}\, \varepsilon > 0\,\text{if, and only if,}\, |p(B^{t'} \& A^t) - p(B^{t'})p(A^t)| < \varepsilon.$$

In the next section we will see how the approximate probabilistic irrelevance relates to chaos, and I will finally propose an answer to our question.

[9] I use what is basically the difference measure in confirmation theory to define the approximate probabilistic irrelevance. I should point out that my claims are independent of the measure involved, i.e., they would remain the same if I used any other measure with the indisputable property that it is continuous when the unpredictability is highest, i.e., when $p(B^{t'} \mid A^t) = p(B^{t'})$. Berkovitz et al. (2006, p. 672) interpret the difference measure of events as a general measure of unpredictability. However, they do not justify this choice or address whether their results are independent of the measure.

4.5.2 Sufficiently past events are approximately probabilistically irrelevant for predictions

The argument I put forward to answer the main question of the chapter is as follows. (P1) *Chaos can be defined in terms of strong mixing.* (P2) *Strongly mixing deterministic systems exhibit a particular pattern of approximate probabilistic irrelevance, which constitutes a form of unpredictability.* Therefore: (C) *a kind of unpredictability specific to chaotic systems is the particular pattern of approximate probabilistic irrelevance arising from strong mixing.*

In subsection 4.3.2 we have seen that premise (P1) is true. Let us now argue for premise (P2). Recall the definition of strong mixing (Definition 27 and Definition 28). I assume without loss of generality that the event we want to predict occurs at time 0. Then, assuming (3.1) and (3.3), it follows that a discrete measure-preserving deterministic system (M, Σ_M, μ, T) or a continuous measure-preserving deterministic system (M, Σ_M, μ, T_t) is strongly mixing if, and only if,

$$\lim_{t\to\infty} p(B^0 \& A^{-t}) - p(B^0)p(A^{-t}) = 0, \tag{4.8}$$

for all $A, B \in \Sigma_M$ with $\mu(A) > 0$. This equation holds for all, i.e., discrete and continuous measure-preserving deterministic systems. Berkovitz et al. (2006, p. 676) also show (4.8), but they interpret their results as applying only to Hamiltonian deterministic systems. Many chaotic systems, e.g., all strange attractors (classes (iv) and (v) of uncontroversially chaotic behaviour), are not Hamiltonian. Since I am interested in the unpredictability implied by chaos, it is important to realise that (4.8) holds for all deterministic systems.

From the definition of the limit, I obtain that (4.8) can be expressed as:

$$\begin{gathered}\text{For any event } B^0,\ \text{any } \varepsilon > 0 \text{ and any } A \in \Sigma_M,\ \mu(A) > 0, \text{ there is}\\ t' \in \mathbb{N}\,\text{or}\,\mathbb{R}_0^+ \text{ such that for all}\, t \geq t' : |p(B^0 \& A^{-t}) - p(B^0)p(A^{-t})| < \varepsilon.\end{gathered} \tag{4.9}$$

Hence strong mixing means that for predicting an arbitrary event at an arbitrary level of precision $\varepsilon > 0$, any sufficiently past event is approximately probabilistically irrelevant. Notice that due to the impossibility of determining initial conditions precisely, scientists always consider regions of phase space corresponding to possible initial conditions. Since these regions are not of measure zero, I am justified assuming that $\mu(A) > 0$. In terms of the distinction introduced in section 4.2, this pattern of probabilistic irrelevance is a version of the second concept of unpredictability. Hence *strongly mixing measure-preserving deterministic systems exhibit a particular pattern of approximate probabilistic irrelevance, which constitutes a form of unpredictability*: i.e., premise (P2) is true.[10]

Now that I have argued for the premises (P1) and (P2) of the above argument, I conclude: (C) *a general kind of unpredictability specific to chaotic systems is that for predicting any event at any level of precision $\varepsilon > 0$, all sufficiently past events are approximately probabilistically irrelevant.*

To fully understand this conclusion, consider the following: for strange attractors this claim applies in a strict sense only to events on the attractor. Yet for practical matters there is chaotic behaviour when solutions are very near to the strange attractor (cf. section 2.1); then my claim means that for predicting any event *on or very near the attractor* Λ at any level of precision $\varepsilon > 0$, all sufficiently past events *in the basin of attraction* $U \supset \Lambda$ are approximately probabilistically irrelevant. For KAM-type systems my claim applies, as one would like it, to each chaotic region. Moreover, as explained in subsection 4.3.2 in discussing the uncontroversially chaotic behaviour, some may want to adopt the broad definition of chaos via strong mixing, i.e., that the measure-preserving deterministic system is ergodic and its phase space is decomposable into $n \geq 1$ regions with disjoint interior such that the n-th iterate is strongly mixing on each set. When $n > 1$, my claim (C) has

[10] This claim can be generalised. The discrete measure-preserving deterministic system (M, Σ_M, μ, T) or the continuous measure-preserving deterministic system (M, Σ_M, μ, T_t) is strongly mixing if, and only if, for any probability measure ρ absolutely continuous with respect to μ and any square integrable function $f \in L^2(M, \Sigma_M, \mu)$:

$$\lim_{t\to\infty} \int f(m) d\rho_t = \int f(m) d\mu, \tag{4.10}$$

where ρ_t is the evolved measure after t units of time ($t \in \mathbb{Z}$ or $\mathbb{R}$). Interpret μ as probability and ρ as measuring our knowledge of the initial condition. Then, assuming absolute continuity of ρ, strong mixing means that for arbitrary knowledge of the initial condition after a sufficiently long time the prediction obtained by evolving the measure is practically no better than if we had no knowledge whatsoever of the initial conditions (cf. Berger 2001, pp. 126–132).

to be adapted in the following way: the unpredictability of strong mixing applies to the n-th iterate on the region of interest. This means that for predicting any event *in the region of interest* at any level of precision $\varepsilon > 0$, all sufficiently past events that *could have evolved to the region of interest* are approximately probabilistically irrelevant.

On the one hand, the unpredictability involved in my answer is strong: sufficiently distant events are *practically as probabilistically independent as coin tosses*. On the other hand, it is weak since only *sufficiently* past measurements are approximately probabilistically irrelevant. Restricting my claim to sufficiently past events is essential: first, many chaotic systems are continuous, and continuity makes it impossible that for all past times, all events are approximately probabilistically irrelevant for predictions. Second, we have seen that to require rapid divergence of nearby solutions for chaotic behaviour is untenable.

What is novel about my claim? Granted, in a few publications on chaos the notion of 'irrelevance' is discussed. In fact, there are two main foci; but none give my claim. First, there is Berkovitz et al.'s (2006) explication of the ergodic hierarchy. Yet recall our main argument (cf. the beginning of this subsection). As pointed out, Berkovitz et al. interpret their results as only applying to Hamiltonian systems. Hence they do not argue for the general premise (P2), and, most importantly, they do not argue for the premise (P1). Therefore, they could not arrive at the conclusion (C). Second, sometimes it is asserted that for chaos the input is irrelevant in the sense that prediction is exponentially expensive in the initial data, meaning that for an input string of length n all information is lost after n steps, at which point we are totally unsure what happens next (Leiber 1998, p. 361; Smith 1998, p. 53). However, as argued in subsection 4.4.2, predictions for chaotic systems need not be exponentially expensive in the initial data; the irrelevance shown by chaotic systems is more subtle.

4.6 Conclusion

The unpredictability of chaotic systems is one of the issues that has attracted most interest in chaos research. Nonetheless, nearly half a century after the start of the systematic investigation of chaos, there has been much confusion about, and no correct answer to, the question: what is the unpredictability specific to chaos? I have tackled this question in this chapter.

After some introductory remarks, in section 4.2 I introduced two conceptual accounts of unpredictability relevant for the discussion. After that, in section 4.3 I showed that chaos can be defined in terms of strong mixing, i.e., that strong mixing captures the main pretheoretic intuitions about chaos and correctly classifies the various classes of uncontroversially chaotic and non-chaotic behaviour. This has never been explicitly argued for in the literature. Then, in section 4.4 I criticised the answers in the literature to the above question. First, I rejected the answer that chaotic systems are asymptotically unpredictable on the grounds that also many non-chaotic deterministic systems are asymptotically unpredictable. Second, I rejected the answer that chaotic systems are unpredictable in the sense of exponential or rapid divergence of nearby solutions (often claimed with the added condition of boundedness). For, when not requiring boundedness, many non-chaotic deterministic systems are also unpredictable in this sense. Furthermore, in the case of requiring boundedness, there are unbounded chaotic systems and, though unacknowledged in the philosophy literature, chaotic systems need not be unpredictable in the sense of having exponential or rapid divergence of solutions. Third, I dismissed the answer that chaotic systems show a specific combination of macro-predictability and micro-unpredictability: there are chaotic systems which are not macro-predictable and non-chaotic systems which also show this combination of macro-predictability and micro-unpredictability. This prompted the search for an alternative answer. In section 4.5, based on defining chaos via strongly mixing, I proposed a novel general answer: a kind of unpredictability specific to chaotic systems is that for predicting any event at any level of precision $\varepsilon > 0$ all sufficiently past events are approximately probabilistically irrelevant. Chaotic behaviour is multi-faceted and takes various forms. Yet if the aim is to identify a general kind of unpredictability specific to chaotic systems, I think this is the best we can get.

In this and the previous chapter we have seen that deterministic systems can be unpredictable and even random. This begs the question of whether measure-theoretic deterministic descriptions and indeterministic descriptions can be observationally equivalent. Let us embark on this question in the next chapter.

Chapter 5

Determinism vs. indeterminism: are deterministic and indeterministic descriptions observationally equivalent?

5.1 Introduction

There has been a lot of philosophical debate about the question of whether the world is deterministic or indeterministic. Within this context, there is often the implicit belief (cf. Weingartner & Schurz 1996, p. 203) that deterministic and indeterministic descriptions are not observationally equivalent. However, the question of whether these descriptions are observationally equivalent has hardly been discussed.

This chapter aims to contribute to fill this gap. Namely, the central questions of this chapter are the following: *are deterministic mathematical descriptions and indeterministic mathematical descriptions observationally equivalent? And what is the philosophical significance of the various results on observational equivalence?* The deterministic and indeterministic descriptions of concern in this chapter are measure-theoretic deterministic systems and stochastic processes, respectively, both of which are ubiquitous in science.

More specifically, by saying that a measure-theoretic deterministic system and a stochastic process are *observationally equivalent*, I will mean the following: the deterministic system, when observed, gives the same predictions as the stochastic process. And when I say that a stochastic process can be *simulated* by a measure-theoretic deterministic system, or conversely, I will mean that it can be simulated by such a deterministic system in the sense that they are observationally equivalent.

This chapter proceeds as follows. In section 5.2 I will show that measure-theoretic deterministic systems and stochastic processes can often be simulated by each other. Despite this, one might guess that it is impossible to simulate stochastic processes of the kinds in fact used in science by measure-theoretic deterministic systems that are used in science. I will show in section 5.3 that this guess is wrong. Given this, one might still guess that it is impossible to simulate measure-theoretic deterministic systems of the kinds in fact used in science at every observation level by stochastic processes that are used in science. By proving some results in ergodic theory, I will show in section 5.4 that this guess is also wrong. Therefore, even stochastic processes and measure-theoretic deterministic system which, intuitively, seem to give very different predictions, are in fact observationally equivalent. Finally, in section 5.5 I will criticise the claims of the previous philosophical papers Suppes (1993), Suppes & de Barros (1996), Suppes (1999) and Winnie (1998) on observational equivalence. Then, in section 5.6 I will summarise my results.

5.2 Basic observational equivalence

I will first discuss some results about observational equivalence which are basic in the sense that they are about the question whether, given a measure-theoretic deterministic system, it is possible to find *any* stochastic process which is observationally equivalent to the measure-theoretic deterministic system, and conversely.

How can a stochastic process and a measure-theoretic deterministic system yield the same predictions? When a measure-theoretic deterministic system is observed, one only sees how one observed value follows the next observed value. Because the observation function can map two or more actual states to the same observed value, the same present observed value can lead to different future observed values. And so a stochastic process can be observationally equivalent to a measure-theoretic deterministic system only if it is assumed that the deterministic system is observed with an observation function which is many to one. Yet this assumption is usually unproblematic: the main reason being perhaps that measure-theoretic deterministic systems used in science typically have an infinitely large phase space, and scientists can only observe finitely many different values.

A probability measure is defined on a measure-theoretic deterministic system. Hence the predictions derived from a deterministic system are the probability distributions over sequences of possible observations. And similarly, the predictions obtained from a stochastic process are the probability distributions over sequences of possible outcomes. Consequently, the most natural meaning of the phrase *'a stochastic process and a measure-theoretic deterministic system are observationally equivalent'* is: *(i) the set of possible outcomes of the stochastic process is identical to the set of possible observed values of the deterministic system*[1]*, and (ii) the realisations of the stochastic process and the solutions of the deterministic system coarse-grained by the observation function have the same probability distribution.*

Let me now investigate when deterministic systems can be simulated by stochastic processes. Then I will investigate when stochastic processes can be simulated by deterministic systems.

5.2.1 Deterministic systems simulated by stochastic processes

Let (M, Σ_M, μ, T) be a discrete measure-theoretic deterministic system. According to the canonical Definition 8, $Z_t(m) = T^t(m)$ is a discrete stochastic process with exactly the same predictions as the discrete deterministic system. Likewise, given a continuous deterministic system (M, Σ_M, μ, T_t),

[1]From a probabilistic viewpoint outcomes with probability zero or observed values with probability zero are irrelevant. Hence, more precisely, condition (i) is: the set of possible outcomes with positive probability is identical to the set of possible observed values with positive probability.

according to the canonical definition Definition 9, $Z_t(m) = T_t(m)$ is a continuous stochastic process with exactly the same predictions as the continuous deterministic system. However, these processes are evidently equivalent to the original deterministic system, and the *transition probabilities*, i.e., the probabilities that one outcome leads to another one, are trivial (0 or 1). Hence they are still really deterministic systems. So this is the mathematical formalisation of the idea known in the philosophy literature that a deterministic system is the special case of a stochastic process where all probabilities are zero or one (cf. Butterfield 2005, Earman 1986).

But one can do better by appealing to observation functions as explained above; and, to my knowledge, these results are unknown in philosophy. Assume the discrete measure-theoretic deterministic system (M, Σ_M, μ, T) is observed with an observation function $\Phi : M \to M_O$. Then $\{Z_t = \Phi(T^t);\ t \in \mathbb{Z}\}$ is a discrete stochastic process. Likewise, assume the continuous measure-theoretic deterministic system (M, Σ_M, μ, T_t) is observed with an observation function $\Phi : M \to M_O$. Then $\{Z_t = \Phi(T_t);\ t \in \mathbb{R}\}$ is a continuous stochastic process. These processes are constructed by applying the observation function to the measure-theoretic deterministic system. Hence for any of these stochastic processes the following holds: the outcomes of the stochastic process are the observed values of the corresponding deterministic system; and the realisations of the stochastic processes and the solutions of the corresponding deterministic system coarse-grained by the observation function have the same probability distribution. Consequently, according to the characterisation above, (M, Σ_M, μ, T) *observed with* Φ *is observationally equivalent to stochastic process* $\{\Phi(T^t);\ t \in \mathbb{Z}\}$, and (M, Σ_M, μ, T_t) *observed with* Φ *is observationally equivalent to stochastic process* $\{\Phi(T_t);\ t \in \mathbb{R}\}$.

But the important question is whether $\{\Phi(T^t);\ t \in \mathbb{Z}\}$ and $\{\Phi(T_t);\ t \in \mathbb{R}\}$ are nontrivial. Indeed, *they are often nontrivial.* I give now a theorem for discrete time and a theorem for continuous time which show this by characterising a class of measure-theoretic deterministic systems as systems that yield stochastic processes which are nontrivial in a certain sense. Besides, several other results also indicate this (cf. Cornfeld et al. 1982, pp. 178–179).[2]

[2]For instance, if discrete Kolmogorov systems or continuous Kolmogorov systems are observed with a finite-valued observation function, one obtains nontrivial stochastic pro-

Recall Definition 7 of a partition. Let me make the realistic assumption that the observations have finite accuracy, i.e., that only finitely many values are observed. Then one has a *finite-valued observation function* Φ; i.e., $\Phi(m) = \sum_{i=1}^{n} o_i \chi_{\alpha_i}(m)$, $M_O = \{o_i \,|\, 1 \leq i \leq n\}$, for some partition α of (M, Σ_M, μ) and some $n \in \mathbb{N}$, where χ_A denotes the characteristic function of A (cf. Cornfeld et al. 1982, p. 179). A finite-valued observation function is called *nontrivial* if, and only if, its corresponding partition is nontrivial.

The following two theorems show that under certain conditions the stochastic process $\{\Phi(T^t);\, t \in \mathbb{Z}\}$ and $\{\Phi(T_t);\, t \in \mathbb{R}\}$ are nontrivial in the following sense: for any time $k \in \mathbb{N}$ or $k \in \mathbb{R}^+$ there is an observed value $o_i \in M_O$ such that for all observed values $o_j \in M_O$ the probability of moving in k time steps from o_i to o_j is smaller than 1. Hence there are two or more observed values that one can reach in k time steps from o_i; and the probability that o_i moves to any of these observed values is between 0 and 1. These are strong results because irrespective of how closely one looks at the measure-theoretic deterministic systems, one always obtains nontrivial stochastic processes.

Theorem 1 *If, and only if, for the discrete measure-preserving deterministic system* (M, Σ_M, μ, T) *there does not exist an* $n \in \mathbb{N}$ *and a* $C \in \Sigma_M$, $0 < \mu(C) < 1$, *such that, except for a set of measure zero,* $T^n(C) = C$, *then the following holds: for every nontrivial finite-valued observation function* $\Phi : M \rightarrow M_O$, $M_O = \cup_{l=1}^{r} o_l$, $r \in \mathbb{N}$, *every* $k \in \mathbb{N}$ *and the stochastic process* $\{Z_t = \Phi(T^t);\, t \in \mathbb{Z}\}$ *there is an* $o_i \in M_O$ *such that for all* $o_j \in M_O$, $P\{Z_{t+k} = o_j \,|\, Z_t = o_i\} < 1$.[3]

For a proof of this theorem, see subsection 5.7.1.

Theorem 2 *If, and only if, for the continuous measure-preserving deterministic system* (M, Σ_M, μ, T_t) *there does not exist a* $n \in \mathbb{R}^+$ *and a* $C \in \Sigma_M$, $0 < \mu(C) < 1$, *such that, except for a set of measure zero,* $T_n(C) = C$, *then*

cesses because for Kolmogorov systems the entropy of any finite partition $H(\alpha, T)$ or $H(\alpha, T_1)$ (see equation (3.6)) is positive (cf. Cornfeld et al. 1982, pp. 280–283; Petersen 1983, p. 83).

[3]For a random variable Z to a measurable space $(\bar{M}, \Sigma_{\bar{M}})$ where $\bar{M}$ is finite the conditional probability is defined as usual as:
$P\{Z \in A \,|\, Z \in B\} = P\{Z \in A \cap B\}/P\{Z \in B\}$ for all $A, B \in \Sigma_{\bar{M}}$ with $P\{Z \in B\} > 0$.

the following holds: for every nontrivial finite-valued observation function $\Phi : M \to M_O$, $M_O = \cup_{l=1}^{r} o_l$, $r \in \mathbb{N}$, *every* $k \in \mathbb{R}^+$ *and the stochastic process* $\{Z_t = \Phi(T_t);\ t \in \mathbb{R}\}$ *there is an outcome* $o_i \in M_O$, *such that for all possible outcomes* $o_j \in M_O$, $P\{Z_{t+k} = o_j \mid Z_t = o_i\} < 1$.

For a proof of this theorem, see subsection 5.7.2.

Now recall Definition 2.5 of being ergodic. An alternative and equivalent definition of ergodicity is the following (Cornfeld et al. 1982, pp. 14–15):

Definition 35 *A discrete measure-preserving deterministic system* (M, Σ_M, μ, T) *is* ergodic *if, and only if, there is no set* $A \in \Sigma_M$, $0 < \mu(A) < 1$, *such that, except for a set of measure zero,* $T(A) = A$.

And note the following: the assumption of Theorem 1 that there does not exist an $n \in \mathbb{N}$ and an $C \in \Sigma_M$, $0 < \mu(C) < 1$, such that, except for a set of measure zero, $T^n(C) = C$ is equivalent to the condition that the discrete measure-preserving deterministic system (M, Σ_M, μ, T^n) is ergodic for all $n \in \mathbb{N}$. And the assumption of Theorem 2 that that there does not exist an $n \in \mathbb{R}^+$ and a $C \in \Sigma_M$, $0 < \mu(C) < 1$, such that, except for a set of measure zero, $T_n(C) = C$, is equivalent to the condition that the discrete measure-preserving deterministic system (M, Σ_M, μ, T_n) is ergodic for all $n \in \mathbb{R}^+$.

Both discrete and continuous measure-preserving deterministic systems are typically what is called 'weakly mixing' (cf. Definition 23 and Definition 24) (Halmos 1944, Halmos 1949). It is easy to see that any discrete weakly mixing deterministic system satisfies the assumption of Theorem 1 (in fact weakly mixing is stronger than this assumption).[4] In the continuous case, as I have explained in subsection 3.4.2, the condition that there does not exist a $n \in \mathbb{R}^+$ and a $C \in \Sigma_M$, $0 < \mu(C) < 1$, such that, except for a set of measure zero, $T^n(C) = C$, is equivalent to the measure-preserving deterministic system being weakly mixing (Hopf 1932*b*). Hence weak mixing

[4] First, assume that for a weakly mixing discrete measure-preserving deterministic system there exists an $n \in \mathbb{N}$ and a $C \in \Sigma_M$, $0 < \mu(C) < 1$, such that, except for a set of measure zero, $T^n(C) = C$. But then equation (23) cannot hold for $A = C$ and $B = C$. In subsection 5.5.2 I will show that the irrational rotation on the circle satisfies the assumption of Theorem 1 but is not weakly mixing.

is strictly stronger than the assumption of Theorem 1 but equivalent to the assumption of Theorem 2, and this indicates a difference between my results for continuous time and my results for discrete time. So we conclude that Theorem 1 and Theorem 2 show that for typical measure-preserving deterministic systems any finite-valued observation function yields a nontrivial stochastic process.

Yet this does not say much about whether the measure-preserving deterministic systems encountered in science fulfill the assumptions of Theorem 1 or Theorem 2 because the measure-preserving deterministic systems encountered in science constitute a small class of all measure-preserving deterministic systems. Indeed, recall the discussion of the KAM theorem in subsection 4.3.2. The KAM theorem says that that the phase space of integrable Hamiltonian deterministic systems which are perturbed by a small nonintegrable perturbation breaks up into stable regions and regions with unpredictable behaviour. With increasing perturbation the regions with unpredictable behaviour become larger and often eventually cover nearly the entire hypersurface of constant energy. Because according to the KAM theorem the solutions of a system are often confined to a region of positive measure smaller than 1, this means that these systems, and their discrete versions, do not satisfy the assumptions of Theorem 1 or Theorem 2 (cf. Berkovitz et al. 2006, section 4).

Despite this, Theorem 1 applies to several deterministic systems encountered in science. For recall that there are several physically relevant discrete and continuous chaotic systems and that chaotic systems are strongly mixing (cf. subsection 4.3.2). It is clear that any strongly mixing measure-preserving deterministic system is also weakly mixing. Therefore, there are several physically relevant discrete and continuous deterministic systems which are weakly mixing (later in subsection 5.3.1 I will say more about which kind of stochastic processes you obtain from observing measure-theoretic deterministic systems encountered in science). For instance, the baker's system (Example 1) is weakly mixing; thus it satisfies the assumption of Theorem 1. Billiards with convex obstacles (Example 2) are also weakly mixing and thus satisfy the assumption of Theorem 2. Consequently, for the baker's system or a billiard system with convex obstacles any finite-valued observation function gives rise to a nontrivial stochastic process. Moreover, in subsection 5.5.2 it will be shown that there are even deterministic systems which are neither chaotic nor chaotic on a region of phase space but which satisfy Theorem 1.

Second, even if the whole measure-theoretic deterministic system does not satisfy the assumption of Theorem 1 or Theorem 2, the motion of the deterministic system restricted to some regions of phase space might well satisfy this assumption. In fact, Theorem 1 and Theorem 2 immediately imply the following results. Assume that for a discrete measure-preserving deterministic system (M, Σ_M, μ, T) there is a $A \in \Sigma_M, \mu(A) > 0$, such that the deterministic system restricted to A[5] fulfills the assumption of Theorem 1. Then all observations which discriminate between values in A lead to nontrivial stochastic processes. That is, for any observation function $\Phi(m) = \sum_{i=1}^{n} o_i \chi_{\alpha_i}(m)$ where there are $h, l,\ h \neq l$, such that $\mu(A \cap \alpha_h) \neq 0$ and $\mu(A \cap \alpha_l) \neq 0$, we have that for all $k \in \mathbb{N}$ there is an outcome $o_i \in M_O$ such that for all outcomes $o_j \in M_O$ it holds that $P\{Z_{t+k} = o_j \mid Z_t = o_i\} < 1$. Likewise, assume that for a continuous measure-preserving deterministic system (M, Σ_M, μ, T_t) there is a $A \in \Sigma_M, \mu(A) > 0$, such that Theorem 2 applies to the deterministic system restricted to A. Then for any $\Phi(m) = \sum_{i=1}^{n} o_i \chi_{\alpha_i}(m)$ where there are $h, l,\ h \neq l$, such that $\mu(A \cap \alpha_h) \neq 0$ and $\mu(A \cap \alpha_l) \neq 0$, we have that for all $k \in \mathbb{R}^+$ there is an $o_i \in M_O$ such that for all $o_j \in M_O$ it holds that $P\{Z_{t+k} = o_j \mid Z_t = o_i\} < 1$.

In particular, although mathematically little is known, it is conjectured that the motion restricted to unstable regions of KAM-type systems is weakly mixing (cf. section 4.3.2). If this is true, then my argument shows that for many observation functions of KAM-type systems one obtains nontrivial stochastic processes.

Theorem 1 and Theorem 2 show that several measure-theoretic deterministic systems, regardless which finite-valued observation function is applied, yield nontrivial stochastic processes. To appreciate this result, and for what follows later, it is important to note the following. For discrete time, assume that the stochastic process $\{\Phi(T^t);\ t \in \mathbb{Z}\}$, where (M, Σ_M, μ, T) is a measure-

[5] That is, the measure-preserving deterministic system $(A, \Sigma_{M\cap A}, \mu_A, T_A)$, where $\Sigma_{M\cap A} = \{B \cap A \mid B \in \Sigma_M\}$, $\mu_A(X) = \frac{\mu(X)}{\mu(A)}$, and T_A denotes T restricted to A. (By assumption, $T_A : A \to A$ is bijective).

theoretic deterministic system and Φ is an observation function, matches our observations and is trivial (the transition probabilities are zero or one); for continuous time assume that the stochastic process $\{\Phi(T_t);\, t \in \mathbb{R}\}$, where (M, Σ_M, μ, T_t) is a measure-theoretic deterministic system and Φ is an observation function, matches our observations and is trivial (the transition probabilities are zero or one). That a trivial stochastic process is obtained does not imply that the observations derive from a deterministic system because the trivial stochastic process may arise from an observed nontrivial stochastic process. *Trivial stochastic processes can derive from both observing deterministic systems and observing nontrivial stochastic processes.* Let me explain this with two examples.

Consider the measure-preserving deterministic system (M, Σ_M, μ, T) consisting of two copies of the baker's system (Example 1) where $M = ([0,1] \times [0,1] \setminus D) \cup ([2,3] \times [0,1] \setminus D')$ with $D' = \{(x,y) \in [2,3] \times [0,1] \mid x = 2 + j/2^n$ or $y = j/2^n$, $n \in \mathbb{N}$, $0 \leq j \leq 2^n\}$, Σ_M is the Lebesgue σ-algebra on M, μ the normalised Lebesgue measure on M, T restricted to $[0,1] \times [0,1] \setminus D$ is the baker's system, and T restricted to $[2,3] \times [0,1] \setminus D'$ is the baker's system shifted to the right by $(0,2)$. Consider the partition $\{\zeta_1, \zeta_2\} = \{[0,1] \times [0,1] \setminus D, [2,3] \times [0,1] \setminus D'\}$, and the observation function $\Phi(m) = o_1\chi_{\zeta_1}(m) + o_2\chi_{\zeta_2}(m)$. So Φ merely reminds us in which of the two copies of the baker's system the state of the system is in. Then, clearly, all transition probabilities of the stochastic process $\{\Phi(T^t);\, t \in \mathbb{Z}\}$ are zero or one. Now let $\gamma = \alpha \cup \beta$ be a partition of M where $\alpha = \{\alpha_1, \ldots, \alpha_n\}$ is a nontrivial partition of $[0,1] \times [0,1] \setminus D$ and $\beta = \{\beta_1, \ldots, \beta_h\}$ is a nontrivial partition of $[2,3] \times [0,1] \setminus D'$ and define $\Psi(m) = \sum_{i=1}^{n} u_i\chi_{\alpha_i}(m) + \sum_{j=1}^{h} v_j\chi_{\beta_j}(m)$. Because the baker's system is weakly mixing, $\{\Psi(T^t);\, t \in \mathbb{Z}\}$ is a nontrivial stochastic process. Now define the observation function $\Gamma : \{u_1, \ldots, u_n, v_1, \ldots, v_h\} \to \{o_1, o_2\}$, $\Gamma(u_i) = o_1$ for all i and $\Gamma(v_j) = o_2$ for all j. Γ tells us whether the outcome is one of the u_i or one of the v_j, and so $\Gamma(\Psi(m))$ tells us which of the two copies of the baker's system the state is in. Therefore, for all $t \in \mathbb{Z}$ we have $\Phi(T^t) = \Gamma(\Psi(T^t))$, and thus $\{\Phi(T^t);\, t \in \mathbb{Z}\}$ is identical to $\{\Gamma(\Psi(T^t));\, t \in \mathbb{Z}\}$. Consequently, the trivial stochastic process $\{\Phi(T^t);\, t \in \mathbb{Z}\}$ is obtained from observing the nontrivial stochastic process $\{\Psi(T^t);\, t \in \mathbb{Z}\}$ with the observation function Γ.

Or, to start from a stochastic process, consider the nontrivial Markov process $\{Z_t;\, t \in \mathbb{Z}\}$ (cf. Example 5) with outcome space $\{s_1, s_2, s_3, s_4\}$ where $P\{Z_t = s_i\} = 1/4$, for all i, $1 \le i \le 4$, $P\{Z_t = s_i \,|\, Z_{t-1} = s_j\} = 1/2$ for all i, j, $1 \le i, j \le 2$, and $P\{Z_t = s_i \,|\, Z_{t-1} = s_j\} = 1/2$ for all i, j, $3 \le i, j \le 4$. This means that the outcomes s_1 and s_2 can be reached from each other but not from the outcomes s_3 or s_4, and, likewise, that the outcomes s_3 and s_4 can be reached from each other but not from the outcomes s_1 or s_2. Thus the Markov process can be split into two parts: the dynamics involving s_1 and s_2 and the dynamics involving s_3 and s_4.[6] Consider the observation function $\Gamma : \{s_1, s_2, s_3, s_4\} \to \{o_1, o_2\}$ where $\Gamma(s_1) = \Gamma(s_2) = o_1$ and $\Gamma(s_3) = \Gamma(s_4) = o_2$. Γ tells us whether the outcome of the Markov process is in $\{s_1, s_2\}$ or in $\{s_3, s_4\}$. So, clearly, $\{\Gamma(Z_t);\, t \in \mathbb{Z}\}$ is a trivial stochastic process (all transition probabilities are 0 or 1). But it is obtained from observing the nontrivial Markov process $\{Z_t;\, t \in \mathbb{Z}\}$ with the observation function Γ.

5.2.2 Stochastic processes simulated by deterministic systems

I have shown that measure-theoretic deterministic systems, when observed, can yield nontrivial stochastic processes. But can one find, for every stochastic process, a measure-theoretic deterministic system which produces this process?

The following idea of how to simulate stochastic processes by deterministic systems is well known in the technical literature (Petersen 1983, pp. 6–7)[7] and is known to philosophers (Butterfield 2005); I also need to discuss it for what follows later. The underlying thought is that for each realisation r_ω, one sets up a deterministic system with phase space $\{r_\omega\}$.

So start with a discrete stochastic process $\{Z_t;\, t \in \mathbb{Z}\}$ from $(\Omega, \Sigma_\Omega, \nu)$ to $(\bar{M}, \Sigma_{\bar{M}})$. Let M be the set of all bi-infinite sequences $m = (\ldots m_{-1} m_0 m_1 \ldots)$ with $m_i \in \bar{M},\ i \in \mathbb{Z}$, and let m_t be the t-th coordinate of $m,\ t \in \mathbb{Z}$. Let Σ_M

[6] Hence, technically, the Markov process $\{Z_t,\, t \in \mathbb{Z}\}$ is not irreducible (see Example 5).

[7] Petersen discusses it only for stationary stochastic processes; I consider generally stochastic processes.

be the σ-algebra generated by the semi-algebra of cylinder-sets

$$C^{A_1\ldots A_n}_{i_1\ldots i_n}=\{m\in M \mid m_{i_1}\in A_1,\ldots,m_{i_n}\in A_n,\, A_j\in\Sigma_{\bar{M}},\, i_j\in\mathbb{Z},\, i_1<\ldots<i_n,\, 1\leq j\leq n\}. \tag{5.1}$$

$\{Z_t; t\in\mathbb{Z}\}$ assigns to each cylinder set $C^{A_1\ldots A_n}_{i_1\ldots i_n}$ a pre-measure, namely the probability $P\{Z_{i_1}\in A_1,\ldots,Z_{i_n}\in A_n\}$. Let μ be the unique extension of this pre-measure to a measure on Σ_M. Let $T: M\rightarrow M$ be the left shift, i.e., $T((\ldots m_{-1}m_0m_1\ldots))=(\ldots m_0m_1m_2\ldots)$. Then one obtains the deterministic system (M,Σ_M,μ,T). Finally, assume one sees only the 0-th coordinate of the sequence m, i.e., one applies the observation function $\Phi_0: M\rightarrow\bar{M}, \Phi_0(m)=m_0$. I now define:

Definition 36 $(M,\Sigma_M,\mu,T,\Phi_0)$ *as constructed above is the* deterministic representation *of the discrete stochastic process* $\{Z_t;\, t\in\mathbb{Z}\}$.

For a continuous stochastic process $\{Z_t;\, t\in\mathbb{R}\}$ from $(\Omega,\Sigma_\Omega,\nu)$ to $(\bar{M},\Sigma_{\bar{M}})$ let M be the set of all functions $m(\tau)$ from $\mathbb{R}$ to $\bar{M}$. Let Σ_M be the σ-algebra on M generated by the cylinder sets (5.1) as defined above where you replace m_{i_j} by $m(i_j)$ and the i_j are arbitrary numbers in $\mathbb{R}$. Again, $\{Z_t;\, t\in\mathbb{R}\}$ assigns to each cylinder set $C^{A_1\ldots A_n}_{i_1\ldots i_n}$ the pre-measure $P\{Z_{i_1}\in A_1,\ldots,Z_{i_n}\in A_n\}$. Let μ be the unique extension of this pre-measure to a measure on Σ_M, and let $T_t(m(\tau))=m(\tau+t)$. Then (M,Σ_M,μ,T_t) is a continuous measure-theoretic deterministic system (cf. Doob 1953, pp. 621–622). Finally, assume one applies the observation function $\Phi_0(m(\tau))=m(0)$. Again, I define:

Definition 37 $(M,\Sigma_M,\mu,T_t,\Phi_0)$ *as constructed above is the* deterministic representation *of the continuous stochastic process* $\{Z_t;\, t\in\mathbb{R}\}$.

For the deterministic representation of a stochastic process it is assumed that the 0-th coordinate is observed. Consequently, the possible outcomes of a stochastic process are the possible observed values of its deterministic representation. Clearly, any realisation r_ω of the stochastic process is contained in M, and observing the solution s_{r_ω} with Φ_0 gives exactly r_ω. Furthermore, the measure μ is defined by the probabilities which are assigned by the stochastic process to each cylinder set. Hence the probability distribution over the realisations of a stochastic process is the same as the one over the sequences of observed values of its deterministic representation. Thus, according to the characterisation at the start of this section, *a stochastic process is observationally equivalent to its deterministic representation. Hence every stochastic process can be simulated by at least one measure-theoretic deterministic system.* (When there is no risk of confusion, I also refer to the measure-theoretic deterministic system (M, Σ_M, μ, T) and to the measure-theoretic deterministic system (M, Σ_M, μ, T_t) of the deterministic representation $(M, \Sigma_M, \mu, T, \Phi_0)$ and $(M, \Sigma_M, \mu, T_t, \Phi_0)$, respectively, as the deterministic representation.)

For instance, for a Bernoulli process with probabilities $(p_1, \ldots, p_N)$ (Example 4) we already encountered the deterministic representation when discussing the meaning of a Bernoulli system (cf. subsection 3.4.1). Namely, the deterministic representation of a Bernoulli process with probabilities $(p_1, \ldots, p_N)$ is the following: (M, Σ_M, μ, T) is the Bernoulli shift $(\Omega, \Sigma_\Omega, \nu, T)$ with probabilities $(p_1, \ldots, p_N)$ corresponding to the Bernoulli process (see Definition 18 of a Bernoulli shift) and $\Phi_0(\omega) = \omega_0$.

From a philosophical perspective the deterministic representation is a cheat because its states are constructed to encode the future and past outcomes of the stochastic process. Despite this, it is important to know that the deterministic representation exists. Of course, there is the question whether deterministic systems which do not involve a cheat can simulate a given stochastic process. I will turn to this question in the sections 5.3 and 5.4, where I will show that for some stochastic processes this is indeed the case. To my knowledge, it is unknown whether *every* stochastic process can be thus simulated.

5.2.3 A mathematical definition of observational equivalence

Let me now mathematically define what it means for a stochastic process and a measure-theoretic deterministic system to be observationally equivalent. Recall the definition of isomorphic measusure-preserving deterministic systems (Definition 19). Isomorphic deterministic systems may have different phase spaces. But if *identical* sets $\hat{M}_1$ and $\hat{M}_2$ can be found, then the measure-preserving deterministic systems are obviously probabilistically equivalent and have, from a probabilistic viewpoint, the same phase space; for this case it will later be convenient to say that the measure-preserving deterministic systems are *manifestly isomorphic.*

According to the characterisation at the beginning of this section, a measure-theoretic deterministic system (M, Σ_M, μ, T) for discrete time, or a measure-theoretic deterministic system (M, Σ_M, μ, T_t) for continuous time, observed with the observation function Φ, gives the same predictions as a stochastic process $\{Z_t;\, t \in \mathbb{Z}\}$ or $\{Z_t;\, t \in \mathbb{R}\}$ just in case the following conditions hold: (i) the set of possible outcomes with positive probability is identical to the set of possible observed values with positive probability, and (ii) the deterministic representation of the stochastic process $\{\Phi(T^t); t \in \mathbb{Z}\}$ for discrete time, or $\{\Phi(T_t); t \in \mathbb{R}\}$ for continuous time, is probabilistically equivalent to the deterministic representation of the stochastic process $\{Z_t; t \in \mathbb{Z}\}$ for discrete time, or $\{Z_t; t \in \mathbb{R}\}$ for continuous time. Hence one arrives at the following definition of 'observational equivalence'; (for what follows, a definition for measure-preserving deterministic systems and, correspondingly, stationary stochastic processes will suffice):[8]

Definition 38 *For discrete time the stationary stochastic process* $\{Z_t;\, t \in \mathbb{Z}\}$ *and the measure-preserving deterministic system* (M, Σ_M, μ, T)*, observed*

[8] For a discrete measure-preserving deterministic system (M, Σ_M, μ, T), the stochastic process $\{\Phi(T^t);\, t \in \mathbb{Z}\}$ is stationary: $\{x \in M \,|\, \Phi(T^{t_1}(x)) \in A_1, \ldots, \Phi(T^{t_n}(x)) \in A_n,\, A_i \in \Sigma_{M_O}, t_i \in \mathbb{Z},\, n \in \mathbb{N}\}$ is identical to $A = T^{-t_1}(\Phi^{-1}(A_1) \cap \ldots \cap T^{t_1-t_n}\Phi^{-1}(A_n))$. Likewise, for any $h \in \mathbb{Z}$, $\{x \in M \,|\, \Phi(T^{t_1+h}(x)) \in A_1, \ldots, \Phi(T^{t_n+h}(x)) \in A_n\}$ is $B = T^{-(t_1+h)}(\Phi^{-1}(A_1) \cap \ldots \cap T^{t_1-t_n}\Phi^{-1}(A_n))$. Because the deterministic system is measure-preserving, $\mu(A) = \mu(B)$, implying that $\{\Phi(T^t);\, t \in \mathbb{Z}\}$ is stationary. Basically the same argument (the only difference being that h and the t_i in $\mathbb{R}$ can be arbitrary) shows that for a continuous measure-preserving deterministic system (M, Σ_M, μ, T_t) the stochastic process $\{\Phi(T_t);\, t \in \mathbb{R}\}$ is stationary. And if a stochastic process is stationary, its deterministic representation is measure-preserving. For, in the discrete case, if $\{Z_t;\, t \in \mathbb{Z}\}$ is stationary, then for the deterministic representation (M, Σ_M, μ, T), $\mu(T(A)) = \mu(A)$ for any cylinder set A and hence $\mu(T(A)) = \mu(A)$ for all $A \in \Sigma_M$. Likewise, in the continuous case, the stationarity of $\{Z_t;\, t \in \mathbb{R}\}$ implies for the deterministic representation (M, Σ_M, μ, T_t) that $\mu(T_t(A)) = \mu(A)$ for any cylinder set A and any $t \in \mathbb{R}$. Therefore, $\mu(T_t(A)) = \mu(A)$ for all $A \in \Sigma_M$ and all $t \in \mathbb{R}$ (cf. Cornfeld et al. 1982, p. 178).

with Φ*, are* observationally equivalent *if, and only if, the deterministic representation of* $\{\Phi(T^t);\, t \in \mathbb{Z}\}$ *is manifestly isomorphic to the deterministic representation of* $\{Z_t;\, t \in \mathbb{Z}\}$*. Likewise, for continuous time the stationary stochastic process* $\{Z_t;\, t \in \mathbb{R}\}$ *and the measure-preserving deterministic system* (M, Σ_M, μ, T_t)*, observed with* Φ*, are* observationally equivalent *if, and only if, the deterministic representation of* $\{\Phi(T_t);\, t \in \mathbb{R}\}$ *is manifestly isomorphic to the deterministic representation of* $\{Z_t;\, t \in \mathbb{R}\}$*.*

For measure-preserving deterministic systems and stationary stochastic processes all the cases of observational equivalence already discussed are cases of observational equivalence in the sense of Definition 38. First, I claimed in subsection 5.2.1 that (M, Σ_M, μ, T) observed with Φ is observationally equivalent to the stochastic process $\{\Phi(T^t);\, t \in \mathbb{Z}\}$ and that (M, Σ_M, μ, T_t) observed with Φ is observationally equivalent to the stochastic process $\{\Phi(T_t);\, t \in \mathbb{R}\}$. This is true because every deterministic system is manifestly isomorphic to itself. Second, I claimed in subsection 5.2.2 that the deterministic representation $(M, \Sigma_M, \mu, T, \Phi_0)$ of $\{Z_t;\, t \in \mathbb{Z}\}$ is observationally equivalent to $\{Z_t;\, t \in \mathbb{Z}\}$ and that the deterministic representation $(M, \Sigma_M, \mu, T_t, \Phi_0)$ of $\{Z_t;\, t \in \mathbb{R}\}$ is observationally equivalent to $\{Z_t;\, t \in \mathbb{R}\}$. This is true because the deterministic representation of $\{\Phi_0(T^t);\, t \in \mathbb{Z}\}$ is $(M, \Sigma_M, \mu, T, \Phi_0)$ and the deterministic representation of $\{\Phi_0(T_t);\, t \in \mathbb{R}\}$ is $(M, \Sigma_M, \mu, T_t, \Phi_0)$.

One important final point: assume that the discrete measure-preserving deterministic system (M, Σ_M, μ, T) is isomorphic (via a function $\phi : \hat{M} \rightarrow \hat{M}_2$) to the deterministic representation $(M_2, \Sigma_{M_2}, \mu_2, T_2, \Phi_0)$ of the stochastic process $\{Z_t;\, t \in \mathbb{Z}\}$. This means that *there is a one-to-one correspondence between the solutions of the deterministic system and the realisations of the stochastic process.* Thus (M, Σ_M, μ, T) observed with $\Phi_0(\phi(m))$, where it does not matter how $\Phi_0(\phi(m))$ is defined for $m \in M \setminus \hat{M}$, is observationally equivalent to $\{Z_t;\, t \in \mathbb{Z}\}$. This is so because the deterministic representation of $\{\Phi_0(\phi(T^t));\, t \in \mathbb{Z}\}$ where T is restricted to $\hat{M}$ is identical to the deterministic representation of $\{\Phi_0(T_2^t);\, t \in \mathbb{Z}\}$ where T_2 is restricted to $\hat{M}_2$. Hence the deterministic representation of $\{\Phi_0(\phi(T^t));\, t \in \mathbb{Z}\}$ is manifestly isomorphic to $(M_2, \Sigma_{M_2}, \mu_2, T_2)$. The same argument shows that if a continuous measure-preserving deterministic system (M, Σ_M, μ, T_t) is isomorphic (via the function $\phi : \hat{M} \to \hat{M_2}$) to the deterministic representation $(M_2, \Sigma_{M_2}, \mu_2, T_t^2, \Phi_0)$ of a continuous stochastic process $\{Z_t;\, t \in \mathbb{R}\}$, then the following holds: the deterministic system (M, Σ_M, μ, T_t) observed with $\Phi_0(\phi(m))$, where it does not matter how $\Phi_0(\phi(m))$ is defined for $m \in M \setminus \hat{M}$, is observationally equivalent to $\{Z_t;\, t \in \mathbb{R}\}$.

Two important instances of this principle are as follows: first, recall Definition 20 of a discrete Bernoulli system. The meaning of discrete Bernoulli systems is clear, viz. the solutions of a discrete Bernoulli system can be put into one-to-one correspondence with the realisations of a Bernoulli process. A discrete Bernoulli system (M, Σ_M, μ, T) is isomorphic (via the function ϕ) to the deterministic representation $(M_2, \Sigma_{M_2}, \mu_2, T_2, \Phi_0)$ of a Bernoulli process with probabilities $(p_1, \ldots, p_n)$. Consequently, (M, Σ_M, μ, T) observed with $\Phi_0(\phi)$ produces a Bernoulli process with probabilities $(p_1, \ldots, p_n)$. At this point it is worth mentioning again the result that two Bernoulli shifts (and hence two discrete Bernoulli systems) are isomorphic if, and only if, they have the same Kolmogorov-Sinai entropy, where the Kolmogorov-Sinai entropy of a Bernoulli shift with probabilities $(p_1, \ldots, p_n)$ is $\sum_{i=1}^{n} -p_i \log p_i$ (see subsection 3.4.1).

Second, recall Definition 30 of a continuous Bernoulli system. As we have seen, two continuous Bernoulli systems (M, Σ_M, μ, T_t) and $(M_2, \Sigma_{M_2}, \mu_2, T_t^2)$ are isomorphic if, and only if, they have the same Kolmgogorov-Sinai entropy. We have seen even more, namely that up to a scaling of time any two continuous Bernoulli systems are isomorphic. That is, given two continuous Bernoulli systems (M, Σ_M, μ, T_t) and $(M_2, \Sigma_{M_2}, \mu_2, T_t^2)$ there is a $c \in \mathbb{R}^+$ such that for $t' = ct$ the measure-preserving deterministic systems (M, Σ_M, μ, T_t) and $(M_2, \Sigma_{M_2}, \mu_2, T_{t'}^2)$ are isomorphic (cf. subsection 3.4.3).

Recall the definition of irrationally related semi-Markov processes (cf. Example 7 where it is also defined what it means for a semi-Markov process to be irrationally related). It can be proven that the deterministic representation $(M_2, \Sigma_{M_2}, \mu_2, T_t^2)$ of any irrationally related semi-Markov process $\{Z_t;\, t \in \mathbb{R}\}$ is a continuous Bernoulli system (Ornstein 1970*b*; Ornstein 1974, pp. 56–61). And, clearly, for any $c \in \mathbb{R}^+$ and $t' = ct$, the measure-preserving deterministic system $(M_2, \Sigma_{M_2}, \mu_2, T_{t'}^2)$ is the deterministic representation of an irrationally related semi-Markov process. Hence given any continuous

Bernoulli system (M, Σ_M, μ, T_t) there is an irrationally related semi-Markov process $\{Z_t;\ t \in \mathbb{R}\}$ whose deterministic representation $(M_2, \Sigma_{M_2}, \mu_2, T_t^2, \Phi_0)$ is isomorphic (via a function ϕ) to (M, Σ_M, μ, T_t). And this means that the continuous Bernoulli system (M, Σ_M, μ, T_t) observed with $\Phi_0(\phi)$ produces the irrationally related semi-Markov process $\{Z_t; t \in \mathbb{R}\}$.

Likewise, the deterministic representation $(M_2, \Sigma_{M_2}, \mu_2, T_t^2)$ of any irrationally related multi-step semi-Markov processes $\{Z_t; t \in \mathbb{R}\}$ (Example 8) is a continuous Bernoulli system (Park 1982). And, clearly, for any $c \in \mathbb{R}^+$ and $t' = ct$, $(M_2, \Sigma_{M_2}, \mu_2, T_{t'}^2)$ is the deterministic representation of an irrationally related multi-step semi-Markov process. Thus for any continuous Bernoulli system (M, Σ_M, μ, T_t) there is an irrationally related multi-step semi-Markov process $\{Z_t; t \in \mathbb{R}\}$ whose deterministic representation $(M_2, \Sigma_{M_2}, \mu_2, T_t^2, \Phi_0)$ is isomorphic (via a function ϕ) to (M, Σ_M, μ, T_t). Therefore, (M, Σ_M, μ, T_t) observed with $\Phi_0(\phi)$ yields $\{Z_t; t \in \mathbb{R}\}$.

5.3 Advanced observational equivalence I

In this section and the following section I will discuss results which are 'advanced' in the sense that they are about the question whether it is possible to simulate measure-theoretic deterministic systems *used in science* with stochastic processes *used in science*. The phrase 'measure-theoretic deterministic systems used in science' (or 'stochastic processes used in science') is a short-hand for measure-theoretic deterministic systems (or stochastic processes) which are used in science to model phenomena.

5.3.1 Deterministic systems used in science which simulate stochastic processes used in science

The deterministic representation does not naturally arise in science (no doubt reflecting the fact that is a philosophical cheat). And the results so far only show that a stochastic process used in science, e.g., a Bernoulli process, can be simulated by its deterministic representation. Hence it seems hard to imagine how measure-theoretic deterministic systems used in science could simulate stochastic processes used in science. In particular, it seems hard to imagine how measure-theoretic deterministic systems in science could be random enough to simulate random stochastic processes such as Bernoulli processes. Thus one might conjecture that *it is impossible to simulate stochastic processes used in science by measure-theoretic deterministic systems used in science.*

Indeed, until the 1960s Kolmogorov and many other scientists believed this. More specifically, Kolmogorov conjectured that while the measure-preserving deterministic systems which simulate the stochastic processes used in science produce positive information, the measure-preserving deterministic systems used in science produce no, i.e., zero information. As I have explained when discussing the Kolmogorov-Sinai entropy (see subsection 3.4.1), the Kolmogorov-Sinai entropy was introduced to capture the information produced by a measure-preserving deterministic system, or equivalently, the amount of uncertainty produced by a measure-preserving deterministic system; and a positive Kolmogorov-Sinai entropy indicates the property that positive information or uncertainty is produced. Kolmogorov and his colleagues expected that this property of producing positive information would accomplish the separation of stochastic processes used in science from measure-preserving deterministic systems used in science. So it was a big surprise when from the 1960s onwards it was found that also many measure-preserving deterministic systems used in science have positive Kolmogorov-Sinai entropy and thus produce positive information (shortly, I will list several examples of continuous and discrete deterministic systems used in science with positive Kolmogorov-Sinai entropy). Hence Kolmogorov's proposed way of separating the measure-preserving deterministic systems used in science from the stochastic processes used in science failed (Radunskaya 1992, chapter 1; Sinai 1989; Sinai 2007; Werndl 2009*a*; Werndl 2009*b*).

Before I proceed, let me mention, for what follows later, that when looking at the measure-preserving deterministic systems used in science, you see that nearly all discrete measure-preserving deterministic systems used in science have *finite* Kolmogorov-Sinai entropy. Also, nearly all continuous measure-preserving deterministic systems used in science have *finite* Kolmogorov-Sinai entropy.[9] Generally, it has been proven that all continu-

[9]'Nearly all' because there are a few deterministic systems which have infinite

ous measure-preserving deterministic system (M, Σ_M, μ, T_t) whose evolution functions T_t are given by Hamilton's equations have finite Kolmogorov-Sinai entropy (Arnold & Avez 1968, pp. 46–47). And, generally, it has been proven that all continuous measure-preserving deterministic systems (M, Σ_M, μ, T_t) where M is a compact manifold and the evolution functions T_t are smooth have finite Kolmogorov-Sinai entropy (Ornstein & Weiss 1991, p. 19).

All discrete and continuous Bernoulli systems have a positive Kolmogorov-Sinai entropy (but having a positive Kolmogorov-Sinai entropy is a weaker condition than being a Bernoulli system; there are many deterministic systems with positive Kolmogorov-Sinai entropy which are not Bernoulli systems) (Cornfeld et al. 1982, p. 283). As stated at the end of the last subsection, continuous Bernoulli systems, when observed with specific observation functions, can yield irrationally related semi-Markov processes and also irrationally related multi-step Markov processes, both of which are often used in science. And discrete Bernoulli systems, when observed with specific observation functions, yield Bernoulli processes (Example 4). Bernoulli processes are widely used in science and are often regarded as the most random discrete-time stochastic processes because their outcomes are probabilistically independent (Ornstein 1989). So let us ask, are there deterministic systems used in science which are discrete Bernoulli systems or continuous Bernoulli systems?

Indeed, there are, namely many of the chaotic measure-preserving deterministic systems listed in subsection 4.3.2. And all these deterministic systems are thus also examples of deterministic systems used in science with positive Kolmogorov-Sinai entropy. To start with, for continuous time: there are systems in Newtonian mechanics which are continuous Bernoulli systems, for instance, first, some hard-sphere systems; as already discussed, hard-sphere systems are important in statistical mechanics because they are a model of the ideal gas (Berkovitz et al. 2006, pp. 679–680; Ornstein 1974, pp. 8–9); second, billiard systems with convex obstacles (Exampe 2) (Ornstein & Galavotti 1974); third, geodesic flows of negative curvature (Ornstein

Kolmogorov-Sinai entropy and which some might want to classify as deterministic systems used in science, such as billiard systems with a countably infinite number of convex obstacles (cf. Haskell 1992).

& Weiss 1991, section 4). Also, there are dissipative continuous measure-preserving deterministic systems which are continuous Bernoulli systems such as the Lorenz system (Example 3) and generally Lorenz-type systems, which have been used to model weather dynamics and the motion of waterwheels (Kolář & Gumbs 1992; Lorenz 1963; Luzzatto et al. 2005; Strogatz 1994). It is usually very hard to prove that measure-preserving deterministic systems are continuous Bernoulli systems. Therefore, for many continuous deterministic systems it is only conjectured, but not proven, that they are continuous Bernoulli systems, e.g., for *all* hard sphere systems and the motion of KAM-type deterministic systems restricted to some regions of phase space (Berkovitz et al. 2006, pp. 679–680; Young 1997; Sźasz 2000).

In the discrete case, the following measure-preserving deterministic systems, for instance, are discrete Bernoulli systems: first, the somewhat artificial example of the baker's system (Example 1); second, generalised versions of the logistic map; the logistic map has been endorsed as a simplified model of population dynamics and climate dynamics (Jacobson 1981; Lorenz 1964; Lyubich 2002; May 1976); third, the Hénon map for certain parameter values and generalised versions thereof; the Hénon map has been proposed as a simplified model of weather dynamics (Benedicks & Young 1993; Hénon 1976). Some of the generalised versions of the logistic map and the Hénon map are dissipative, showing that there are dissipative discrete Bernoulli systems. Furthermore, it follows from Definition 30 of a continuous Bernoulli system that the discrete versions of any continuous Bernoulli system are discrete Bernoulli systems; hence the discrete versions of any of the continuous deterministic systems used in science listed above are discrete Bernoulli systems. And again, there are several discrete deterministic systems which are only conjectured, but not proven, to be discrete Bernoulli systems such as the Hénon map for certain parameter values (Benedicks & Young 1993; Young 1997).

In some contexts some of the continuous and discrete Bernoulli systems which I have listed give relatively accurate predictions, e.g., the Lorenz system as a model for waterwheels (cf. Example 3). Yet sometimes these deterministic systems are motivated as simple models which help us to understand phenomena, and not so much to predict them: e.g., the Hénon map and the

Lorenz system for weather dynamics (Lorenz 1963; Smith 1998, chapter 8; Strogatz 1994).

For reasons of illustration, let me show that the the baker's system (Example 1) is a discrete Bernoulli system. Assign to each (x, y) in M the sequence $\phi(x, y) = \ldots \omega_{-2}\omega_{-1}\omega_0\omega_1\omega_2 \ldots$ defined by the binary expansion of the coordinates:

$$x = 0.\omega_0\omega_1 \ldots = \sum_{i=1}^{\infty} \frac{\omega_{i-1}}{2^i}; \quad y = 0.\omega_{-1}\omega_{-2} \ldots = \sum_{i=1}^{\infty} \frac{\omega_{-i}}{2^i}. \tag{5.2}$$

Consider the Bernoulli shift $(M_2, \Sigma_{M_2}, \mu_2, T_2)$ with outcomes s_1, s_2 and probabilities $(\frac{1}{2}, \frac{1}{2})$. Let $\hat{M}_2$ be the subset of M_2 excluding all states beginning or ending with an infinite sequence of zeros or ones; note that $\mu_2(\hat{M}_2) = 1$. One easily verifies that $\phi : M \to \hat{M}_2$ gives an isomorphism from (M, Σ_M, μ, T) to $(M_2, \Sigma_{M_2}, \mu_2, T_2)$. Hence the baker's system with the observation function $\Phi((x, y)) = s_1\chi_{\alpha_1}((x, y)) + s_2\chi_{\alpha_2}((x, y))$, where $\alpha = \{\alpha_1, \alpha_2\} = \{[0, \frac{1}{2}) \times [0, 1] \setminus D, [\frac{1}{2}, 1] \times [0, 1] \setminus D\}$ yields the Bernoulli process with outcomes s_1, s_2 and probabilities $(\frac{1}{2}, \frac{1}{2})$.

Note that continuous and discrete Bernoulli systems are weakly mixing (Petersen 1983, p. 58). Hence Theorem 1 applies to all discrete Bernoulli systems and Theorem 2 applies to all continuous Bernoulli systems. That is, provided a discrete or continuous Bernoulli system is observed with a finite-valued observation function, one always obtains a nontrivial stochastic process.

What is the significance of the these results? They show that several continuous measure-theoretic deterministic systems, when observed, yield irrationally related semi-Markov processes and also irrationally related multistep semi-Markov processes. And they show that several discrete measure-theoretic deterministic systems used in science, when observed, produce Bernoulli processes, which are usually, we recall, regarded as the most random stochastic processes. Consequently, the conjecture advanced at the beginning of this subsection is wrong: *it is possible to simulate stochastic processes used in science by deterministic systems used in science.*[10]

[10]The arguments in this section allow any meaning of 'deterministic systems used in science' that is wide enough to include some Bernoulli systems but narrow enough to exclude deterministic systems such as the deterministic representation.

Of course, the question arises whether for measure-theoretic deterministic systems used in science which are observationally equivalent to stochastic processes used in science the corresponding observation function is *natural* in the sense that one might encounter it when modeling phenomena. The answer depends on the deterministic system, the stochastic process and the phenomenon under consideration. Sometimes the required observation function seems very involved and thus it seems unlikely that a natural interpretation can be found. But in other cases the observation function corresponds to a realistic way of observing the system.

For instance, recall that the baker's system models a particle bouncing on several mirrors where (x, y) denotes the position of the particle on a square (Example 1). Here an observer might well only be interested in whether the position of the particle is to the left or to the right of the square. Then the observation function $\Phi((x, y)) = s_1\chi_{\alpha_1}((x, y)) + s_2\chi_{\alpha_2}((x, y))$, above, which indeed produces a Bernoulli process, would be natural.[11]

5.4 Advanced observational equivalence II

The previous discussion showed that for several measure-theoretic deterministic systems used in science, regardless of which finite-valued observation function one applies, one always obtains a nontrivial stochastic process. But to obtain stochastic processes used in science such as Bernoulli processes, it seems crucial that *coarse* observation functions are applied. Hence it is hard to imagine that by taking finer and finer observations of measure-theoretic deterministic systems used in science one still obtains stochastic processes used in science. In particular, it is hard to imagine that one still obtains random stochastic processes. Therefore, one might conjecture that *it is im-*

[11]You might wonder whether it is possible to argue that all observation functions of measure-theoretic deterministic systems used in science which produce stochastic processes used in science are not natural, and hence that it is not really true that measure-theoretic deterministic systems used in science produce stochastic processes used in science. It is not possible to argue this because of fundamental results, discussed in section 5.4, which basically say that some measure-theoretic deterministic systems used in science, regardless how they are observed, yield stochastic processes used in science.

possible to simulate measure-theoretic deterministic systems used in science at every observation level by stochastic processes used in science.

5.4.1 The meaning of simulation at every observation level

What does it mean to say that 'a stochastic process of a certain type simulates a measure-theoretic deterministic system *at every arbitrary observation level*'? Let me introduce three possible meanings of this phrase. Because the previous discussion concerns, and all the examples I will be discussing are, measure-preserving deterministic systems, I will assume that a measure-preserving deterministic system is given.

The usual meaning based on ε-congruence

To introduce the first possible meaning of this phrase, I have to start by explaining what it means for a measure-preserving deterministic system and a stochastic process to give the same predictions at an observation level $\varepsilon > 0$, $\varepsilon \in \mathbb{R}$. There are two aspects. First, one imagines that in practice, for sufficiently small ε_1, one cannot distinguish states of the deterministic system which are less than the distance ε_1 apart. The second aspect concerns probabilities: in practice, for sufficiently small ε_2, one will not be able to observe differences in probabilities of less than ε_2. Assume that ε is smaller than ε_1 and ε_2. Then we can define a measure-preserving deterministic system and a stochastic process to give the same predictions at observation level ε if the following holds: the solutions of the measure-preserving deterministic system can be put into one-to-one correspondence with the realisations of the stochastic process in such a way that the actual state of the deterministic system and the corresponding outcome of the stochastic process are at each time point less then ε apart except for a set whose probability is smaller than ε. One can think of this notion of giving the same predictions at observation level ε as a kind of *shadowing* result: for each solution of the measure-preserving deterministic system the corresponding realisation of the stochastic process shadows this solution in the sense that at each time point the state of the deterministic system and the outcome of the stochastic process are within ε (except for a set whose probability is smaller than ε).

Mathematically, this idea is captured by the notion of ε-congruence. To define it, one needs to speak of distances between states in the phase space M of the deterministic system; hence one assumes a metric d_M defined on M. So we need to find a stochastic process whose outcome is within distance ε of the actual state of the deterministic system. Hence one assumes that each possible outcome of the stochastic process is a subset of the phase space of the deterministic system. Now recall Definition 36 and Definition 37 of the deterministic representation and Definition 19 of being isomorphic. So finally, I can define:

Definition 39 *Let* (M, Σ_M, μ, T) *be a discrete measure-preserving deterministic system, where* (M, d_M) *is a metric space. Let* $(M_2, \Sigma_{M_2}, \mu_2, T_2, \Phi_0)$ *be the deterministic representation of the stationary stochastic process* $\{Z_t;\, t \in \mathbb{Z}\}$ *with outcomes in* (M, d_M)*, i.e.,* $\Phi_0 : M_2 \to M$. (M, Σ_M, μ, T) *is* ε-congruent *to* $\{Z_t;\, t \in \mathbb{Z}\}$ *if, and only if,* (M, Σ_M, μ, T) *is isomorphic via a function* $\phi : M \to M_2$ *to* $(M_2, \Sigma_{M_2}, \mu_2, T_2)$ *and* $d_M(m, \Phi_0(\phi(m))) < \varepsilon$ *for all* $m \in M$ *except for a set of measure* $< \varepsilon$ *in* M*. For continuous measure-preserving deterministic systems, where* (M, d_M) *is a metric space, and continuous stationary stochastic processes* $\{Z_t;\, t \in \mathbb{R}\}$ *with outcomes in* (M, d_M) ε-congruence *is defined analogously (cf. Ornstein & Weiss 1991, pp. 22–23).*

By generalising over ε, one obtains a natural meaning of the phrase that stochastic processes of a certain type simulate a measure-preserving deterministic system at every observation level, namely: for every $\varepsilon > 0$ there is a stochastic process of this type which gives the same predictions at observation level ε. Or technically: *for every* $\varepsilon > 0$ *there exists a stochastic process of this type which is* ε*-congruent to the measure-preserving deterministic system.* This notion, referring to ε-congruence, is the standard notion of simulation at every observation level discussed in the literature (Ornstein & Weiss 1991; Suppes 1999).

Note that ε-congruence does not assume that the measure-preserving deterministic system is observed with an observation function: the *actual* states of the deterministic system, and not states observed with an observation function, are compared with the outcomes of the stochastic process. To arrive at a notion of observational equivalence no observation functions are invoked, but it is asked whether the actual state of the deterministic system and the corresponding outcome of the stochastic process are less than distance ε apart.

At this point I should mention that it follows from the discussion at the end of subsection 5.2.3 that if a discrete measure-preserving deterministic system and a discrete stochastic process $\{Z_t;\ t \in \mathbb{Z}\}$ are ε-congruent, then $\{\Phi_0(\phi(T^t));\ t \in \mathbb{Z}\}$, where $\Phi_0(\phi(T^t))$ can take any arbitrary values in $\bar{M}$ for $m \in M \setminus \hat{M}$, is the stochastic process $\{Z_t;\ t \in \mathbb{Z}\}$. Likewise, for continuous time it follows that if a continuous measure-preserving deterministic system and a continuous stochastic process $\{Z_t;\ t \in \mathbb{R}\}$ are ε-congruent, then $\{\Phi_0(\phi(T_t));\ t \in \mathbb{R}\}$, where $\Phi_0(\phi(T_t))$ can take any arbitrary value in $\bar{M}$ for $m \in M \setminus \hat{M}$, is the stochastic process $\{Z_t;\ t \in \mathbb{R}\}$. Technically, $\Phi_0(\phi)$ is an observation function of the deterministic system but for ε-congruence it is *not* interpreted in this way. Instead, the meaning of $\Phi_0(\phi)$ is as follows: when it is applied to the deterministic system, the resulting process is the stochastic process whose realisations shadow the solutions of the measure-preserving deterministic system (at precision ε).

Sometimes we might want to know what stochastic processes are obtained if specific observation functions are applied to a deterministic system, and, as explained, the notion of ε-congruence does not help us in answering this question. For this reason, I will now introduce two other meanings of simulation at every observation level which compare measure-preserving deterministic systems as observed with observation functions to stochastic processes. Whether a notion of simulation at every observation level is preferable that (i) is based on the assumption that you cannot compare states which are within distance ε (such as the notion based on Definition 39) or (ii) a notion that tells us what stochastic processes are obtained if specific observation functions are applied to a deterministic system (such as the notion based on Definition 40 or the notion based on Definition 41), will depend on the modeling process and the phenomenon under consideration.

A new meaning based on strong (Φ, ε)-simulation

To introduce the second meaning of simulation at every observation level, I have to start by explaining what it means for a stochastic process and a measure-preserving deterministic system as observed with an observation function Φ to give the same predictions relative to accuracy $\varepsilon > 0$, $\varepsilon \in \mathbb{R}^+$, where ε indicates that we cannot distinguish differences in probabilistic predictions of less than ε. It is plausible that this means that the possible observed values of the measure-preserving deterministic system and the possible outcomes of the stochastic process are the same, and that the probabilistic predictions of the deterministic system as observed with Φ and the probabilistic predictions of the stochastic process are the same or differ by less than ε.

Technically, this idea is captured by the definition of strong (Φ, ε)-simulation; (the reason for 'strong' will become clear soon). Since in practice scientists can only observe finitely many values, I will assume that Φ is a finite-valued observation function.

Definition 40 *A discrete stochastic process* $\{Z_t;\, t \in \mathbb{Z}\}$ strongly (Φ, ε)-simulates *a discrete measure-preserving deterministic system* (M, Σ_M, μ, T) *observed with* Φ*, where* $\Phi : M \to \bar{M}$ *is a surjective finite-valued observation function, if, and only if, there is a surjective measurable function* $\Psi : M \to \bar{M}$ *such that (i)* $Z_t = \Psi(T^t)$ *for all* $t \in \mathbb{Z}$*, and (ii)* $\mu(\{m \in M \mid \Psi(m) \neq \Phi(m)\}) < \varepsilon$*. That a continuous stochastic process* $\{Z_t;\, t \in \mathbb{R}\}$ strongly (Φ, ε)-simulates *a continuous measure-preserving deterministic system* (M, Σ_M, μ, T_t) *observed with* Φ*, where* $\Phi : M \to \bar{M}$ *is a surjective finite-valued observation function, is defined analogously.*

If ε is small enough, the notion of strong (Φ, ε)-simulation captures the idea that in practice the observed measure-preserving deterministic system and the stochastic process give the same predictions. By generalising over Φ and ε, we obtain a plausible meaning of the phrase that stochastic processes of a certain type simulate a measure-preserving deterministic system at any observation level, namely: for every finite-valued observation function Φ and every ε there is a stochastic process of this type which strongly

(Φ, ε)-simulates the deterministic system. This notion of simulation at every observation seems very natural because it allows that the deterministic system is observed with any finite-valued observation function. Yet, to my knowledge, it has not really been discussed in the literature.

A new meaning based on weak (Φ, ε)-simulation

The notion of strong (Φ, ε)-simulation tells us what stochastic process we obtain when we apply an observation function Φ to the measure-preserving deterministic system. This notion can be relaxed by allowing that what you obtain when you observe the measure-preserving deterministic system with an observation function Φ is an *observed* stochastic process. That is, I require that there is a stochastic process and an observation function Γ of this stochastic process such that the stochastic process as observed with Γ gives the same predictions as the measure-preserving deterministic system as observed with Φ for accuracy $\varepsilon > 0$, where $\varepsilon \in \mathbb{R}^+$ (as before, ε indicates that we cannot distinguish differences in probabilistic predictions of less than ε). More specifically, I require that there is an observation of the stochastic process such that the possible observed outcomes of the stochastic process are the possible observed values of the measure-preserving deterministic system, and that the probabilistic predictions of the stochastic process observed with Γ and the probabilistic predictions of the deterministic system observed with Φ are the same or differ by less than ε.

Technically, this idea is captured by the notion of weak (Φ, ε)-simulation. Again, since in practice scientists can only observe finitely many values, I will assume that Φ is a finite-valued observation function.

Definition 41 *A discrete stochastic process* $\{Z_t;\, t \in \mathbb{Z}\}$ weakly (Φ, ε)-simulates *a discrete measure-preserving deterministic system* (M, Σ_M, μ, T) *observed with* Φ*, where* $\Phi : M \to \bar{M}$ *is a surjective finite-valued observation function if, and only if, there is a surjective measurable function* $\Psi : M \to S$ *and a surjective observation function* $\Gamma : S \to \bar{M}$ *such that (i)* $\Gamma(Z_t) = \Psi(T^t)$ *for all* $t \in \mathbb{Z}$*, and (ii)* $\mu(\{m \in M \,|\, \Psi(m) \neq \Phi(m)\}) < \varepsilon$*. That a continuous stochastic process* $\{Z_t;\, t \in \mathbb{R}\}$ weakly (Φ, ε)-simulates *a continuous measure-preserving deterministic system* (M, Σ_M, μ, T_t) *observed* *with* Φ*, where* $\Phi : M \to \bar{M}$ *is a surjective finite-valued observation function, is defined analogously.*

I call it *weak* (Φ, ε)-simulation because if a stochastic process strongly (Φ, ε)-simulates a measure-preserving deterministic system, then it is apparent that it also weakly (Φ, ε)-simulates the deterministic system; (we can simply choose Γ be to be the identity function, that is, we let $\Gamma : S \to S$, $\Gamma(s) = s$). It is also clear that the converse is generally not true.

If ε is small enough, weak (Φ, ε)-simulation captures the idea that the observed stochastic process and the deterministic system as observed with Φ give the same predictions. Again, by generalising over Φ and ε, we obtain a plausible meaning of the phrase that stochastic processes of a certain type simulate a measure-preserving deterministic system at any observation level, namely: for every finite-valued observation function Φ and every ε there is a stochastic process of this type which weakly (Φ, ε)-simulates the deterministic system. To my knowledge, this notion of simulation at every observation level has not been discussed in the literature before.

Compared to the second notion of simulation at every observation level, this third notion only requires that the data could derive from *some observed* stochastic process. For this reason, the second notion might look more attractive. Still, according to all three notions of simulation at every observation level, regardless how the measure-preserving deterministic system is observed, the data could derive from the measure-preserving deterministic system or a stochastic process of a certain type. Hence for all three notions it will be worthwhile to see in the next subsection what results we obtain.

5.4.2 Stochastic processes used in science which simulate deterministic systems used in science at every observation level

The discrete-time case

For a Bernoulli process the next outcome of the process is probabilistically independent of its previous outcome. So, intuitively, it seems clear that *discrete measure-preserving deterministic systems used in science, for which* *the next state of the system is constrained by its previous state (because of the underlying determinism at the level of states), cannot be simulated by Bernoulli processes at every observation level.* Smith (1998, pp. 160–162) also hints at this idea but does not substantiate it with a proof. The following two theorems and the following proposition show that for our three notions of simulation at every observation level (respectively) this idea is indeed correct. Consequently, these results show a limitation on the observational equivalence of discrete measure-theoretic deterministic systems and discrete stochastic processes.

Theorem 3 *Let (M, Σ_M, μ, T) be a discrete measure-preserving deterministic system where Σ_M contains all open balls of the metric space (M, d_M)*[12]*, T is continuous at some point $x \in M$, every open ball around x has positive measure, and there is a set $D \in \Sigma_M$, $\mu(D) > 0$, with $d(T(x), D) = \inf\{d(T(x), m) \,|\, m \in D\} > 0$. Then there is some $\varepsilon > 0$ for which there is no Bernoulli process to which (M, Σ_M, μ, T) is ε-congruent.*

For a proof of this theorem, see subsection 5.7.3. The assumptions of this theorem are very mild and always hold for measure-preserving deterministic systems used in science.

Theorem 4 *Let (M, Σ_M, μ, T) be a discrete measure-preserving deterministic system. Then there is a finite-valued observation function Φ and an $\varepsilon > 0$ such that no Bernoulli process strongly (Φ, ε)-simulates (M, Σ_M, μ, T).*

For a proof, see subsection 5.7.4.

Proposition 1 *Let (M, Σ_M, μ, T) be a discrete measure-preserving deterministic system. Then there is a finite-valued observation function Φ and an $\varepsilon > 0$ such that no Bernoulli process weakly (Φ, ε)-simulates (M, Σ_M, μ, T).*

For a proof of Proposition 1, see subsection 5.7.5.

Given these results, it is natural to ask (which, incidentally, Smith 1998 does not do) whether discrete measure-preserving deterministic systems used

[12] An open ball with centre y and radius $\varepsilon > 0$, $y \in M$, is defined as the set $\{m \in M \mid d(m, y) < \varepsilon\}$.

in science can be simulated at every observation level by other stochastic processes used in science. The answer is 'yes'. Besides, all one needs are Markov processes (Example 5) or multi-step Markov processes (Example 6), which are widely used in science. Markov processes are often regarded as random; in particular, Bernoulli processes are regarded as the most random stochastic processes and Markov processes as the next most random (Eagle 2005; Ornstein & Weiss 1991, p. 38 and p. 66). The following two theorems and one proposition show that *discrete Bernoulli systems (cf. Definition 20) can be simulated at every observation level by irreducible and aperiodic Markov processes or by irreducible and aperiodic multi-step Markov processes* (concerning respectively the three notions defined in subsection 5.4.1).

Theorem 5 *Let* (M, Σ_M, μ, T) *be a discrete Bernoulli system where the metric space* (M, d_M) *is separable*[13] *and where* Σ_M *contains all open balls of* (M, d_M)*. Then for any* $\varepsilon > 0$ *there is an irreducible and aperiodic Markov process such that* (M, Σ_M, μ, T) *is* ε*-congruent to this Markov process.*

For a proof, see subsection 5.7.6. The assumptions in this theorem are fulfilled by all discrete Bernoulli systems used in science.

Theorem 6 *Let* (M, Σ_M, μ, T) *be a discrete Bernoulli system. Then for every finite-valued observation function* Φ *and every* $\varepsilon > 0$ *there is an* n *such that an irreducible and aperiodic Markov process of order* n *strongly* (Φ, ε)*-simulates* (M, Σ_M, μ, T).

For a proof of this theorem, see Radunskaya (1992, chapter 4).

Proposition 2 *Let* (M, Σ_M, μ, T) *be a discrete Bernoulli system. Then for every finite-valued observation function* Φ *and every* $\varepsilon > 0$ *there is an irreducible and aperiodic Markov process which weakly* (Φ, ε)*-simulates* (M, Σ_M, μ, T).

[13] (M, d_M) is *separable* if, and only if, there exists a countable set $\ddot{M} = \{m_n \,|n \in \mathbb{N}\}$ with $m_n \in M$ such that every nonempty open subset of M contains at least one element of $\ddot{M}$.

For a proof of this proposition, see subsection 5.7.7.

For example, consider the baker's system (M, Σ_M, μ, T) (Example 1) with the Euclidean metric d_M. It is a discrete Bernoulli system. Thus for every $\varepsilon > 0$ there is a Markov process such that the baker's system is ε-congruent to this Markov process. And for every finite-valued observation function Φ and every $\varepsilon > 0$ there is an n such that an irreducible and aperiodic Markov process of order n strongly (Φ, ε)-simulates the baker's system. And finally, for every finite-valued observation function Φ and every $\varepsilon > 0$ there is an irreducible and aperiodic Markov process which weakly (Φ, ε)-simulates the baker's system.

Now one might ask whether not only discrete Bernoulli systems but maybe also other discrete measure-preserving deterministic systems used in science can be simulated at every observation level by irreducible and aperiodic Markov processes or by irreducible and aperiodic multi-step Markov processes. As the following theorem (Theorem 7) shows, according to our first notion of simulation at every observation level, indeed only Bernoulli systems can be simulated at every observation level by irreducible and aperiodic Markov processes. For the second and third notion of simulation of every observation level the complete picture is unknown. But I will give two theorems (Theorem 8 and Theorem 9) which show that two important classes of discrete measure-preserving deterministic systems *cannot* be simulated at every observation level by irreducible and aperiodic Markov processes or by irreducible and aperiodic multi-step Markov processes. Namely, these classes are: (i) discrete measure-preserving deterministic systems with zero Kolmogorov-Sinai entropy, and (ii) discrete measure-preserving deterministic systems which are ergodic, which have finite Kolmogorov-Sinai entropy and which are not discrete Bernoulli systems (recall that nearly all deterministic systems in science have finite Kolmogorov-Sinai entropy, see subsection 5.3.1). The classes (i) and (ii) include many discrete deterministic systems used in science, e.g., all discrete versions of integrable Hamiltonian systems, all discrete versions of the motion on clearly non-chaotic regions of KAM-type systems, periodic motion and fixed points (Arnold & Avez 1968, pp. 86–90 and pp. 210–214; Lichtenberg & Lieberman 1992, chapter 3–5; Petersen 1983, p. 245).

So let me first state the theorem about the first notion of simulation at every observation level.

Theorem 7 *The deterministic representation of any irreducible and aperiodic multi-step Markov process (and thus the deterministic representation of any irreducible and aperiodic Markov process) is a discrete Bernoulli system.*

For a proof of this deep theorem, see Ornstein (1974, pp. 45–47). Let (M, Σ_M, μ, T) be a discrete measure-preserving deterministic system, and assume that for all $\varepsilon > 0$ there is an irreducible and aperiodic Markov process which is ε-congruent to (M, Σ_M, μ, T). Then the deterministic representation of any of these Markov processes is isomorphic to (M, Σ_M, μ, T). Hence Theorem 7 implies that (M, Σ_M, μ, T) is a discrete Bernoulli system.

Let me now state the theorems about the second and third notion of simulation at every observation level.

Theorem 8 *Assume that (M, Σ_M, μ, T) is a discrete measure-preserving deterministic system with zero Kolmogorov-Sinai entropy or a discrete ergodic measure-preserving deterministic system with finite Kolmogorov-Sinai entropy which is not a discrete Bernoulli system. Then there is a finite-valued observation function Φ and an $\varepsilon > 0$ such that no irreducible and aperiodic multi-step Markov process strongly (Φ, ε)-simulates (M, Σ_M, μ, T).*

See subsection 5.7.8 for a proof of this theorem.

Theorem 9 *Assume that (M, Σ_M, μ, T) is a discrete measure-preserving deterministic system with zero Kolmogorov-Sinai entropy or a discrete ergodic measure-preserving deterministic system with finite Kolmogorov-Sinai entropy which is not a discrete Bernoulli system. Then there is a finite-valued observation function Φ and an $\varepsilon > 0$ such that no irreducible and aperiodic Markov process weakly (Φ, ε)-simulates (M, Σ_M, μ, T).*

For a proof of this Theorem, see subsection 5.7.9.

The continuous-time case

So far we have only discussed discrete stochastic processes and discrete measure-preserving deterministic systems. What about continuous stochastic processes and continuous measure-preserving deterministic systems? It turns out that analogous results hold here too. Namely, as the following three theorems show, according to our three notions of simulation at every observation level, *continuous Bernoulli systems can be simulated at every observation level by irrationally related semi-Markov processes (Example 7) or by irrationally related multi-step semi-Markov processes (Example 8).*

Theorem 10 *Let (M, Σ_M, μ, T_t) be a continuous Bernoulli system where the metric space (M, d_M) is separable and Σ_M contains all open balls of (M, d_M). Then for any $\varepsilon > 0$ there is an irrationally related semi-Markov process such that (M, Σ_M, μ, T_t) is ε-congruent to this semi-Markov process.*

For a proof of this theorem, see Ornstein & Weiss (1991, pp. 93–94). The assumptions in this theorem are fulfilled by all continuous Bernoulli systems used in science.

Theorem 11 *Let (M, Σ_M, μ, T_t) be a continuous Bernoulli system. Then for every finite-valued observation function Φ and every $\varepsilon > 0$ there is an n such that an irrationally related semi-Markov process of order n strongly (Φ, ε)-simulates (M, Σ_M, μ, T_t).*

For a proof of this theorem, see Ornstein & Weiss (1991, pp. 94–95).

Theorem 12 *Let (M, Σ_M, μ, T_t) be a continuous Bernoulli system. Then for every finite-valued observation function Φ and every $\varepsilon > 0$ there is an irrationally related semi-Markov process $\{Z_t,\, t \in \mathbb{R}\}$ which weakly (Φ, ε)-simulates (M, Σ_M, μ, T_t).*

For a proof of this theorem, see subsection 5.7.10.

For instance, consider a billiard systems with convex obstacles (Example 2) with the Euclidean metric d_M, and recall that it is a continuous Bernoulli system. Hence for every $\varepsilon > 0$ there is an irrationally related semi-Markov process such that the billiard system with convex obstacles is ε-congruent to this semi-Markov process. And it holds that for every finite-valued observation function Φ and every $\varepsilon > 0$ there is an n such that an irrationally related semi-Markov process of order n strongly (Φ, ε)-simulates the billiard system with convex obstacles. And finally, for every finite-valued observation function Φ and every $\varepsilon > 0$ there is an irrationally related semi-Markov process which weakly (Φ, ε)-simulates the billiard system with convex obstacles.

As in the discrete case, you might wonder whether not only continuous Bernoulli systems but maybe also other measure-preserving deterministic systems used in science can be simulated at every observation level by irrationally related semi-Markov processes or by irrationally related multi-step semi-Markov processes. Again, here results analogous to the ones for discrete time can be shown. Namely, as the following theorem (Theorem 13) shows, according to the first notion of simulation at every observation level, only continuous Bernoulli systems can be simulated at every observation level by irrationally related semi-Markov processes. For the second and third notion of simulation of every observation level the complete picture is unknown. But below are two theorems (Theorem 14 and Theorem 15) which show that two important classes of continuous measure-preserving deterministic systems cannot be simulated at every observation level by irrationally related multi-step semi-Markov processes or by irrationally related semi-Markov processes. Namely, these classes are: (i) continuous measure-preserving deterministic systems with zero Kolmogorov-Sinai entropy, and (ii) continuous measure-preserving deterministic systems (M, Σ_M, μ, T_t) with finite Kolmogorov-Sinai entropy which are not continuous Bernoulli systems and where for some $t_0 \in \mathbb{R}$, $t_0 \neq 0$, the discrete deterministic system $(M, \Sigma_M, \mu, T_{t_0})$ is ergodic (recall that nearly all deterministic systems in science have finite Kolmogorov-Sinai entropy, see subsection 5.3.1). The classes (i) and (ii) include many continuous deterministic systems used in science, e.g., all integrable Hamiltonian systems, the motion on clearly non-chaotic regions of KAM-type systems and any periodic motion (Arnold & Avez 1968, pp. 86–90 and pp. 210–214; Lichtenberg & Lieberman 1992, chapter 3–5).

Let me first present the theorem about the first notion of simulation at every observation level.

Theorem 13 *The deterministic representation of every irrationally related multi-step semi-Markov process (and thus the deterministic representation of any irrationally related semi-Markov process) is a continuous Bernoulli system.*

See Park (1982) and Ornstein (1974, pp. 56–61) for a proof of this theorem. Let (M, Σ_M, μ, T_t) be a continuous measure-preserving deterministic system. Assume that for all $\varepsilon > 0$ there is an irrationally related semi-Markov process which is ε-congruent to (M, Σ_M, μ, T_t). Then the deterministic representation of any of these semi-Markov processes is isomorphic to (M, Σ_M, μ, T_t). Consequently, it follows from Theorem 13 that (M, Σ_M, μ, T_t) is a continuous Bernoulli system.

Let me now present the theorems about the second and third notion of simulation at every observation level.

Theorem 14 *Let (M, Σ_M, μ, T_t) be a continuous measure-preserving deterministic system with zero Kolmogorov-Sinai entropy or a continuous measure-preserving deterministic system which is not a continuous Bernoulli system and where for some $t_0 \in \mathbb{R}\backslash\{0\}$ the discrete measure-preserving deterministic system $(M, \Sigma_M, \mu, T_{t_0})$ is ergodic. Then there is a finite-valued observation function Φ and an $\varepsilon > 0$ such that no irrationally related multi-step semi-Markov process strongly (Φ, ε)-simulates (M, Σ_M, μ, T_t).*

For a proof of this theorem, see subsection 5.7.11.

Theorem 15 *Let (M, Σ_M, μ, T_t) be a continuous measure-preserving deterministic system with zero Kolmogorov-Sinai entropy or a continuous measure-preserving deterministic system which is not a continuous Bernoulli system and where for some $t_0 \in \mathbb{R}\backslash\{0\}$ the discrete measure-preserving deterministic system $(M, \Sigma_M, \mu, T_{t_0})$ is ergodic. Then there is a finite-valued observation function Φ and an $\varepsilon > 0$ such that no irrationally related semi-Markov process weakly (Φ, ε)-simulates (M, Σ_M, μ, T_t).*

See subsection 5.7.12 for a proof of this theorem.

To summarise the most important points: the results of this section show that discrete Bernoulli systems can be simulated at every observation level by irreducible and aperiodic Markov processes and by irreducible and aperiodic multi-step Markov processes, respectively. And continuous Bernoulli systems can be simulated at every observation level by irrationally related semi-Markov processes and by irrationally related multi-step semi-Markov processes, respectively. Recall that Markov processes, multi-step Markov processes, semi-Markov processes and multi-step semi Markov processes are widely used in science to model phenomena. Also recall that several discrete deterministic systems used in science are discrete Bernoulli systems and that several continuous deterministic systems used in science are continuous Bernoulli systems (see subsection 5.3.1). Consequently, I conclude that the conjecture advanced at the beginning of this subsection is wrong: *it is possible to simulate measure-theoretic deterministic systems used in science at every observation level by stochastic processes used in science;* sometimes even by Markov processes, which are regarded as the next most random stochastic processes after Bernoulli processes. *All this shows that even kinds of stochastic processes and kinds of deterministic systems which intuitively seem to give very different predictions can be observationally equivalent.*

5.5 Previous philosophical discussion

Let me discuss the previous philosophical papers about the topic of this chapter that I have been able to find. Suppes & de Barros (1996) and Suppes (1999) discuss an instance of Theorem 5, namely that for discrete versions of billiard systems with convex obstacles and for every $\varepsilon > 0$ there is a Markov process such that the billiard system is ε-congruent to this Markov process. Suppes (1993) (albeit with only half a page on the topic of this chapter) and Winnie (1998) discuss the theorem that for continuous Bernoulli systems and for every $\varepsilon > 0$ there is an irrationally related semi-Markov process which is ε-congruent to the deterministic system (Theorem 10). And Hoefer's (2008) entry briefly summarises and comments on the debate between Suppes (1993) and Winnie (1998).

My discussion of the previous philosophy literature will focus on three issues: the significance of Theorem 5 and Theorem 10, the role of chaos in results on observational equivalence, and the question of whether the deterministic or the stochastic description is the better one. Let me start with the first issue.

5.5.1 The significance of Theorem 5 and Theorem 10

Suppes & de Barros (1996, p. 196), Suppes (1999, pp. 181–182) and Winnie (1998, p. 317) claim that the philosophical significance of Theorem 10 and of the above-mentioned instance of Theorem 5 is that for chaotic motion and every observation level one can choose between a deterministic description used in science and a stochastic description. For instance, Suppes & de Barros (1996, p. 196) comment on the significance of these results:

> What is fundamental is that independent of this variation of choice of examples or experiments is that [*sic*] when we do have chaotic phenomena [...] then we are in a position to choose either a deterministic or stochastic model.

However, I submit that these claims are weak, and Theorem 5 and Theorem 10 show more. As discussed in subsection 5.2.1, the basic results on observational equivalence already show that for many measure-preserving deterministic systems, including several deterministic systems used in science, the following holds: for every finite-valued observation function one can choose between a nontrivial stochastic description or a deterministic description (cf. Theorem 1 and Theorem 2). And as one would expect, the following two propositions show that this implies the following: according to our first notion of simulation at every observation level, many deterministic systems, namely all those to which either Theorem 1 or Theorem 2 applies and which additionally have finite Kolmogorov-Sinai entropy, can be simulated at every observation level by nontrivial stochastic processes. (As discussed in subsection 5.3.1, nearly all measure-preserving deterministic systems used in science have finite Kolmogorov-Sinai entropy).

Proposition 3 *Let (M, Σ_M, μ, T) be a discrete measure-preserving deterministic system where (M, d_M) is separable and where Σ_M contains all open balls of (M, d_M). Assume that (M, Σ_M, μ, T) satisfies the assumption of Theorem 1 and has finite Kolmogorov-Sinai entropy. Then for every $\varepsilon > 0$ there is a stochastic process $\{Z_t;\, t \in \mathbb{Z}\}$ with outcome space $\bar{M} = \cup_{l=1}^{h} o_l$, $h \in \mathbb{N}$, such that $\{Z_t;\, t \in \mathbb{Z}\}$ is ε-congruent to (M, Σ_M, μ, T), and for all $k \in \mathbb{N}$ there is an outcome $o_i \in \bar{M}$ such that for all $o_j \in \bar{M}$, $1 \leq j \leq h$, $P\{Z_{t+k} = o_j \mid Z_t = o_i\} < 1$.*

This proposition is easy to establish. For a proof, see subsection 5.7.13.

Proposition 4 *Let* (M, Σ_M, μ, T_t) *be a continuous measure-preserving deterministic system where* (M, d_M) *is separable and where* Σ_M *contains all open balls of* (M, d_M)*. Assume that* (M, Σ_M, μ, T_t) *satisfies the assumption of Theorem 2 and has finite Kolmogorov-Sinai entropy. Then for every* $\varepsilon > 0$ *there is a stochastic process* $\{Z_t;\, t \in \mathbb{R}\}$ *with outcome space* $M_O = \cup_{l=1}^{h} o_l$*,* $h \in \mathbb{N}$*, such that* $\{Z_t;\, t \in \mathbb{R}\}$ *is* ε*-congruent to* (M, Σ_M, μ, T_t)*, and for all* $k \in \mathbb{R}^+$ *there is an outcome* $o_i \in M_O$ *such that for all* $o_j \in M_O$*,* $1 \leq j \leq h$*,* $P\{Z_{t+k} = o_j \mid Z_t = o_i\} < 1$*.*

Again, this proposition is easy to establish. See subsection 5.7.14 for a proof.

Also, clearly, the basic results immediately imply the following for the second and third notion of simulation at every observation level: every measure-preserving deterministic system to which Theorem 1 or Theorem 2 applies, and thus many measure-preserving deterministic systems (including deterministic systems used in science), can be simulated at every observation level by nontrivial stochastic processes. This is so because the definition of the second or third notion of simulation at every observation level quantifies over all finite-valued observation functions Φ. Given a finite-valued observation function Φ and a discrete measure-preserving deterministic system (M, Σ_M, μ, T), the stochastic process $\{\Phi(T^t); t \in \mathbb{Z}\}$ is nontrivial by Theorem 1. And, obviously, $\{\Phi(T^t); t \in \mathbb{Z}\}$ strongly (Φ, ε)-simulates (M, Σ_M, μ, T) and weakly (Φ, ε)-simulates (M, Σ_M, μ, T). Likewise, given a continuous measure-preserving deterministic system (M, Σ_M, μ, T_t), the stochastic process $\{\Phi(T_t); t \in \mathbb{R}\}$ is nontrivial by Theorem 2, and $\{\Phi(T_t); t \in \mathbb{R}\}$ strongly (Φ, ε)-simulates (M, Σ_M, μ, T_t) and weakly (Φ, ε)-simulates (M, Σ_M, μ, T_t). Hence all measure-preserving deterministic systems to which Theorem 1 or Theorem 2 applies are simulated at every observation level by nontrivial stochastic processes.

And similar results for chaotic systems were known long before the ε-congruence results were proved (cf. subsection 5.2.1). *Hence the fact that at every observation level one has a choice between a measure-preserving deterministic system used in science and a stochastic process was known long before the* ε*-congruence results (the instance of Theorem 5 and Theorem 10)* *were proved; and so this cannot be the philosophical significance of these results as claimed by these authors.* As I have argued in subsection 5.4.1, the significance of these results is something stronger: namely that it is possible to simulate measure-preserving deterministic systems used in science at every observation level by stochastic processes used in science.

Moreover, Suppes & de Barros (1996, p. 196–198) and Suppes (1999, p. 189 and p. 192) wrongly think that what it means for a measure-preserving deterministic system to be ε-congruent to a certain type of stochastic process for every $\varepsilon > 0$ (the first notion of simulation at every observation level) is the following: the deterministic system observed with any finite-valued observation function yields a stochastic process of a certain type (that is, something like my second notion of simulation at every observation level). As discussed in subsection 5.4.1, the first and the second notion of simulation at every observation level are quite different (for instance, only the latter tells us what happens if we apply any arbitrary observation function to a deterministic system). And in particular, as we have seen in subsection 5.4.2, the first and second notion give rise to different results.[14]

There is hardly any conceptual or philosophical discussion in the mathematics literature on those mathematical results presented in this chapter which were already proven before. The main exception is the following comment by Ornstein & Weiss (1991, pp. 39–40):

> Our theorem [Theorem 10] also tells us that certain semi-Markov systems could be thought of as being produced by Newton's laws (billiards seen through a deterministic viewer) or by coin-flipping. This may mean that there is no philosophical distinction between

[14] The reader should also be warned that there are several technical lacunae in Suppes & de Barros (1996) and Suppes (1999). For instance, according to their definition, any two measure-preserving deterministic systems whatsoever are ε-congruent (let the metric space simply consist of one element). Also, these authors do not seem to be aware that the results about simulation at every observation level by semi-Markov processes (Theorem 10) require the measure-preserving deterministic system to be a Bernoulli system and so do not generally hold for ergodic measure-preserving deterministic systems. And in these papers it is wrongly assumed that the notion of isomorphism requires that the measure-preserving deterministic system is looked at through a finite-valued observation function (Suppes & de Barros 1996, p. 198; Suppes 1999, pp. 189–192).

> processes governed by roulette wheels and processes governed by Newton's laws. {The popular literature emphasizes the distinction between "deterministic chaos" and "real randomness".} In this connection we should note that our model for a stationary process (§ 1.2) [the deterministic representation] means that random processes have a deterministic model. This model, however, is abstract, and there is no reason to believe that it can be endowed with any special additional structure. Our point is that we are comparing, in a strong sense, Newton's laws and coin flipping.[15]

It is hard to tell what this comment expresses because it is vague and unclear. For instance, why do Ornstein & Weiss highlight coin flipping even though Theorem 10 does not tell us anything about Bernoulli processes but only about semi-Markov processes? Disregarding that, possibly, Ornstein and Weiss think that semi-Markov processes are random and hence this comment expresses that deterministic systems as well as stochastic processes can be random. This is true and in fact widely acknowledged in the philosophy literature (e.g., Eagle 2005). Or maybe Ornstein & Weiss want to say that measure-preserving deterministic systems used in science, when observed with specific observation functions, can be observationally equivalent to stochastic processes used in science or, if semi-Markov processes are random, even random stochastic processes.[16] This is true and an important insight. Yet, as discussed in subsection 5.3.1, this insight was generally known before Theorem 10 and related results were proven, and it has been established by theorems which are weaker than Theorem 10. One might have expected Ornstein & Weiss to say that Theorem 10 shows that measure-preserving deterministic systems used in science can be simulated at every observation level by stochastic processes used in science (cf. subsection 5.4.2). But they do not seem to say this here: because, if they did,

[15] The text enclosed in braces is in a footnote.

[16] As explained in subsection 5.4.1, if a continuous measure-preserving deterministic system (M, Σ_M, μ, T_t) and a semi-Markov process $\{Z_t;\ t \in \mathbb{R}\}$ are ε-congruent, then there is a finite-valued observation function Φ such that $\{\Phi(T_t);\ t \in \mathbb{R}\}$ is the same semi-Markov process.

it would be unclear why the deterministic representation is mentioned; and also they do not talk about all possible observation levels.

In any case, it goes without saying that even if Theorem 10 shows that deterministic and stochastic descriptions are observationally equivalent in some sense, it is not true that "this may mean that there is no philosophical distinction between processes governed by roulette wheels and processes governed by Newton's laws" in the sense that this may mean that there is no conceptual distinction between a deterministic description and a stochastic description (as a kind of indeterministic description). Regardless of any results on observational equivalence, there will remain this conceptual distinction.

5.5.2 The role of chaotic behaviour

Let us now turn to the second issue, namely the role of chaos in results on observational equivalence. Hoefer (2008) is not aware, and Suppes & de Barros (1996), Suppes (1999) and Winnie (1998) do not seem to be aware, that *also for non-chaotic systems there is a choice between a deterministic and a stochastic description (at every observation level).* To show this, it will suffice to show that Theorem 1 also applies to deterministic systems which are uncontroversially neither chaotic nor chaotic restricted to a region of phase space. Consider the measure-preserving deterministic system (M, Σ_M, μ, T) where $M = [0, 1)$ represents the unit circle, i.e., each $m \in M$ represents the point $e^{2\pi mi}$, Σ_M is the Lebesgue σ-algebra on M, μ is the Lebesgue measure, and T is the rotation $T(m) = m + \alpha \pmod 1$, where $\alpha \in \mathbb{R}$ is irrational. (M, Σ_M, μ, T) is called an *irrational rotation on the circle.* It is uncontroversial that this measure-preserving deterministic system is neither chaotic nor chaotic on a region of phase space because all solutions are stable, i.e., nearby solutions stay close for all times. However, one easily sees that it satisfies the assumption of Theorem 1.[17] Consequently, for any nontrivial

[17] Any irrational rotation on the circle is ergodic (cf. Definition 2.5) (Petersen 1983, p. 49). Hence there can be no $n \in \mathbb{N}$ and $C \in \Sigma_M$, $0 < \mu(C) < 1$, such that, except for a set of measure zero, $T^n(C) = C$ since this would imply that there is an irrational rotation on the circle which is not ergodic.

finite-valued observation function the measure-preserving deterministic system (M, Σ_M, μ, T) yields a nontrivial stochastic process.

Furthermore any irrational rotation on a circle has zero entropy (Petersen 1983, p. 245). Thus, according to any of our three notions of simulation at every observation level, any irrational rotation (M, Σ_M, μ, T) is simulated at every observation level by nontrivial stochastic processes (see Proposition 3 and the paragraph following this proposition).[18]

5.5.3 Is the deterministic or the indeterministic description better?

Let me now turn to the third issue, namely if there is a choice between a deterministic and a stochastic description, which one is better or preferable? In a way, if you aim to describe the world at a specific level, it is uncontroversial that if the phenomenon under consideration is really stochastic at this level, the stochastic description is preferable; and if the phenomenon is really deterministic at this level, the deterministic description is preferable.

But really of concern here is the question of which description is preferable when you cannot know for sure whether the phenomenon is deterministic or stochastic. So which description is then preferable in the sense of being preferable relative to our current knowledge and evidence? This question has not been the topic of this chapter. Rather, the topic of this chapter has been whether measure-theoretic deterministic systems and stochastic processes are observationally equivalent, and whether even kinds of stochastic processes and kinds of deterministic systems which intuitively seem to give very different predictions can be observationally equivalent. Still, this question arises from our discussion, and so I will address it. Because of lack of space, it will not be possible to treat the question in all its details. But I will criticise the previous literature about this question, namely Hoefer (2008), Suppes (1993) and Winnie (1998), and I will conclude that a more careful treatment is needed.

[18]This example can be generalised: any rationally independent rotation on a torus is uncontroversially non-chaotic but fulfills the assumption of Theorem 1 (cf. Petersen 1983, p. 51).

Before I turn to the previous literature on this question, note the following. Consider a discrete measure-theoretic deterministic system (M, Σ_M, μ, T) or a continuous measure-theoretic deterministic system (M, Σ_M, μ, T_t), and consider an observation function $\Phi : M \to M_O$ which is many to one. Then the deterministic description ((M, Σ_M, μ, T) or (M, Σ_M, μ, T_t) observed with Φ) is more informative than the stochastic description ($\{Z_t = \Phi(T^t);\, t \in \mathbb{Z}\}$ or $\{Z_t = \Phi(T_t);\, t \in \mathbb{R}\}$) in the following sense: while (M, Σ_M, μ, T) or (M, Σ_M, μ, T_t) tells us where each state $m \in M$ evolves, $\{Z_t;\, t \in \mathbb{Z}\}$ or $\{Z_t;\, t \in \mathbb{R}\}$ only gives us the probability distributions over all possible sequences of outcomes in M_O. Yet this extra information might not be desirable or relevant as, for instance, for the deterministic representation. Thus, suppose you have a choice between a stochastic process and its deterministic representation. Even though the deterministic representation is more informative in this sense, you might argue that the stochastic process is preferable because, from a philosophical perspective, the deterministic representation is a cheat (cf. subsection 5.2.2). Thus the fact that, in this sense, the deterministic description is more informative than the stochastic one does not imply that the deterministic description is the better description.

Let me now consider arguments in the literature which purport to show that by observing a phenomenon at different observation levels, you can find out that the measure-preserving deterministic system is the correct description. Consider the following claim by Hoefer (2008):

> It may well be true that there are some deterministic dynamical systems that, *when viewed properly*, display behavior indistinguishable from that of a genuinely stochastic process. For example, using the billiard table above [a billiard system with convex obstacles], if one divides its surface into quadrants and looks at which quadrant the ball is in at 30-second intervals, the resulting sequence is no doubt highly random. But this does not mean that the same system, when viewed in a *different* way (perhaps at a higher degree of precision) does not cease to look random and instead betrays its deterministic nature [original emphasis].[19]

[19] Hoefer (2008) uses the word 'random' synonymously to 'stochastic'.

Our previous discussion shows that this claim is misguided for two reasons. First, for any discretised version of any billiard system with convex obstacles every finite-valued observation function yields a nontrivial stochastic process (cf. Theorem 1). Hence *there will never be trivial transition probabilities, contrary to what Hoefer suggests*. Second, assume that the stochastic process $\{\Phi(T^t);\, t \in \mathbb{Z}\}$, where (M, Σ_M, μ, T) is a discrete measure-theoretic deterministic system and Φ is an observation function, is in accordance with the observations and is trivial (the transition probabilities are zero or one). Or assume that the discrete stochastic process $\{\Phi(T^t_{t_0});\, t \in \mathbb{Z}\}$, where (M, Σ_M, μ, T_t) is a continuous measure-theoretic deterministic system, Φ is an observation function and $t_0 \in \mathbb{R}^+$, is in accordance with the observations and is trivial. *This does not imply, as the quote suggests, that the observations derive from a deterministic system.* As argued, trivial stochastic processes can also derive from observing a nontrivial stochastic process (cf. the end of subsection 5.2.1).

Another argument in this direction has been put forward by Winnie (1998).[20] For the baker's system (M, Σ_M, μ, T) (Example 1) we consider the relation between two observations on the system. Consider the observation function $\Phi(m) = o_1\chi_{\alpha_1}(m) + o_2\chi_{\alpha_2}(m)$ where $\alpha_1 = [0,1] \times [0,1/2] \setminus D, \alpha_2 = [0,1] \times (1/2,1] \setminus D$ and consider the observation function $\Psi(m) = \sum_{i=1}^{4} q_i\chi_{\beta_i}$ where $\beta_1 = [0,1/2] \times [0,1/2] \setminus D, \beta_2 = (1/2,1] \times [0,1/2] \setminus D, \beta_3 = [0,1/2] \times (1/2,1] \setminus D, \beta_4 = (1/2,1] \times (1/2,1] \setminus D$. It is clear that if you observe q_1 (with Ψ), the probability that you will next observe o_1 (with Φ) is 1; if you observe q_2, the probability that you will next observe o_2 is 1; if you observe q_3, the probability that you will next observe o_1 is 1; and if you observe q_4, the probability that you will next observe o_2 is 1. Thus there are trivial transition probabilities from the observation modeled by Ψ to the coarser observation modeled by Φ. Winnie (1998, pp. 314–315) comments on this:

[20] Winnie (1998) does not clearly distinguish between random and stochastic behaviour as a form of indeterministic behaviour. As a consequence, the discussion sometimes suffers from ambiguities. It is uncontroversial that stochastic processes are processes governed by probabilistic laws. Random behaviour is usually regarded as different from stochastic behaviour, but there are various different accounts about what randomness amounts to (see, for instance, the recent survey Eagle 2005).

> Thus, the fact that a chaotic deterministic system [...] has *some* partitioning that yields a set of random or stochastic observations in no way undermines the distinction between deterministic and stochastic behaviour for such systems. [...] As successive partitionings are exemplified [...] the determinism underlying the preceding, coarser observations emerges. To be sure, at any state of the above process, the system may be modeled stochastically, but the successive stages of that modeling process provide ample—inductive—reason for believing that the deterministic model is correct [original emphasis].

In order to understand this quote, note the following. From the fact that, in the discrete case, there are trivial transition probabilities from an observation (modeled by Ψ) to a coarser observation (modeled by Φ), or that, in the continuous case, there are trivial transition probabilities from an observation (modeled by Ψ) to a coarser observation (modeled by Φ) when the observations are made at the time points nt_0, $n \in \mathbb{Z}$, $t_0 \in \mathbb{R}^+$, it does not follow that the observed phenomenon is deterministic and Winnie also does not claim this. It may well be that $\{\Psi(T^t);\, t \in \mathbb{Z}\}$ or $\{\Psi(T_t);\, t \in \mathbb{R}\}$, or any stochastic process at a smaller scale, really governs the phenomenon under consideration.

The argument Winnie seems to make in the quote is the following. Assume that you can make observations at finer levels (that is, observations where there is at least one value of the coarser observation function such that two or more values of the finer observation function corresponds to one observed value of the coarser observation function). Further, assume that you find that for observations at finer levels you need stochastic processes at a smaller scale to explain the observational data (that is, stochastic processes where there is at least one outcome of the stochastic process at a larger scale such that two or more outcomes of the stochastic process at a smaller scale corresponds to one outcome of the stochastic process at a larger scale). Then this provides inductive evidence that the phenomenon under consideration is deterministic, and hence that the deterministic description is better. Let me call this argument the ‘nesting argument’. I think that, unlike suggested by Winnie’s quote, *the nesting argument is independent of whether* *there are trivial transition probabilities from an observation to a coarser observation.* For instance, consider again the baker's system (M, Σ_M, μ, T) (Example 1). Let the observation function $\Psi(m) = \sum_{i=1}^{4} q_i \chi_{\beta_i}(m)$ be as above, and consider the observation function $\Phi(m) = o_1 \chi_{\gamma_1}(m) + o_2 \chi_{\gamma_2}(m)$ where $\gamma_1 = [0, 1/2] \times [0, 1] \setminus D, \gamma_2 = (1/2, 1] \times [0, 1] \setminus D$. Clearly, for all i, $1 \leq i \leq 4$, and all j, $1 \leq i \leq 2$, the probability that q_i will be followed by o_j is 1/2. Still Φ is coarser that Ψ, and all that matters for the nesting argument is that for observations at finer levels you need stochastic processes at a smaller scale to explain the data.

Before I continue the discussion on the nesting argument, let me mention another view in the literature about which description is preferable. Namely, Suppes (1993, p. 254), without providing any arguments, simply claims that if there is a choice between a deterministic description used in science and a stochastic description, both descriptions are equally good. And Winnie presents the nesting argument also as a criticism of this claim by Suppes.

I want to argue that neither Suppes (1993) nor Winnie's (1998) view is tenable. Note that both Suppes and Winnie's claims are very general and are not based on any arguments about the state of the art of which scientific theories best describe the observed phenomena or which interpretation of a scientific theory is correct. Thus to refute these claims, it will suffice to show that there *could* be situations in science (regardless of whether this is the current situation in science) where (contra Suppes) *not* both descriptions are equally good and where the premises of the nesting argument are true but where (contra Winnie) *not* the deterministic description is preferable.

As already pointed out above, in a way, if the aim is to describe the world at a specific level and if the phenomenon under consideration is really stochastic at this level, the stochastic description is preferable, even if the stochasticity is at a very small scale, and thus you find that for observations at a finer level you need stochastic processes at a smaller scale to explain the data. Likewise, if the phenomenon is really deterministic at this level, the deterministic description is preferable. But really of concern is the following question: which description is preferable in the sense of being preferable relative to our current knowledge and evidence?

Before I can explain why I think that also here neither Winnie's nor Suppes' view is tenable, let me point out that an answer to this question depends on many factors, such as the kind of phenomenon under consideration, the state of the art of scientific theories, the metaphysical predilections and, as part of this, the views about how models relate to reality.

For instance, first, a stochastic description can be preferable if the following holds: there is no theory from which the deterministic description is derivable; the stochastic description is derivable from a well-confirmed theory T; there is evidence which is not derivable from the specific deterministic or stochastic description but which confirms the stochastic theory T and hence provides evidence for the stochastic description. Or second, suppose that a discrete measure-theoretic deterministic system (M, Σ_M, μ, T) or a continuous measure-theoretic deterministic system (M, Σ_M, μ, T_t) can be be derived from Newton's equations of motion. And suppose that there is confusion about the more fundamental theory but there is the general consensus that it might well be that in reality there is a stochastic process at a small scale of the form $\{\Phi(T^t); t \in \mathbb{Z}\}$ or $\{\Phi(T_t); t \in \mathbb{R}\}$. Because it is unknown which exact stochastic process might be an alternative description, the scientist might reasonably decide to work with the deterministic description.

The first example, i.e. that a well-confirmed theory suggest that stochastic process is correct and hence that the stochastic description is preferable, *provides a counterexample to both Suppes's (1993) and Winnie's (1998) claims.* Here the stochastic process which is believed to be the real one might be at a very small scale, and thus you find that for observations at a finer level you need stochastic processes at a smaller scale to explain the data. That is, the premises of the nesting argument are true (but the conclusion is not).

At one point in the text Winnie (1998, p. 318) says that if there were some in principle limitations on observational accuracy, then the deterministic description might not be the better one. But he quickly dismisses this thought, arguing that the deterministic descriptions in dynamical systems theory are deterministic descriptions in Newtonian mechanics and there are no in principle limitations on observational accuracy in Newtonian mechanics. But this misses the point: even if there are no such limitations in Newtonian mechanics, there might be, or there might be evidence for, in principle limitations on observational accuracy in the actual world; for instance, because in the actual world the phenomenon is governed, or believed to be governed, by a stochastic process at a very small scale.[21]

To conclude, the question of whether the deterministic or the stochastic description is preferable depends on many factors. Neither Hoefer's (2008), Suppes' (1993) nor Winnie's (1998) view is tenable, and a more careful treatment of this question is needed.

5.6 Conclusion

The central question of this chapter has been: are deterministic and indeterministic descriptions observationally equivalent in the sense that deterministic descriptions, when observed, and indeterministic descriptions give the same predictions?

After some introductory remarks, in section 5.2 I demonstrated that every stochastic process is observationally equivalent to a measure-theoretic deterministic system, and that many measure-theoretic deterministic systems are observationally equivalent to stochastic processes; and I formally defined what it means for a measure-preserving deterministic system, observed with an observation function, and a stochastic process to be observationally equivalent. Still, one might guess that the measure-theoretic deterministic systems which are observationally equivalent to stochastic processes used in science do not include any measure-theoretic deterministic systems used in science. In section 5.3 I showed this to be false because some discrete measure-theoretic deterministic systems used in science even produce Bernoulli processes and some continuous measure-theoretic deterministic systems even produce semi-Markov processes. Despite this, one might guess that measure-theoretic deterministic systems used in science cannot give the same predictions at every observation level as stochastic processes used in science. I have introduced three plausible technical notions of simulation at every observation level. In section 5.4 I showed that there is indeed a limitation on observational equivalence, namely discrete measure-preserving deterministic systems used

[21]Furthermore, dynamical systems theory is applied not only in Newtonian mechanics but in many other scientific fields. Hence Winnie would have to extend his argument to all the other applications of dynamical systems theory.

in science cannot give the same predictions at every observation level as Bernoulli processes. However, the guess is still wrong because I have shown the following: several discrete measure-theoretic deterministic systems used in science give the same predictions at every observation level as Markov processes or multi-step Markov processes; and several continuous measure-theoretic deterministic systems used in science, including Newtonian systems, give the same predictions at every observation level as semi-Markov processes or multi-step semi-Markov processes. The general insight of all these results is that even kinds of deterministic systems and kinds of stochastic processes which, intuitively, seem to give very different predictions, are observationally equivalent. Finally, in section 5.5 I criticised the previous philosophical literature. Suppes & de Barros (1996), Suppes (1999) and Winnie (1998) argue that the philosophical significance of the result which says that some continuous measure-preserving deterministic systems can be simulated at every observation level by semi-Markov processes is that for chaotic motion one can choose at every observation level between a stochastic or a deterministic description. However, this is already shown by the basic results in section 5.2. The philosophical significance of these results is really something stronger, namely that there are measure-preserving deterministic systems used in science that give the same predictions at every observation level as stochastic processes used in science. Moreover, these authors seem not to be aware that there are also uncontroversially non-chaotic deterministic systems which can be simulated at every observation level by nontrivial stochastic processes. Furthermore, I argued that the viewpoints in the literature on the question of whether the deterministic or the stochastic description is preferable, namely Hoefer (2008), Suppes (1993), Winnie (1998), are untenable. I concluded that this question needs more careful consideration.

5.7 Appendix: Proofs

5.7.1 Proof of Theorem 1

Theorem 1 *If, and only if, for the discrete measure-preserving deterministic system* (M, Σ_M, μ, T) *there does not exist an* $n \in \mathbb{N}$ *and a* $C \in \Sigma_M$, $0 <$

$\mu(C) < 1$, such that, except for a set of measure zero, $T^n(C) = C$, then the following holds: for every nontrivial finite-valued observation function $\Phi : M \to M_O$, $M_O = \cup_{l=1}^{r} o_l$, $r \in \mathbb{N}$, every $k \in \mathbb{N}$ and the stochastic process $\{Z_t = \Phi(T^t);\ t \in \mathbb{Z}\}$ there is an $o_i \in M_O$ such that for all $o_j \in M_O$, $P\{Z_{t+k}{=}o_j \mid Z_t{=}o_i\} < 1$.

Proof: Notice that it suffices to prove the following:

> ($*$) If, and only if, for (M, Σ_M, μ, T) it is not the case that there exists an $n \in \mathbb{N}$ and a $C \in \Sigma_M$, $0 < \mu(C) < 1$, such that, except for a set of measure zero (esmz.), $T^n(C) = C$, then the following holds: for any nontrivial partition $\alpha = \{\alpha_1, \ldots, \alpha_r\}$, $r \in \mathbb{N}$, and any $k \in \mathbb{N}$ there is an $i \in \{1, \ldots, r\}$ such that for all j, $1 \le j \le r$, $\mu(T^k(\alpha_i) \setminus \alpha_j) > 0$.

Recall that finite-valued observation functions are of the form $\sum_{l=1}^{r} o_l \chi_{\alpha_l}(m)$, where $\alpha = \{\alpha_1, \ldots, \alpha_r\}$ is a partition and $M_O = \cup_{l=1}^{r} o_l$ (cf. subsection 5.2.1). Hence the conclusion of ($*$) says that for any nontrivial finite-valued observation function $\Phi : M \to M_O$ and any $k \in \mathbb{N}$ there is an outcome $o_i \in M_O$ such that for all possible outcomes $o_j \in M_O$ it holds that $P\{Z_{t+k} = o_j \mid Z_t = o_i\} < 1$, $t \in \mathbb{Z}$ arbitrary.

$\Leftarrow$: Assume that there is an $n \in \mathbb{N}$ and a $C \in \Sigma_M$, $0 < \mu(C) < 1$, such that, esmz., $T^n(C) = C$. Then for the partition $\alpha = \{C, M \setminus C\}$ we have $\mu(T^n(C) \setminus C) = 0$ and $\mu(T^n(M \setminus C) \setminus (M \setminus C)) = 0$.

$\Rightarrow$: So assume that the conclusion of ($*$) does not hold, i.e., there exists a nontrivial partition α and a $k \in \mathbb{N}$ such that for each α_i there exists an α_j with, esmz., $T^k(\alpha_i) \subseteq \alpha_j$. Now recall the definition of a deterministic system being ergodic (Definition 35). It can be shown (cf. Petersen 1983, section 2.4) that a a discrete measure-preserving deterministic system (M, Σ_M, μ, T) is ergodic if, and only if, for all $A, B \in \Sigma_M$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} (\mu(T^n(A) \cap B) - \mu(A)\mu(B)) = 0. \tag{5.3}$$

As already pointed out, the assumption that there exists an $n \in \mathbb{N}$ and a $C \in \Sigma_M$, $0 < \mu(C) < 1$, such that, esmz., $T^n(C) = C$ implies that (M, Σ_M, μ, T^k) is ergodic for all $k \in \mathbb{N}$.

Case 1: For all i there is a j such that, esmz., $T^k(\alpha_i) = \alpha_j$. Then ergodicity of (M, Σ_M, μ, T^k) (equation (5.3)) implies that there is an $h \in \mathbb{N}$ such that, esmz., $T^{kh}(\alpha_1) = \alpha_1$. But this contradicts the assumption that it is not the case that there exists an $n \in \mathbb{N}$ and a $C \in \Sigma_M$, $0 < \mu(C) < 1$, such that, esmz., $T^n(C) = C$.

Case 2: For some i there is a j with, esmz., $T^k(\alpha_i) \subset \alpha_j$ and $\mu(\alpha_i) < \mu(\alpha_j)$. Ergodicity of (M, Σ_M, μ, T^k) (equation (5.3)) implies that there exists a $h \in \mathbb{N}$ such that, esmz., $T^{hk}(\alpha_j) \subseteq \alpha_i$. Hence it holds that $\mu(\alpha_j) \leq \mu(\alpha_i)$, yielding a contradiction, viz. $\mu(\alpha_i) < \mu(\alpha_j) \leq \mu(\alpha_i)$.

5.7.2 Proof of Theorem 2

Theorem 2 *If, and only if, for the continuous measure-preserving deterministic system (M, Σ_M, μ, T_t) there does not exist a $n \in \mathbb{R}^+$ and a $C \in \Sigma_M$, $0 < \mu(C) < 1$, such that, except for a set of measure zero, $T_n(C) = C$, then the following holds: for every nontrivial finite-valued observation function $\Phi : M \to M_O$, $M_O = \cup_{l=1}^r o_l$, $r \in \mathbb{N}$, every $k \in \mathbb{R}^+$ and the stochastic process $\{Z_t = \Phi(T_t);\, t \in \mathbb{R}\}$ there is an outcome $o_i \in M_O$ such that for all possible outcomes $o_j \in M_O$, $P\{Z_{t+k} = o_j \mid Z_t = o_i\} < 1$.*

Proof: This proof uses the same ideas as the proof for the analogous discrete-time result (Theorem 1). It suffices to prove the following:

> $(**)$ If, and only if, for (M, Σ_M, μ, T_t) there does not exist an $n \in \mathbb{R}^+$ and a $C \in \Sigma_M$, $0 < \mu(C) < 1$, such that, esmz., $T_n(C) = C$, then the following holds: for any nontrivial partition $\alpha = \{\alpha_1, \ldots, \alpha_r\}$, $r \in \mathbb{N}$, and all $k \in \mathbb{R}^+$ there is an $i \in \{1, \ldots, r\}$ such that for all j, $1 \leq j \leq r$, $\mu(T_k(\alpha_i) \setminus \alpha_j) > 0$.

Recall that finite-valued observation functions are of the form $\sum_{l=1}^r o_l \chi_{\alpha_l}(m)$, where $\alpha = \{\alpha_1, \ldots, \alpha_r\}$ is a partition and $M_O = \cup_{l=1}^r o_l$ (cf. subsection 5.2.1). Consequently, the conclusion of $(**)$ expresses that for any nontrivial finite-valued observation function $\Phi : M \to M_O$ and all $k \in \mathbb{R}^+$ there is an outcome

$o_i \in M_O$ such that for all outcomes $o_j \in M_O$, $P\{Z_{t+k} = o_j \mid Z_t = o_i\} < 1$.

$\Leftarrow$: Assume that there is an $n \in \mathbb{R}^+$ and a $C \in \Sigma_M$, $0 < \mu(C) < 1$, such that, esmz., $T_n(C) = C$. Then for the partition $\alpha = \{C, M \setminus C\}$ it holds that $\mu(T_n(C) \setminus C) = 0$ and $\mu(T_n(M \setminus C) \setminus (M \setminus C)) = 0$.

$\Rightarrow$: So assume that the conclusion of (**) does not hold, and hence that there is a nontrivial partition α and a $k \in \mathbb{R}^+$ such that for each α_i there is an α_j with, esmz., $T_k(\alpha_i) \subseteq \alpha_j$. From the assumptions it follows that for every $k \in \mathbb{R}^+$ the discrete measure-preserving deterministic system (M, Σ_M, μ, T_k) is ergodic (cf. Definition 35).

Case 1: For all i there is a j such that, esmz., $T_k(\alpha_i) = \alpha_j$. Because the discrete measure-preserving deterministic system (M, Σ_M, μ, T_k) is ergodic (equation (5.3)), it follows that there is an $h \in \mathbb{N}$ such that, esmz., $T_{kh}(\alpha_1) = \alpha_1$. But this is in contradiction with the assumption that it is not the case that there exists an $n \in \mathbb{R}^+$ and a $C \in \Sigma_M$, $0 < \mu(C) < 1$, such that, esmz., $T_n(C) = C$.

Case 2: For some i there is a j with, esmz., $T_k(\alpha_i) \subset \alpha_j$ and with $\mu(\alpha_i) < \mu(\alpha_j)$. Because the discrete time deterministic system (M, Σ_M, μ, T_k) is ergodic (equation (5.3)), it holds that there is a $h \in \mathbb{N}$ such that, esmz., $T_{hk}(\alpha_j) \subseteq \alpha_i$. Hence it follows that $\mu(\alpha_j) \leq \mu(\alpha_i)$. But this yields the contradiction $\mu(\alpha_i) < \mu(\alpha_j) \leq \mu(\alpha_i)$.

5.7.3 Proof of Theorem 3

Theorem 3 *Let (M, Σ_M, μ, T) be a discrete measure-preserving deterministic system where Σ_M contains all open balls of the metric space (M, d_M), T is continuous at a point $x \in M$, every open ball around x has positive measure, and there is a set $D \in \Sigma_M$, $\mu(D) > 0$, with $d(T(x), D) = \inf\{d(T(x), m) \mid m \in D\} > 0$. Then there is some $\varepsilon > 0$ for which there is no Bernoulli process to which (M, Σ_M, μ, T) is ε-congruent.*

Proof: For $m \in M$, $E \subseteq M$ and $\varepsilon > 0$ let the ball of radius ε around m be $B(m, \varepsilon) = \{y \in M \mid d(y, m) < \varepsilon\}$ and let $B(E, \varepsilon) = \cup_{m \in E} B(m, \varepsilon)$. Since

$d(T(x), D) > 0$, one can choose $\gamma > 0$ and $\beta > 0$ such that $B(T(x), 2\gamma) \cap B(D, 2\beta) = \emptyset$. Because T is continuous at x, one can choose $\delta > 0$ such that $T(B(x, 4\delta)) \subseteq B(T(x), \gamma)$. Recall that $\mu(B(x, 2\delta)) = \rho_1 > 0$ and that $\mu(D) = \rho_2 > 0$. Let $\varepsilon > 0$ be such that $\varepsilon < \frac{\rho_1\rho_2}{8}$, $\varepsilon < \delta$, $\varepsilon < \beta$ and $\varepsilon < \gamma$. I am going to show that there is no Bernoulli process such that (M, Σ_M, μ, T) is ε-congruent to this Bernoulli process.

Assume that (M, Σ_M, μ, T) is ε-congruent to a Bernoulli process, and let $(\Omega, \Sigma_\Omega, \nu, S, \Phi_0)$ be the deterministic representation of this Bernoulli process. This implies that (M, Σ_M, μ, T) is isomorphic (via $\phi : \hat{M} \to \hat{\Omega}$) to the Bernoulli shift $(\Omega, \Sigma_\Omega, \nu, S)$ and hence that (M, Σ_M, μ, T) is a discrete Bernoulli system. Let $\alpha_{\Phi_0} = \{\alpha^1_{\Phi_0} \dots \alpha^s_{\Phi_0}\}$, $s \in \mathbb{N}$, be the partition of $(\Omega, \Sigma_\Omega, \nu)$ corresponding to the observation function Φ_0 (cf. subsection 5.2.1). Let $\check{M} = M \setminus \hat{M}$ and $\check{\Omega} = \Omega \setminus \hat{\Omega}$. Clearly, $\phi^{-1}(\alpha_{\Phi_0}) = \{\phi^{-1}(\alpha^1_{\Phi_0} \setminus \check{\Omega}) \cup \check{M}, \phi^{-1}(\alpha^2_{\Phi_0} \setminus \check{\Omega}), \dots, \phi^{-1}(\alpha^s_{\Phi_0} \setminus \check{\Omega})\}$ is a partition of (M, Σ_M, μ).

Consider all the sets in $\phi^{-1}(\alpha_{\Phi_0})$ which are assigned values in $B(x, 3\delta)$, i.e., all the sets $a \in \phi^{-1}(\alpha_{\Phi_0})$ with $\Phi_0(\phi(m)) \in B(x, 3\delta)$ for almost all $m \in a$. Denote these sets by $A_1, \dots, A_n$, $n \in \mathbb{N}$, and let $A = \cup_{i=1}^n A_i$. Because (M, Σ_M, μ, T) is ε-congruent to $(\Omega, \Sigma_\Omega, \nu, S, \Phi_0)$, it follows that $\mu(A \setminus B(x, 4\delta)) < \varepsilon$ and $\mu(A \cap B(x, 2\delta)) \geq \rho_1/2$.

Now consider all the sets in $\phi^{-1}(\alpha_{\Phi_0})$ which are assigned values in $B(D, \beta)$, i.e., all the sets $c \in \phi^{-1}(\alpha_{\Phi_0})$ where $\Phi_0(\phi(m)) \in B(D, \beta)$ for almost all $m \in c$. Denote these sets by $C_1, \dots, C_k$, $k \in \mathbb{N}$, and let $C = \cup_{i=1}^k C_i$. Because (M, Σ_M, μ, T) is ε-congruent to $(\Omega, \Sigma_\Omega, \nu, S, \Phi_0)$, I have $\mu(C \cap D) \geq \rho_2/2$ and $\mu(C \cap B(T(x), \gamma)) < \varepsilon$.

Because $(\Omega, \Sigma_\Omega, \nu, S)$ is a Bernoulli shift isomorphic to (M, Σ_M, μ, T), it must hold that $\mu(T(A_i) \cap C_j) = \mu(A_i)\mu(C_j)$ for all i, j, $1 \leq i \leq n$, $1 \leq j \leq k$. Hence also $\mu(T(A) \cap C) = \mu(A)\mu(C)$. But it follows that $\mu(A)\mu(C) \geq \frac{\rho_1\rho_2}{4}$ and that $\mu(T(A) \cap C) < \varepsilon + \varepsilon$, and this yields the contradiction $\frac{\rho_1\rho_2}{4} < 2\varepsilon < \frac{\rho_1\rho_2}{4}$ since it was assumed that $\varepsilon < \frac{\rho_1\rho_2}{8}$.

5.7.4 Proof of Theorem 4

Theorem 4 *Let (M, Σ_M, μ, T) be a discrete measure-preserving deterministic system. Then there is a finite-valued observation function Φ and an $\varepsilon > 0$*

such that no Bernoulli process strongly (Φ, ε)*-simulates* (M, Σ_M, μ, T).

Proof: Assume you observe the deterministic system (M, Σ_M, μ, T) with a surjective finite-valued observation function $\Phi : M \to \{o_1, o_2\}$. Then either for every $\varepsilon > 0$ there is a Bernoulli process which strongly (Φ, ε)-simulates (M, Σ_M, μ, T) or not. In the latter case we are done. In the former case there is a $\Theta(m) = o_1\chi_{\alpha_1}(m) + o_2\chi_{\alpha_2}(m)$, $\{\alpha_1, \alpha_2\}$ a partition of (M, Σ_M, μ), such that $\{X_t = \Theta(T^t);\ t \in \mathbb{Z}\}$ is a Bernoulli process with probabilities $p_1 = \mu(\alpha_1), p_2 = \mu(\alpha_2)$. Now consider the partition $\beta = \{\beta_1, \ldots, \beta_l\} = \alpha \vee T\alpha \vee T^{-1}\alpha$ and an observation function $\Phi(m) = \sum_{i=1}^{l} q_i\chi_{\beta_i}(m)$ where $q_i \neq q_j$ for $i \neq j$, $1 \leq i, j \leq l$. I now show that the stochastic process $\{Z_t = \Phi(T^t);\ t \in \mathbb{Z}\}$ is no Bernoulli process. First note that for all t it holds that

$$P\{X_{t+1} = o_1, X_t = o_1, X_{t-1} = o_1\} = P\{Z_t = q_i\} \text{ for some } q_i,\ 1 \leq i \leq l. \tag{5.4}$$

It follows that

$$\begin{aligned} &P\{Z_t = q_i\} = P\{X_{t+1} = o_1, X_t = o_1, X_{t-1} = o_1\} = p_1^3 < \qquad (5.5)\\ &p_1 = \frac{p_1^4}{p_1^3} = \frac{P\{X_{t+1}=o_1, X_t=o_1, X_{t-1}=o_1, X_{t-2}=o_1\}}{P\{X_t=o_1, X_{t-1}=o_1, X_{t-2}=o_1\}} = P\{Z_t=q_i|Z_{t-1}=q_i\}, \end{aligned}$$

and hence that $\{Z_t; t \in \mathbb{Z}\}$ is no Bernoulli process.

And we cannot change Φ on a set of arbitrary small measure such that the resulting stochastic process is a Bernoulli process. For let $\varepsilon > 0$, and consider an arbitrary surjective measurable function $\Psi : M \to \{q_1, \ldots, q_l\}$ with $\mu(\{m \in M \,|\, \Psi(m) \neq \Phi(m)\}) < \varepsilon$. For the stochastic process $\{Y_t = \Psi(T^t);\ t \in \mathbb{Z}\}$ it holds that

$$P\{Y_t = q_i |\, Y_{t-1} = q_i\} > \frac{p_1^4 - 2\varepsilon}{p_1^3 + 2\varepsilon} \text{ and that } P\{Y_t = q_i\} < p_1^3 + \varepsilon. \tag{5.6}$$

Because $p_1 > p_1^3$, it follows that for sufficiently small $\varepsilon > 0$:

$$\frac{p_1^4 - 2\varepsilon}{p_1^3 + 2\varepsilon} > p_1^3 + \varepsilon. \tag{5.7}$$

Hence I can conclude that $P\{Y_t = q_i\} < P\{Y_t = q_i|Y_{t-1} = q_i\}$ and that $\{Y_t;\ t \in \mathbb{Z}\}$ cannot be a Bernoulli process.

5.7.5 Proof of Proposition 1

Proposition 1 *Let (M, Σ_M, μ, T) be a discrete measure-preserving deterministic system. Then there is a finite-valued observation function Φ and an $\varepsilon > 0$ such that no Bernoulli process weakly (Φ, ε)-simulates (M, Σ_M, μ, T).*

Proof: Assume that $\{Z_t;\, t \in \mathbb{Z}\}$ is a Bernoulli process with outcome space S. Let $\Gamma : S \to \bar{M}$, where $\bar{M} = \{q_1, \ldots, q_N\}$, $N \in \mathbb{N}$, be a surjective observation function. I will now show that $\{Y_t = \Gamma(Z_t);\, t \in \mathbb{Z}\}$ is a Bernoulli process too. Clearly, this result and Theorem 4 immediately imply that for the deterministic system (M, Σ_M, μ, T) there is a finite-valued observation function Φ and an $\varepsilon > 0$ such that no Bernoulli process weakly (Φ, ε)-simulates (M, Σ_M, μ, T).

All I have to show is that $\{Y_t;\, t \in \mathbb{Z}\}$ are probabilistically independent. Label the elements $S = \{s_{1,1}, s_{1,2}, \ldots, s_{1,l_1}, \ldots, s_{N,1}, \ldots, s_{N,l_N}\}$, $l_i \in \mathbb{N}$, $1 \leq i \leq N$, such that

$$\Gamma(s_{1,1})=q_1, \Gamma(s_{1,2})=q_1, \ldots, \Gamma(s_{1,l_1})=q_1, \ldots, \Gamma(s_{N,1})=q_N, \ldots, \Gamma(s_{N,l_N})=q_N. \tag{5.8}$$

Now for all $m \in \mathbb{N}$, all $t_1, \ldots, t_m \in \mathbb{Z}$ and all $q_{j_1}, \ldots, q_{j_m} \in \bar{M}$

$$P\{Y_{t_1} = q_{j_1}, \ldots, Y_{t_m} = q_{j_m}\} = \sum_{\text{all possible } k_1,\ldots,k_m} P\{Z_{t_1} = s_{j_1,k_1}, \ldots, Z_{t_m} = s_{j_m,k_m}\} \tag{5.9}$$

$$= \sum_{\text{all possible } k_1,\ldots,k_m} P\{Z_{t_1} = s_{j_1,k_1}\} \cdots P\{Z_{t_m} = s_{j_m,k_m}\} =$$

$$P\{Y_{t_1} = q_{j_1}\} \sum_{\text{all possible } k_2,\ldots,k_m} P\{Z_{t_2}=s_{j_2,k_2}\} \cdots P\{Z_{t_m}=s_{j_m,k_m}\} = \ldots = P\{Y_{t_1}=q_{j_1}\} \cdots P\{Y_{t_m}=q_{j_m}\},$$

and from this follows that $\{Y_t;\, t \in \mathbb{Z}\}$ are probabilistically independent.

5.7.6 Proof of Theorem 5

Theorem 5 *Let (M, Σ_M, μ, T) be a discrete Bernoulli system where the metric space (M, d_M) is separable and where Σ_M contains all open balls of (M, d_M). Then for any $\varepsilon > 0$ there is an irreducible and aperiodic Markov process such that (M, Σ_M, μ, T) is ε-congruent to this Markov process.*

Proof: I need the following definition.

Definition 42 *A partition α of (M, Σ_M, μ) is* generating *for (M, Σ_M, μ, T) if, and only if, for every $A \in \Sigma_M$ there is an $n \in \mathbb{N}$ and a set C of unions of elements in $\vee_{j=-n}^{n} T^j(\alpha)$ such that $\mu((A \setminus C) \cup (C \setminus A)) < \varepsilon$ (cf. Petersen 1983, p. 244).*

By assumption, the deterministic system (M, Σ_M, μ, T) is isomorphic via a function $\phi : \hat{M} \to \hat{\Omega}$ to the deterministic representation $(\Omega, \Sigma_\Omega, \nu, S, \Phi_0)$ of a Bernoulli process with outcome space $\bar{M}$. Let $\alpha_{\Phi_0} = \{\alpha_{\Phi_0}^1, \ldots, \alpha_{\Phi_0}^k\}$, $k \in \mathbb{N}$, be the partition of $(\Omega, \Sigma_\Omega, \nu)$ corresponding to the observation function Φ_0 (cf. subsection 5.2.1). Let $\check{M} = M \setminus \hat{M}$ and $\check{\Omega} = \Omega \setminus \hat{\Omega}$. $\phi^{-1}(\alpha_{\Phi_0}) = \{\phi^{-1}(\alpha_{\Phi_0}^1 \setminus \check{\Omega}) \cup \check{M}, \phi^{-1}(\alpha_{\Phi_0}^2 \setminus \check{\Omega}), \ldots, \phi^{-1}(\alpha_{\Phi_0}^k \setminus \check{\Omega})\}$ is a partition of (M, Σ_M, μ).

Since (M, d_M) is separable, there exists an $r \in \mathbb{N}$ and $m_i \in M$, $1 \leq i \leq r$, such that $\mu(M \setminus \cup_{i=1}^{r} B(m_i, \frac{\varepsilon}{2})) < \frac{\varepsilon}{2}$. Because for a discrete Bernoulli system $\phi^{-1}(\alpha_{\Phi_0})$ is generating for (M, Σ_M, μ, T) (Petersen 1983, p. 275), for each $B(m_i, \frac{\varepsilon}{2})$ there is an $n_i \in \mathbb{N}$ and a C_i of union of elements in $\vee_{j=-n_i}^{n_i} T^j(\phi^{-1}(\alpha_{\Phi_0}))$ such that $\mu(D_i) < \frac{\varepsilon}{2r}$, where $D_i = (B(m_i, \frac{\varepsilon}{2}) \setminus C_i) \cup (C_i \setminus B(m_i, \frac{\varepsilon}{2}))$. Define $n = \max\{n_i\}$. For $Q = \{q_1, \ldots, q_l\} = \vee_{j=-n}^{n} S^j(\alpha_{\Phi_0})$ let $\Phi_0^Q : \Omega \to M, \Phi_0^Q(\omega) = \sum_{i=1}^{l} o_i \chi_{q_i}(\omega)$, where $o_i \in \phi^{-1}(q_i \setminus \check{\Omega})$. Note that $o_i \neq o_j$ for $i \neq j$, $1 \leq i, j \leq l$. Then

$$d_M(m, \Phi_0^Q(\phi(m))) < \varepsilon \text{ except for a set in } M \text{ of measure } < \varepsilon. \qquad (5.10)$$

$\{\Phi_0^Q(S^t);\, t \in \mathbb{Z}\}$ is a stochastic process from $(\Omega, \Sigma_\Omega, \nu)$ to (M, Σ_M), and let $(X, \Sigma_X, \lambda, R, \Theta_0)$ be its deterministic representation. This process is a Markov process since for any $k \in \mathbb{N}$ and any $A, B_1, \ldots, B_k \in \bar{M}^{2n+1}$,

$$\frac{\nu(\{\omega \in \Omega \,|\, (\omega_{-n} \ldots \omega_n) = A \text{ and } (\omega_{-n+1} \ldots \omega_{n+1}) = B_1\})}{\nu(\{\omega \in \Omega \,|\, (\omega_{-n+1} \ldots \omega_{n+1}) = B_1\})} = \qquad (5.11)$$
$$\frac{\nu(\{\omega \in \Omega | (\omega_{-n} \ldots \omega_n) = A \text{ and} (\omega_{-n+1} \ldots \omega_{n+1}) = B_1, \ldots, (\omega_{-n+k} \ldots \omega_{n+k}) = B_k\})}{\nu(\{\omega \in \Omega | (\omega_{-n+1} \ldots \omega_{n+1}) = B_1, \ldots, (\omega_{-n+k} \ldots \omega_{n+k}) = B_k\})},$$

if $\nu(\{\omega \in \Omega | (\omega_{-n} \ldots \omega_n) = A \text{ and} (\omega_{-n+1} \ldots \omega_{n+1}) = B_1, \ldots, (\omega_{-n+k} \ldots \omega_{n+k}) = B_k\}) > 0$.

Because S is a shift, one sees that for all i, j, $1 \leq i, j \leq l$, there is a $k \geq 1$ such that $P^k(o_i, o_j) > 0$, and hence that the Markov process is irreducible. One also sees that there exists an outcome o_i, $1 \leq i \leq l$, such that $P^1(o_i, o_i) > 0$. Hence $d_{o_i} = 1$; and since all outcomes of an irreducible Markov process have the same periodicity (Cinlar 1975, p. 131), it follows that the Markov process is also aperiodic.

Consider $\psi : \Omega \to X$, $\psi(\omega) = \ldots \Phi_0^Q(S^{-1}(\omega)), \Phi_0^Q(\omega), \Phi_0^Q(S(\omega)) \ldots$, for $\omega \in \Omega$. Clearly, there is a $\hat{X} \subseteq X$ with $\lambda(\hat{X}) = 1$ such that $\psi : \Omega \to \hat{X}$ is bijective and measure-preserving and $R(\psi(\omega)) = \psi(S(\omega))$ for all $\omega \in \Omega$. Hence $(\Omega, \Sigma_\Omega, \nu, S)$ is isomorphic to $(X, \Sigma_X, \lambda, R)$ via ψ, and thus (M, Σ_M, μ, T) is isomorphic to $(X, \Sigma_X, \lambda, R)$ via $\theta = \psi(\phi)$. Now because of (5.10):

$$d_M(m, \Theta_0(\theta(m))) < \varepsilon \text{ except for a set in } M \text{ of measure } < \varepsilon. \qquad (5.12)$$

5.7.7 Proof of Proposition 2

Proposition 2 *Let (M, Σ_M, μ, T) be a discrete Bernoulli system. Then for every finite-valued observation function Φ and every $\varepsilon > 0$ there is an irreducible and aperiodic Markov process which weakly (Φ, ε)-simulates (M, Σ_M, μ, T).*

Proof: Let (M, Σ_M, μ, T) be a discrete Bernoulli system. Let $\Phi : M \to \{q_1, \ldots, q_N\}$, $N \in \mathbb{N}$, be an arbitrary surjective finite-valued observation function and let $\varepsilon > 0$ be arbitrary. Theorem 6 implies that there is an n and a surjective measurable function $\Theta : M \to Q$, $\Theta(m) = \sum_{i=1}^{N} q_i \chi_{\alpha_i}(m)$, for a partition α, such that $\{Z_t = \Theta(T^t);\, t \in \mathbb{Z}\}$ is a Markov process of order n which strongly (Φ, ε)-simulates (M, Σ_M, μ, T). Define $\beta = \{\beta_1 \ldots, \beta_l\} = \alpha \vee T\alpha \vee \ldots \vee T^{n-1}\alpha$, and let $\Psi : M \to \{o_1, \ldots o_l\}$, $\Psi(m) = \sum_{j=1}^{l} o_j \chi_{\beta_j}(m)$ with $o_i \neq o_j$ for $i \neq j$, $1 \leq i, j \leq l$. Let the surjective observation function $\Gamma : \{o_1, \ldots, o_l\} \to Q$ be defined as follows: for any arbitrary r, $1 \leq r \leq N$, any o_i and any o_j, $1 \leq i, j \leq l$, such that $\beta_i \subseteq \alpha_r$ and $\beta_j \subseteq \alpha_r$ are assigned the same value, namely $\Gamma(o_i) = \Gamma(o_j) = q_r$, where q_r is the value Θ takes for all states in α_r. By construction, $Z_t = \Gamma(\Psi(T^t))$ and, since $\{Z_t;\, t \in \mathbb{Z}\}$ strongly (Φ, ε)-simulates (M, Σ_M, μ, T), $\mu(\{m \in M \mid \Gamma(\Psi(m) \neq \Phi(m)\}) < \varepsilon$. Consequently, $\{Y_t = \Psi(T^t);\, t \in \mathbb{Z}\}$ weakly (Φ, ε)-simulates (M, Σ_M, μ, T).

So it remains only to show that $\{Y_t;\, t \in \mathbb{Z}\}$ is an irreducible and aperiodic Markov process. By construction, for all t and all i, $1 \leq i \leq l$, there are $q_{i,0}, \ldots, q_{i,n-1} \in Q$ such that

$$P\{Y_t = o_i\} = P\{Z_t = q_{i,0}, Z_{t+1} = q_{i,1}, \ldots, Z_{t+n-1} = q_{i,n-1}\}. \tag{5.13}$$

Therefore, for all $k \in \mathbb{N}$ and all $i, j_1, \ldots, j_k$, $1 \leq i, j_1, \ldots, j_k \leq l$:

$$P\{Y_{t+1} = o_i \,|\, Y_t = o_{j_1}, \ldots, Y_{t-k+1} = o_{j_k}\} = \tag{5.14}$$

$$P\{Z_{t+1}{=}q_{i,0},\ldots, Z_{t+n}{=}q_{i,n-1}|Z_t{=}q_{j_1,0},\ldots, Z_{t+n-1}{=}q_{i,n-2}, Z_{t-1}{=}q_{j_2,0},\ldots, Z_{t-k+1}{=}q_{j_k,0}\}$$
$$= P\{Z_{t+1} = q_{i,0}, \ldots, Z_{t+n} = q_{i,n-1} \,|\, Z_t = q_{j_1,0}, \ldots, Z_{t+n-1} = q_{i,n-2}\}$$
$$= P\{Y_{t+1} = o_i \,|\, Y_t = o_{j_1}\},$$

if $P\{Y_{t+1} = o_i, Y_t = o_{j_1}, \ldots, Y_{t-k+1} = o_{j_k}\} > 0$. Hence $\{Y_t;\, t \in \mathbb{Z}\}$ is a Markov process. Every discrete Bernoulli system is strongly mixing (cf. Definition 27) (Petersen 1983, p. 58). Consequently, (M, Σ_M, μ, T) is strongly mixing, and this immediately implies that the Markov process $\{Y_t;\, t \in \mathbb{Z}\}$ is irreducible and aperiodic.

5.7.8 Proof of Theorem 8

Theorem 8 *Assume that* (M, Σ_M, μ, T) *is a discrete measure-preserving deterministic system with zero Kolmogorov-Sinai entropy or an ergodic discrete measure-preserving deterministic system with finite Kolmogorov-Sinai entropy which is not a discrete Bernoulli system. Then there is a finite-valued observation function* Φ *and an* $\varepsilon > 0$ *such that no irreducible and aperiodic multi-step Markov process strongly* (Φ, ε)*-simulates* (M, Σ_M, μ, T)*.*

Proof:

Case 1: Assume that (M, Σ_M, μ, T) is a discrete measure-preserving deterministic system with zero Kolmogorov-Sinai entropy. Assume that for some finite-valued observation function $\Psi(m) = \sum_{i=1}^{n} o_i \chi_{\alpha_i}$, where α is a partition, $\{\Psi(T^t);\ t \in \mathbb{Z}\}$ is an irreducible and aperiodic multi-step Markov process. The deterministic representation of this Markov process has Kolmogorov-Sinai entropy $E > 0$ because the deterministic representation of any irreducible and aperiodic multi-step Markov process is a Bernoulli system (cf. Theorem 7). This implies that $H(\alpha, T) \geq E > 0$ (where $H(\alpha, T)$ is the entropy relative to the partition α; see equation (3.6) in subsection 3.4.1). Hence the Kolmogorov-Sinai entropy of (M, Σ_M, μ, T) is positive. But this cannot be the case. Therefore, there can be no finite-valued observation function Ψ such that $\{\Psi(T^t);\ t \in \mathbb{Z}\}$ is an irreducible and aperiodic multi-step Markov process. Consequently, there is a finite-valued observation function Φ and an $\varepsilon > 0$ such that no irreducible and aperiodic multi-step Markov process strongly (Φ, ε)-simulates (M, Σ_M, μ, T).

Case 2: Assume that (M, Σ_M, μ, T) is an ergodic discrete measure-preserving deterministic system with finite Kolmogorov-Sinai entropy which is not a discrete Bernoulli system. I have to show that there is a finite-valued observation function Φ and an $\varepsilon > 0$ such that no irreducible and aperiodic multi-step Markov process strongly (Φ, ε)-simulates (M, Σ_M, μ, T). I can equally show the following claim (C): assume that an ergodic discrete measure-preserving deterministic system with finite Kolmogorov-Sinai entropy is given where for every $\varepsilon > 0$ and every finite-valued observation function Φ there is an n such that an irreducible and aperiodic Markov process of order n strongly (Φ, ε)-simulates (M, Σ_M, μ, T). Then (M, Σ_M, μ, T) is a discrete Bernoulli system.

So assume the assumptions of claim (C). A theorem by Krieger (1970) implies that there is a partition $\beta = \{\beta_1, \ldots, \beta_r\}$, $r \in \mathbb{N}$, of (M, Σ_M, μ) which is generating for (M, Σ_M, μ, T) (cf. Definition 42) (Krieger 1970). I need the following theorem (Ornstein 1973*a*; Petersen 1983, pp. 274–275):

> (+) Let (K, Σ_K, μ_K, R) be a discrete measure-preserving deterministic system, and let $\Pi(k) = \sum_{i=1}^{l} o_i \chi_{\alpha_i}(k)$, $l \in \mathbb{N}$, $o_i \neq o_j$ for $i \neq j$, $1 \leq i, j \leq l$, where the partition $\alpha = \{\alpha_1, \ldots, \alpha_l\}$ is generating for (K, Σ_K, μ_K, R). Assume that for all $\varepsilon > 0$ there is a surjective measurable function $\Theta : K \to \{u_1, \ldots, u_s\}$, $s \geq l$, and a surjective measurable function $\Gamma : \{u_1, \ldots, u_s\} \to \{o_1, \ldots o_l\}$ with $\mu_K(\{k \in K \mid \Pi(k) \neq \Gamma(\Theta(k))\}) < \varepsilon$ such that the deterministic representation of $\{\Theta(R^t);\ t \in \mathbb{Z}\}$ is a discrete Bernoulli system. Then (K, Σ_K, μ_K, R) is a discrete Bernoulli system.

Let $\Phi(m) = \sum_{i=1}^{r} q_r \chi_{\beta_r}(m)$, $q_i \neq q_j$ for $i \neq j$, $1 \leq i, j \leq r$. Then for every $\varepsilon > 0$ there is an n and an irreducible and aperiodic Markov process of order n which strongly (Φ, ε)-simulates (M, Σ_M, μ, T). The deterministic representation of every irreducible and aperiodic multi-step Markov process is a discrete Bernoulli system (Theorem 7). Consequently, Theorem (+) implies that (M, Σ_M, μ, T) is a discrete Bernoulli system.

5.7.9 Proof of Theorem 9

Theorem 9 *Assume that* (M, Σ_M, μ, T) *is a discrete measure-preserving deterministic system with zero Kolmogorov-Sinai entropy or an ergodic discrete measure-preserving deterministic system with finite Kolmogorov-Sinai entropy which is not a discrete Bernoulli system. Then there is a finite-valued observation function* Φ *and an* $\varepsilon > 0$ *such that no irreducible and aperiodic Markov process weakly* (Φ, ε)*-simulates* (M, Σ_M, μ, T).

Proof: The proof is essentially the same as the proof of Theorem 8 (cf. subsection 5.7.8).

Case 1: Assume that (M, Σ_M, μ, T) is a discrete measure-preserving deterministic system with zero Kolmogorov-Sinai entropy. An irreducible and aperiodic Markov process is an irreducible and aperiodic Markov process of order 1. Hence, Case 1 of the proof of Theorem 8 shows that there is a finite-valued observation function Φ and an $\varepsilon > 0$ such that no irreducible and aperiodic Markov process weakly (Φ, ε)-simulates (M, Σ_M, μ, T).

Case 2: Let (M, Σ_M, μ, T) be an ergodic measure-preserving deterministic system with finite Kolmogorov-Sinai entropy which is not a discrete Bernoulli system. I have to show that there is a finite-valued observation function Φ and an $\varepsilon > 0$ such that no irreducible and aperiodic Markov process weakly (Φ, ε)-simulates (M, Σ_M, μ, T). Again it suffices to show the following claim (C): assume that an ergodic discrete measure-preserving deterministic system (M, Σ_M, μ, T) with finite Kolmogorov-Sinai entropy is given where for every $\varepsilon > 0$ and every finite-valued observation function Φ an irreducible and aperiodic Markov process weakly (Φ, ε)-simulates (M, Σ_M, μ, T). Then (M, Σ_M, μ, T) is a discrete Bernoulli system.

So assume that the assumptions of claim (C) are fulfilled. The theorem by Krieger (1970) implies that there is a partition $\beta = \{\beta_1, \ldots, \beta_r\}$, $r \in \mathbb{N}$, which is generating for (M, Σ_M, μ, T). Define $\Phi(m) = \sum_{i=1}^{r} q_r \chi_{\beta_r}(m)$, $q_i \neq q_j$ for $i \neq j$, $1 \leq i, j \leq r$. Then for every $\varepsilon > 0$ there is an irreducible and aperiodic Markov process which weakly (Φ, ε)-simulates (M, Σ_M, μ, T). Therefore, from Theorem (+) (as stated in the proof of Theorem 8) and the fact that the deterministic representation of every irreducible and aperiodic Markov process is discrete Bernoulli system, it follows that (M, Σ_M, μ, T) is a discrete Bernoulli system.

5.7.10 Proof of Theorem 12

Theorem 12 *Let (M, Σ_M, μ, T_t) be a continuous Bernoulli system. Then for every finite-valued observation function Φ and every $\varepsilon > 0$ there is an irrationally related semi-Markov process $\{Z_t;\, t \in \mathbb{R}\}$ which weakly (Φ, ε)-simulates (M, Σ_M, μ, T_t).*

Proof: Let (M, Σ_M, μ, T_t) be a continuous Bernoulli system, let $\Phi : M \rightarrow$

$S, S = \{s_1, \ldots, s_N\}$, $N \in \mathbb{N}$, be an arbitrary surjective finite-valued observation function, and let $\varepsilon > 0$ be arbitrary. Theorem 11 implies that there is an $n \in \mathbb{N}$ and a surjective observation function $\Theta : M \to S$, $\Theta(m) = \sum_{i=1}^{N} s_i \chi_{\alpha_i}(m)$, for a partition α, such that $\{Y_t = \Theta(T_t);\, t \in \mathbb{R}\}$ is an irrationally related semi-Markov process of order n with outcomes s_i and corresponding times $u(s_i)$, $1 \leq i \leq N$, which strongly (Φ, ε)-simulates (M, Σ_M, μ, T_t).

I need the following definition:

Definition 43 *The discrete deterministic system* $(M_2, \Sigma_{M_2}, \mu_2, T_2)$ *is a* factor *of the discrete deterministic system* $(M_1, \Sigma_{M_1}, \mu_1, T_1)$ *(where both systems are assumed to be measure-preserving) if, and only if, there are measurable sets* $\hat{M}_i \subseteq M_i$ *with* $\mu_i(M_i \setminus \hat{M}_i) = 0$ *and* $T_i \hat{M}_i \subseteq \hat{M}_i$ $(i = 1, 2)$ *and there is a function* $\phi : \hat{M}_1 \to \hat{M}_2$ *such that (i)* $\phi^{-1}(B) \in \Sigma_{M_1}$ *for all* $B \in \Sigma_{M_2}, A \subseteq \hat{M}_2$*; (ii)* $\mu_1(\phi^{-1}(B)) = \mu_2(B)$ *for all* $B \in \Sigma_{M_2}$, $B \subseteq \hat{M}_2$*; (iii)* $\phi(T_1(m)) = T_2(\phi(m))$ *for all* $m \in \hat{M}_1$*. For continuous measure-preserving deterministic systems* $(M_1, \Sigma_{M_1}, \mu_1, T_t^1)$ *and* $(M_2, \Sigma_{M_2}, \mu_2, T_t^2)$ *the definition of a* factor *is the same except that condition (iii) is* $\phi(T_t^1(m)) = T_t^2(\phi(m))$ *for all* $m \in \hat{M}_1$ *and all* $t \in \mathbb{R}$ *(cf. Petersen 1983, p. 11).*[22]

Note that the deterministic representation $(X, \Sigma_X, \mu_X, W_t, \Lambda_0)$ of this semi-Markov process of order n is a factor of (M, Σ_M, μ, T_t) (via the function $\phi(m) = r_m$, where r_m is the realisation of m of the stochastic process $\{Y_t;\, t \in \mathbb{R}\}$) (cf. Ornstein & Weiss 1991, p. 18).

Now I construct a continuous measure-preserving deterministic system $(K, \Sigma_K, \mu_K, R_t)$ as follows. Let $(\Omega, \Sigma_\Omega, \mu_\Omega, V, \Xi_0)$, $\Xi_0(\omega) = \sum_{i=1}^{N} s_i \chi_{\beta_i}(\omega)$, where β is a partition, be the deterministic representation of $\{S_k;\, k \in \mathbb{Z}\}$, the irreducible and aperiodic Markov process of order n corresponding to $\{Y_t;\, t \in \mathbb{R}\}$ (see Example 5). Let $f : \Omega \to \{u_1, \ldots, u_N\}$, $f(\omega) = u(\Xi_0(\omega))$. Define K as $\cup_{i=1}^{N} K_i = \cup_{i=1}^{N} (\beta_i \times [0, u(s_i)))$. Let Σ_{K_i}, $1 \leq i \leq N$, be the product σ-algebra $(\Sigma_\Omega \cap \beta_i) \times L([0, u(s_i)))$ where $L([0, u(s_i)))$ is the Lebesgue

[22] Clearly, if measure-preserving deterministic systems are isomorphic (Definition 19), then they are a factor of each other; but if a measure-preserving deterministic system is a factor of another deterministic system, this does not imply that they are isomorphic.

σ-algebra of $[0, u(s_i))$. Let μ_{K_i} be the product measure

$$(\mu_\Omega^{\Sigma_\Omega \cap \beta_i} \times \lambda([0, u(s_i))))/\sum_{j=1}^{N} u(s_j)\mu_\Omega(\beta_j), \tag{5.15}$$

where $\lambda([0, u(s_i)))$ is the Lebesgue measure on $[0, u(s_i))$ and $\mu_\Omega^{\Sigma_\Omega \cap \beta_i}$ is the measure μ_Ω restricted to $\Sigma_\Omega \cap \beta_i$. Now define Σ_K as the completion of the σ-algebra generated by $\cup_{i=1}^{N} \Sigma_{K_i}$. Define a pre-measure $\bar{\mu}_K$ on the semi-algebra

$$H = (\cup_{i=1}^{N}(\Sigma_\Omega \cap \beta_i \times L([0, s_i)))) \cup K, \tag{5.16}$$

by $\bar{\mu}_K(K) = 1$ and $\bar{\mu}_K(A) = \mu_{K_i}(A)$ for $A \in \Sigma_{K_i}$, and let μ_K be the unique extension of this pre-measure to a measure on Σ_K. Finally, R_t is defined as follows: let the state of the deterministic system at time zero be $(k, v) \in K$, $k \in \Omega, v < f(k)$; the state moves vertically with unit velocity, and just before it reaches $(k, f(k))$ it jumps to $(V(k), 0)$ at time $f(k)-v$; then it again moves vertically with unit velocity, and just before it reaches $(V(k), f(V(k))))$ it jumps to $(V^2(k), 0)$ at time $f(V(k)) + f(k) - v$, and so on. $(K, \Sigma_K, \mu_K, R_t)$ is a continuous measure-preserving deterministic system (called a 'flow built under the function f'), and it has been shown that $(X, \Sigma_X, \mu_X, W_t)$ is isomorphic (via a function ψ) to $(K, \Sigma_K, \mu_K, R_t)$ (Ambrose 1941; Park 1982; Rudolph 1976).

Exactly as in the proof of Proposition 2 we see that for $\gamma = \{\gamma_1, \ldots, \gamma_l\} = \beta \vee V\beta \vee \ldots \vee V^{n-1}\beta$ and $\Pi(\omega) = \sum_{j=1}^{l} q_j \chi_{\gamma_j}(\omega)$, $q_j \neq q_i$ for $i \neq j$, $1 \leq i, j \leq l$, the discrete stochastic process $\{B_t = \Pi(V^t(\omega))\}$ is an irreducible and aperiodic Markov process. Now consider $\Delta(k) = \sum_{i=1}^{l} q_i \chi_{\gamma_i \times [0, u(q_i))}(k)$, where $u(q_i)$, $1 \leq i \leq l$, is defined as follows: $u(q_i) = u(s_r)$ where $\gamma_i \subseteq \beta_r$. Then it follows immediately that the stochastic process $\{X_t = \Delta(R_t);\, t \in \mathbb{R}\}$ is an irrationally related semi-Markov process.

Let the surjective measurable function $\Psi : M \rightarrow \{q_1, \ldots, q_l\}$ be defined as follows: $\Psi(m) = \Delta(\psi(\phi(m)))$ for $m \in \hat{M}$ and q_1 otherwise. Recall that $(X, \Sigma_X, \mu_X, W_t)$ is a factor (via ϕ) of (M, Σ_M, μ, T_t) and that $(X, \Sigma_X, \mu_X, W_t)$ is isomorphic (via ψ) to $(K, \Sigma_K, \mu_K, R_t)$. Therefore, it follows that $\{Z_t = \Psi(T_t);\, t \in \mathbb{R}\}$ is an irrationally related semi-Markov process with outcomes q_i and corresponding times $u(q_i)$, $1 \leq i \leq l$. Consider the surjective finite-valued observation function $\Gamma : \{q_1, \ldots, q_l\} \rightarrow S$, where $\Gamma(q_i)$, $1 \leq i \leq l$, is defined as follows: $\Gamma(q_i) = s_r$ where $\gamma_i \subseteq \beta_r$. By construction, we obtain that, esmz., $\Gamma(\Psi(T_t(m))) = Y_t(m)$ for all $t \in \mathbb{R}$. Hence, because $\{Y_t;\, t \in \mathbb{Z}\}$ strongly (Φ, ε)-simulates (M, Σ_M, μ, T_t), $\mu(\{m \in M \,|\, \Gamma(\Psi(m)) \neq \Phi(m)\}) < \varepsilon$.

5.7.11 Proof of Theorem 14

Theorem 14 *Let (M, Σ_M, μ, T_t) be a continuous measure-preserving deterministic system with zero Kolmogorov-Sinai entropy or a continuous measure-preserving deterministic system which is not a continuous Bernoulli system and where for some $t_0 \in \mathbb{R}\backslash\{0\}$ the discrete measure-preserving deterministic system $(M, \Sigma_M, \mu, T_{t_0})$ is ergodic. Then there is a finite-valued observation function Φ and an $\varepsilon > 0$ such that no irrationally related multi-step semi-Markov process strongly (Φ, ε)-simulates (M, Σ_M, μ, T_t).*

Proof: The proof paralells the proof of the analogous discrete-time result (Theorem 8).

Case 1: Let (M, Σ_M, μ, T_t) be a continuous measure-preserving deterministic system with zero Kolmogorov-Sinai entropy. Assume that there is a finite-valued observation function $\Psi(m) = \sum_{i=1}^{n} o_i \chi_{\alpha_i}$, where α is a partition, such that $\{\Psi(T_t);\ t \in \mathbb{R}\}$ is an irrationally related multi-step semi-Markov process. The deterministic representation of this multi-step semi-Markov process has Kolmogorov-Sinai entropy $E > 0$ because the deterministic representation of any irrationally related multi-step semi-Markov process is a continuous Bernoulli system (cf. Theorem 13). Hence $H(\alpha, T_1) \geq E > 0$ (where $H(\alpha, T_1)$ is the entropy relative to the partition α; see equation (3.6) in subsection 3.4.1). But this means that the Kolmogorov-Sinai entropy of (M, Σ_M, μ, T_t) is positive, which contradicts the assumption. Therefore, there can be no finite-valued observation function Ψ such that $\{\Psi(T_t);\ t \in \mathbb{R}\}$ is an irrationally related multi-step semi-Markov process. Consequently, there is a finite-valued observation function Φ and an $\varepsilon > 0$ such that no irrationally related multi-step semi-Markov process strongly (Φ, ε)-simulates (M, Σ_M, μ, T_t).

Case 2: Assume that the continuous deterministic system (M, Σ_M, μ, T_t)

has finite Kolmogorov-Sinai entropy, is not a continuous Bernoulli system, and that for some $t_0 \in \mathbb{R} \setminus \{0\}$ the discrete system $(M, \Sigma_M, \mu, T_{t_0})$ is ergodic. I have to show that there is a finite-valued observation function Φ and an $\varepsilon > 0$ such that no irrationally related multi-step semi-Markov process strongly (Φ, ε)-simulates (M, Σ_M, μ, T_t). For this I can equally show the following claim (C): assume that a continuous measure-preserving deterministic system (M, Σ_M, μ, T_t) is given which has finite Kolmogorov-Sinai entropy and where for some $t_0 \in \mathbb{R} \setminus \{0\}$ the discrete deterministic system $(M, \Sigma_M, \mu, T_{t_0})$ is ergodic. Further assume that for every $\varepsilon > 0$ and every finite-valued observation function Φ there is an $n \in \mathbb{N}$ such that an irrationally related semi-Markov process of order n strongly (Φ, ε)-simulates (M, Σ_M, μ, T_t). Then (M, Σ_M, μ, T_t) is a continuous Bernoulli system.

So assume that the assumptions of claim (C) are satisfied. I need the following definition:

Definition 44 *A partition $\alpha = \{\alpha_1, \ldots, \alpha_n\}$ of (M, Σ_M, μ) is* generating for (M, Σ_M, μ, T_t) *if, and only if, for every $A \in \Sigma_M$ there is a $\tau \in \mathbb{R}^+$ and a set C of unions of elements in $\bigcup_{\text{all } m} \bigcap_{t=-\tau}^{\tau} (T^{-t}(\alpha(T^t(m))))$, where $\alpha(m)$ is defined as the set $\alpha_j \in \alpha$ with $m \in \alpha_j$, such that $\mu((A \setminus C) \cup (C \setminus A)) < \varepsilon$.*

Because the discrete deterministic system $(M, \Sigma_M, \mu, T_{t_0})$ is ergodic, the theorem by Krieger (1970) implies that there is a partition $\beta = \{\beta_1, \ldots, \beta_r\}$, $r \in \mathbb{N}$, which is generating for $(M, \Sigma_M, \mu, T_{t_0})$ and thus also generating for the continuous deterministic system (M, Σ_M, μ, T_t). I need the following theorem (Ornstein & Weiss 1991, p. 66; Petersen 1983, pp. 274–275):

> (++) Let $(K, \Sigma_K, \mu_K, R_t)$ be a continuous measure-preserving deterministic system, and let $\Pi(k) = \sum_{i=1}^{l} o_i \chi_{\alpha_i}(k)$, $l \in \mathbb{N}$, $o_i \neq o_j$ for $i \neq j$, $1 \leq i, j \leq l$, where the partition $\alpha = \{\alpha_1, \ldots, \alpha_l\}$ is generating. Assume that for all $\varepsilon > 0$ there is a surjective measurable function $\Theta : K \to \{u_1, \ldots, u_s\}$, $s \geq l$, and a surjective measurable function $\Gamma : \{u_1, \ldots, u_s\} \to \{o_1, \ldots o_l\}$ with $\mu_K(\{k \in K \,|\, \Pi(k) \neq \Gamma(\Theta(k))\}) < \varepsilon$ such that the deterministic representation of $\{\Theta(R_t);\ t \in \mathbb{R}\}$ is a continuous Bernoulli system. Then $(K, \Sigma_K, \mu_K, R_t)$ is a continuous Bernoulli system.

Let $\Phi(m) = \sum_{i=1}^{r} q_r \chi_{\beta_r}(m)$, $q_i \neq q_j$ for $i \neq j$, $1 \leq i, j \leq r$. It follows that for every $\varepsilon > 0$ there is an $n \in \mathbb{N}$ and an irrationally related semi-Markov process of order n which strongly (Φ, ε)-simulates (M, Σ_M, μ, T_t). The deterministic representation of every irrationally related multi-step semi-Markov process is a continuous Bernoulli system (Theorem 13). Consequently, Theorem (++) implies that (M, Σ_M, μ, T_t) is a continuous Bernoulli system.

5.7.12 Proof of Theorem 15

Theorem 15 *Let (M, Σ_M, μ, T_t) be a continuous measure-preserving deterministic system with zero Kolmogorov-Sinai entropy or a continuous measure-preserving deterministic system which is not a continuous Bernoulli system and where for some $t_0 \in \mathbb{R} \setminus \{0\}$ the discrete measure-preserving deterministic system $(M, \Sigma_M, \mu, T_{t_0})$ is ergodic. Then there is a finite-valued observation function Φ and an $\varepsilon > 0$ such that no irrationally related semi-Markov process weakly (Φ, ε)-simulates (M, Σ_M, μ, T_t).*

Proof: The proof is essentially the same as the proof of Theorem 14 (cf. subsection 5.7.11).

Case 1: Because an irrationally related semi-Markov process is an irrationally related semi-Markov process of order 1, Case 1 of the proof of Theorem 14 shows that there is a finite-valued observation function Φ and an $\varepsilon > 0$ such that no irrationally related semi-Markov process strongly (Φ, ε)-simulates (M, Σ_M, μ, T_t).

Case 2: Let (M, Σ_M, μ, T_t) be a continuous measure-preserving deterministic system with finite Kolmogorov-Sinai entropy which is not a continuous Bernoulli system. Assume that for some $t_0 \in \mathbb{R} \setminus \{0\}$ the discrete deterministic system $(M, \Sigma_M, \mu, T_{t_0})$ is ergodic. It needs to be shown that there is a finite-valued observation function Φ and an $\varepsilon > 0$ such that no irrationally related semi-Markov process strongly (Φ, ε)-simulates (M, Σ_M, μ, T_t). For this I only have to show the following claim (C): assume that a continuous measure-preserving deterministic system (M, Σ_M, μ, T_t) is given which has finite Kolmogorov-Sinai entropy and where for some $t_0 \in \mathbb{R} \setminus \{0\}$ the discrete system $(M, \Sigma_M, \mu, T_{t_0})$ is ergodic. Further assume that for every $\varepsilon > 0$

and every finite-valued observation function Φ an irrationally related semi-Markov process weakly (Φ, ε)-simulates (M, Σ_M, μ, T_t). Then (M, Σ_M, μ, T_t) is a continuous Bernoulli system.

So assume that the assumptions of claim (C) are satisfied. According to the theorem by Krieger (1970), there is a partition $\beta = \{\beta_1, \ldots, \beta_r\}$, $r \in \mathbb{N}$, which is generating for $(M, \Sigma_M, \mu, T_{t_0})$ and thus also generating for (M, Σ_M, μ, T_t). Let $\Phi(m) = \sum_{i=1}^{r} q_r \chi_{\beta_r}(m)$, $q_i \neq q_j$ for $i \neq j$, $1 \leq i, j \leq r$. It follows that for every $\varepsilon > 0$ there is an irrationally related semi-Markov process which weakly (Φ, ε)-simulates (M, Σ_M, μ, T_t). The deterministic representation of every irrationally related semi-Markov process is a continuous Bernoulli system. Consequently, Theorem (++) (as stated in the proof of Theorem 14) implies that (M, Σ_M, μ, T_t) is a continuous Bernoulli system.

5.7.13 Proof of Proposition 3

Proposition 3 *Let (M, Σ_M, μ, T) be a discrete measure-preserving deterministic system where (M, d_M) is separable and where Σ_M contains all open balls of (M, d_M). Assume that (M, Σ_M, μ, T) satisfies the assumption of Theorem 1 and has finite Kolmogorov-Sinai entropy. Then for every $\varepsilon > 0$ there is a stochastic process $\{Z_t;\, t \in \mathbb{Z}\}$ with outcome space $\bar{M} = \cup_{l=1}^{h} o_l$, $h \in \mathbb{N}$, such that $\{Z_t;\, t \in \mathbb{Z}\}$ is ε-congruent to (M, Σ_M, μ, T), and for all $k \in \mathbb{N}$ there is an outcome $o_i \in \bar{M}$ such that for all $o_j \in \bar{M}$, $1 \leq j \leq h$, $P\{Z_{t+k} = o_j \mid Z_t = o_i\} < 1$.*

Proof: Recall that if (M, Σ_M, μ, T) satisfies the assumptions of Theorem 1, then (M, Σ_M, μ, T) is ergodic (cf. subsection 5.7.1). Hence the theorem by Krieger (1970) implies that there is a partition α which is generating for (M, Σ_M, μ, T) (cf. Definition 42). Let $\varepsilon > 0$. Since (M, d_M) is separable, there exists a $r \in \mathbb{N}$ and $m_i \in M$, $1 \leq i \leq r$, such that $\mu(M \setminus \cup_{i=1}^{r} B(m_i, \frac{\varepsilon}{2})) < \frac{\varepsilon}{2}$. Because α is generating, for each $B(m_i, \frac{\varepsilon}{2})$ there is an $n_i \in \mathbb{N}$ and a C_i of union of elements in $\vee_{j=-n_i}^{n_i} T^j(\alpha)$ such that $\mu((B(m_i, \frac{\varepsilon}{2}) \setminus C_i) \cup (C_i \setminus B(m_i, \frac{\varepsilon}{2})) < \frac{\varepsilon}{2r}$. Define $n = \max\{n_i\}$, $\beta = \{\beta_1, \ldots, \beta_l\} = \vee_{j=-n}^{n} T^j(\alpha)$ and $\Psi(m) = \sum_{i=1}^{l} o_i \chi_{\beta_i}(m)$ with $o_i \in \beta_i$. Ψ is finite-valued, and Theorem 1 implies that for the process $Z_t = \{\Psi(T^t);\, t \in \mathbb{Z}\}$, for all $k \in \mathbb{N}$ there is an outcome

o_i such that for all o_j, $1 \leq j \leq l$, $P\{Z_{t+k} = o_j \,|\, Z_t = o_i\} < 1$. Furthermore, because α is generating, β is generating. Therefore, (M, Σ_M, μ, T) is isomorphic (via a function ϕ) to the deterministic representation $(M_2, \Sigma_{M_2}, \mu_2, T_2, \Phi_0)$ of $\{Z_t; t \in \mathbb{Z}\}$ (Petersen 1983, p. 274). By construction, $d_M(m, \Phi_0(\phi(m))) < \varepsilon$ except for a set in M of measure smaller than ε.

5.7.14 Proof of Proposition 4

Proposition 4 *Let (M, Σ_M, μ, T_t) be a continuous measure-preserving deterministic system where (M, d_M) is separable and where Σ_M contains all open balls of (M, d_M). Assume that (M, Σ_M, μ, T_t) satisfies the assumption of Theorem 2 and has finite Kolmogorov-Sinai entropy. Then for every $\varepsilon > 0$ there is a stochastic process $\{Z_t;\ t \in \mathbb{R}\}$ with outcome space $M_O = \cup_{l=1}^{h} o_l$, $h \in \mathbb{N}$, such that $\{Z_t;\ t \in \mathbb{R}\}$ is ε-congruent to (M, Σ_M, μ, T_t), and for all $k \in \mathbb{R}^+$ there is an outcome $o_i \in M_O$ such that for all $o_j \in M_O$, $1 \leq j \leq h$, $P\{Z_{t+k} = o_j \,|\, Z_t = o_i\} < 1$.*

Proof: The proof uses the same ideas as the proof for the analogous discrete-time result. By assumption, there is a $t_0 \in \mathbb{R} \setminus \{0\}$ such that the discrete deterministic system $(M, \Sigma_M, \mu, T_{t_0})$ is ergodic. Then the theorem by Krieger (1970) implies that there is a partition α which is generating for $(M, \Sigma_M, \mu, T_{t_0})$ and thus also generating for (M, Σ_M, μ, T_t) (cf. Definition 44). Since (M, d_M) is separable, for every $\varepsilon > 0$ there is a $r \in \mathbb{N}$ and $m_i \in M$, $1 \leq i \leq r$, such that $\mu(M \setminus \cup_{i=1}^{r} B(m_i, \frac{\varepsilon}{2})) < \frac{\varepsilon}{2}$. Because α is generating for $(M, \Sigma_M, \mu, T_{t_0})$, for each $B(m_i, \frac{\varepsilon}{2})$ there is an $n_i \in \mathbb{N}$ and a C_i of union of elements in $\vee_{j=-n_i}^{n_i} T_{jt_0}(\alpha)$ such that $\mu((B(m_i, \frac{\varepsilon}{2}) \setminus C_i) \cup (C_i \setminus B(m_i, \frac{\varepsilon}{2}))) < \frac{\varepsilon}{2r}$. Let $n = \max\{n_i\}$, $\beta = \{\beta_1, \ldots, \beta_l\} = \vee_{j=-n}^{n} T_{jt_0}(\alpha)$ and $\Psi(m) = \sum_{i=1}^{l} o_i \chi_{\beta_i}(m)$ with $o_i \in \beta_i$. Since Ψ is a finite-valued observation function, Theorem 2 implies that for the stochastic process $Z_t = \{\Psi(T_t); t \in \mathbb{R}\}$, for all $k \in \mathbb{R}^+$ there is an outcome o_i, $1 \leq i \leq l$, such that for all o_j, $1 \leq j \leq l$, $P\{Z_{t+k} = o_j \,|\, Z_t = o_i\} < 1$. Because β is generating for (M, Σ_M, μ, T_t), (M, Σ_M, μ, T_t) is isomorphic (via a function ϕ) to the deterministic representation $(M_2, \Sigma_{M_2}, \mu_2, T_t^2, \Phi_0)$ of $\{Z_t; t \in \mathbb{R}\}$ (Petersen 1983, p. 274). And, by construction, $d_M(m, \Phi_0(\phi(m))) < \varepsilon$ except for a set in M smaller than ε.

Chapter 6

Concluding remarks

This dissertation has been about some of the most important philosophical aspects of chaos research, a famous recent area of research about deterministic yet unpredictable and irregular, or even random behaviour. I have treated chaos from a measure-theoretic point of view because only this viewpoint provides a connection to probability theory and to the theory of stochastic processes, contributing to many topics of philosophical relevance. Let me briefly summarise this dissertation.

I started by examining mathematical notions of unpredictability in ergodic theory. On this basis, I drew conclusions about the actual practice of how mathematical definitions are justified. More specifically, I introduced the main account of this issue, namely Lakatos's (1976, 1978) proof-generated definitions. After that I presented two previously unidentified but common ways of justifying definitions which play an important role for notions of unpredictability in ergodic theory, namely condition-justification and redundancy-justification. I argued that these two kinds of justification are among the most important ones in mathematics. Also, I analysed the interrelationships between the different kinds of justification. Then I criticised Lakatos's theory. I argued that it does not acknowledge the interrelationships between the different kinds of justification, and that it ignores the fact that various kinds of justification—not only proof-generation—are important.

With this background on notions of unpredictability, we were ready to tackle the question of what is the unpredictability specific to chaos. There is a widespread belief that chaotic systems are unpredictable in a way that other deterministic systems are not. Hence one might expect that this question has already been answered in a satisfactory way. However, I argued that this is not so: the answers in the literature are defective. This prompted the search for a better answer. An event is called 'probabilistically irrelevant' for predicting another event if knowledge of the latter event neither heightens nor lowers the probability of the former event. Based on defining chaos via strongly mixing, I proposed a novel answer: the unpredictability specific to chaotic systems is that for predicting any event at any level of precision, all sufficiently past events are approximately probabilistically irrelevant.

Finally, the fact that some deterministic systems are unpredictable and random raised the question of whether deterministic systems and stochastic processes can be observationally equivalent. I showed that for many measure-theoretic deterministic systems there is a stochastic process which is observationally equivalent to the deterministic system; and conversely, that for all stochastic processes there is a measure-theoretic deterministic system which is observationally equivalent to the stochastic process. Still, one might guess that the deterministic systems which are observationally equivalent to stochastic processes used in science do not include any deterministic systems used in science. I argued that this is not so because deterministic systems used in science give rise to Bernoulli processes and to semi-Markov processes. Despite this, one might guess that deterministic systems used in science cannot give the same predictions at every observation level as stochastic processes used in science. By proving new results in ergodic theory, I showed that also this guess is misguided: there are deterministic systems used in science which give the same predictions at every observation level as Markov processes or n-step Markov processes (for discrete time) and semi-Markov processes or n-step semi-Markov processes (for continuous time). Therefore, even kinds of stochastic processes and kinds of deterministic systems which intuitively seem to give very different predictions are observationally equivalent. Furthermore, I criticised the previous philosophical literature on observational equivalence, namely Hoefer (2008), Suppes (1993), Suppes & de Barros (1996), Suppes (1999) and Winnie (1998). These authors fail to see the philosophical significance of the results on observational equivalence, and they do not seem to be aware that also non-chaotic deterministic systems can be simulated at every observation level by stochastic processes. Furthermore, the viewpoints of these authors on the question of whether the deterministic or the stochastic description is preferable are untenable, and I have argued that this question needs more careful consideration.

This summary illustrates that this dissertation makes a contribution to the literature at two levels. First, the mathematical theorems and the discussion about how to define chaos contributes to the general mathematical field of dynamical systems theory, and hence is also of relevance to the special sciences where dynamical systems theory is applied, from physics and biology to the social sciences. But, of course, the contribution of this dissertation are not only of mathematical nature. Primarily, this dissertation with its conceptual reflection about the mathematical results advances our knowledge of important philosophical themes such as the justification of definitions, unpredictability, and the question of whether phenomena are deterministic or indeterministic.

To conclude this dissertation, let me give an outlook of important open questions related to my dissertation. Let me first point out four issues which are directly related to the topics I have treated. First, there has traditionally been little philosophical reflection on the actual practice of mathematics, and in particular about the mathematical practice of justifying definitions (for some recent notable work on the actual practice of mathematics, see, for instance, Corfield 2003, Larvor 2001, Leng 2002, Mancosu 2008). So I think that there is much more of philosophical interest that could be said about the justification of definitions, and more generally about mathematical practice, such as what makes theorems deep as opposed to shallow.

Second, philosophers distinguish between process randomness, i.e., randomness of the dynamics of a system, and product randomness, i.e., randomness of its output (Earman 1986, p. 145). Ergodic theorists agree that chaotic processes (and not just outputs) can be random. For instance, the ergodic hierarchy, a series of mathematical definitions, is often claimed to provide a hierarchy of increasing levels of deterministic process randomness (for more on some of the notions of the ergodic hierarchy, see section 3.3 and

4.3). Yet there is hardly any philosophical literature on deterministic process randomness. There is the question of what the account of randomness is endorsed in ergodic theory, and how this account adds to our philosophical understanding. To my knowledge, this question has not been treated apart from Berkovitz et al.'s (2006) analysis of the ergodic hierarchy. Yet two of the levels of the ergodic hierarchy do not correspond to the mathematical characterisation of randomness they propose. Therefore, I have doubts that their characterisation of the randomness involved in the ergodic hierarchy succeeds. The underlying thought in ergodic theory seems to be that there are certain properties which make stochastic processes random, and that chaotic deterministic systems can share these properties and hence can be random. But the details are unclear and worthy of exploration.

Third, I think that there is scope for proving further philosophically relevant mathematical results on the observational equivalence of deterministic and indeterministic descriptions. For instance, one might prove further results about limitations on observational equivalence, similar to my theorems saying that discrete deterministic systems used in science cannot be simulated at every observation level by Bernoulli processes. Furthermore, if there is a choice between a deterministic and an indeterministic description, the question arises which description is preferable. As already highlighted in subsection 5.5, this question deserves a more careful treatment.

Fourth, as explained in some detail in section 2.1, invariant measures are often interpreted as probability densities. There are still many open questions about this issue. For instance, there are interpretations of measures as probability densities which, to the best of my knowledge, have not been philosophically assessed, such as the so-called Kolmogorov measures. These measures are defined as follows: add to a given deterministic system a small random noise ε. The resulting stochastic process usually has just one stationary measure μ_{ε}. The invariant measure $\mu = \lim_{\varepsilon\to 0}\mu_{\varepsilon}$ often exists and is interpreted as probability density since it derives from stochastic processes (Eckmann & Ruelle 1985, p. 626), but it is still unclear whether these measures justify the appellation 'probability'. Also, there has been no philosophical work on the interesting question of which measure one should choose if two methods of identifying invariant measures suggest different measures.

Furthermore, there has been relatively little philosophical discussion even about the most popular interpretations of invariant measures as probability densities, such as the time-average interpretation. Hence also here there is a need for further research, such as on the topic of how the time-average interpretation is best understood for nonergodic systems (cf. Lavis 2010). This gap is all the more important as all the extant philosophical literature on this issue is about classical statistical mechanics, which lacks the more exotic measures of dynamical systems theory, such as physical measures on strange attractors.

Finally, let me point to three open questions more generally about dynamical systems theory and chaos. First, our understanding of how chaotic behaviour arises from the quantum world is still incomplete (there is, of course, a vast literature; for two recent articles see Belot & Earman 1997 and Landsman 2007, section 5-7). Thus it would be desirable to see more foundational work on this issue.

Second, there are many open questions about the significance of chaotic behaviour in statistical mechanics. Generally, it is still debated what exactly statistical mechanics accomplishes, and it is poorly understood why the various schemes of statistical mechanics such as Gibbs' phase space averaging work (Uffink 2007). In particular, there are many open questions about the role of chaotic behaviour in explaining the second law of thermodynamics or in explaining why in Gibbsian mechanics one can take phase averages of observables. For instance, recent accounts of typicality purport to derive an analogue of the second law of thermodynamics by appealing to chaotic behaviour and ergodicity (Goldstein 2001, Lebowitz 1993): yet it remains unclear whether this derivation indeed goes through (Frigg 2009*a*, Frigg 2009*b*).

Third, chaos research, and more generally dynamical systems theory, is applied in disciplines such as as meteorology and the climate sciences. Policy recommendations and also policies are sometimes based on predictions which were derived from models of dynamical systems theory. Yet this only makes sense if these models are not prone to model error, that is, if approximately the same results are obtained when the model is changed slightly. If model error prevails, then we need to be cautious with conclusions based on these models. Leading climate researchers are aware of this issue (e.g., Smith 2007) and would like to see philosophical as well as mathematical research on the role of model error.

Most of these open questions are also of theoretical importance for the specific sciences, and some of them are relevant to policy. Yet because these questions are conceptual or foundational, scientists tend not to reflect on them carefully. Philosophical research, in particular research in the philosophy of science including the philosophy of the special sciences, can and should fill these gaps. To conclude, there is still much interesting work to be done about the philosophical aspects of chaos and the topics of this dissertation. Exciting work for the future!

List of Figures

Bibliography

Ambrose, W. (1941), ‘Representation of ergodic flows’, *Annals of Mathematics, 2nd Series* **42**, 723–739.

Arnold, V. & Avez, A. (1968), *Ergodic Problems of Classical Mechanics*, W. A. Benjamin, New York.

Ash, R. (1972), *Measure, Integration and Functional Analysis*, Academic Press, New York and London.

Aubin, D. & Dahan-Dalmedico, A. (2002), ‘Writing the history of dynamical systems and chaos: Longue durée and revolution, disciplines and cultures’, *Historia Mathematica* **29**, 273–339.

Batterman, R. (1991), ‘Randomness and probability in dynamical theories: on the proposals of the Prigogine school’, *Philosophy of Science* **58**, 241–263.

Batterman, R. & White, H. (1996), ‘Chaos and algorithmic complexity’, *Foundations of Physics* **26**, 307–336.

Belot, G. & Earman, J. (1997), ‘Chaos out of order: quantum mechanics, the correspondence principle and chaos’, *Studies in History and Philosophy of Modern Physics* **28**, 147–182.

Benedicks, M. & Young, L.-S. (1993), ‘Sinai-Bowen-Ruelle-measures for certain Hénon maps’, *Inventiones Mathematicae* **112**, 541–567.

Berger, A. (2001), *Chaos and Chance, an Introduction to Stochastic Aspects of Dynamics*, De Gruyter, New York.

Berkovitz, J., Frigg, R. & Kronz, F. (2006), 'The ergodic hierarchy, randomness and Hamiltonian chaos', *Studies in History and Philosophy of Modern Physics* **37**, 661–691.

Billingsley, P. (1965), *Ergodic Theory and Information*, John Wiley and Sons, New York.

Birkhoff, G. (1931), 'Proof of the ergodic theorem', *Proceedings of the National Academy of Sciences of the United States of America* **17**, 656–660.

Bishop, R. (2003), 'On separating predictability and determinism', *Erkenntnis* **58**, 169–188.

Bishop, R. (2008), 'What could be worse than the butterfly effect?', *Canadian Journal of Philosophy* **38**, 519–547.

Bowen, R. (1977), 'Bernoulli maps of the interval', *Israel Journal of Mathematics* **28**, 161–168.

Brin, M. & Stuck, G. (2002), *Introduction to Dynamical Systems*, Cambridge, Cambridge University Press.

Brown, J. (1999), *Philosophy of Mathematics: an Introduction to the World of Proofs and Pictures*, Routledge, London.

Butterfield, J. (2005), 'Determinism and indeterminism'. Routledge Encyclopaedia of Philosophy Online.

Carathéodory, C. (1914), 'Über das lineare Maß von Punktmengen—eine Verallgemeinerung des Längenbegriffs', *Nachrichten der königlichen Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse* pp. 404–426.

Chernov, N. & Markarian, R. (2006), *Chaotic Billiards*, American Mathematical Society, Providence.

Chirikov, B. (1979), 'A universal instability of many-dimensional oscillator systems', *Physics Reports* **52**, 264–379.

Cinlar, E. (1975), *Introduction to Stochastic Processes*, Prentice Hall, Englewood Cliffs, New Jersey.

Cohn, D. (1980), *Measure Theory*, Birkhäuser, Boston.

Corfield, D. (1997), 'Assaying Lakatos's philosophy of mathematics', *Studies in History and Philosophy of Science* **28**, 99–121.

Corfield, D. (2003), *Towards a Philosophy of Real Mathematics*, Cambridge University Press, Cambridge.

Cornfeld, I., Fomin, S. & Sinai, Y. (1982), *Ergodic Theory*, Springer, Berlin.

Cover, T. & Thomas, J. (2006), *Elements of Information Theory*, second edn, Wiley, New York.

Dahan-Dalmedico, A. (2004), Chaos, disorder, and mixing: a new Fin-de-siécle image of science?, *in* M. Wise, ed., 'Growing Explanations: Historical Perspective on the Sciences of Complexity', Duke University Press, Durham, pp. 67–94.

Devaney, R. (1986), *An Introduction to Chaotic Dynamical Systems*, Addison-Wesley, New York et al.

Doob, J. L. (1953), *Stochastic Processes*, John Wiley & Sons, New York.

Eagle, A. (2005), 'Randomness is unpredictability', *The British Journal for the Philosophy of Science* **56**, 749–790.

Earman, J. (1971), 'Laplacian determinism, or is this any way to run a universe?', *Journal of Philosophy* **68**, 729–744.

Earman, J. (1986), *A Primer on Determinism*, D. Reidel, Dotrecht.

Eckmann, J.-P. & Ruelle, D. (1985), 'Ergodic theory of chaos and strange attractors', *Reviews of Modern Physics* **57**, 617–654.

Falconer, K. (1990), *Fractal Geometry: Mathematical Foundations and Applications*, John Wiley & Sons, New York.

Ford, J. (1989), What is chaos that we should be mindful of it, *in* P. Davies, ed., 'The New Physics', Cambridge, Cambridge University Press, pp. 348–371.

Frigg, R. (2004), 'In what sense is the Kolmogorov-Sinai entropy a measure for chaotic behaviour?—bridging the gap between dynamical systems theory and communication theory', *The British Journal for the Philosophy of Science* **55**, 411–434.

Frigg, R. (2006), 'Chaos and randomness: An equivalence proof of a generalised version of the Shannon entropy and the Kolmogorov-Sinai entropy for Hamiltonian dynamical systems', *Chaos, Solitons and Fractals* **28**, 26–31.

Frigg, R. (2009*a*), 'Typicality and the approach to equilibrium in Boltzmannian statistical mechanics', *Philosophy of Science (Supplement)* **forthcoming**.

Frigg, R. (2009*b*), Why typicality does not explain the approach to equilibrium, *in* M. Suárez, ed., 'Probabilities, Causes and Propensitites in Physics', forthcoming, Springer, Berlin.

Frigg, R. & Hartmann, S. (2006), Models in science, *in* E. Zalta, ed., 'The Stanford Encyclopedia of Philosophy (Spring 2006 Edition)', http://plato.stanford.edu/archives/spr2006/entries/models-science/, Stanford.

Frigg, R. & Werndl, C. (2010), Entropy – a guide for the perplexed, *in* C. Beisbart & S. Hartmann, eds, 'Probabilities in Physics', forthcoming, Oxford University Press, Oxford.

Goldstein, S. (2001), Boltzmann's approach to statistical mechanics, *in* J. Bricmont, D. Dürr, M. Galavotti, G. Ghirardi, F. Pettrucione & N. Zanghi, eds, 'Chance in Physics: Foundations and Perspectives', Springer, Berlin and New York, pp. 39–54.

Halmos, P. (1944), 'In general a measure-preserving transformation is mixing', *The Annals of Mathematics* **45**, 786–792.

Halmos, P. (1949), 'Measurable transformations', *Bulletin of the American Mathematical Society* **55**, 1015–1043.

Halmos, P. (1950), *Measure Theory*, Van Nostrand, New York and London.

Halmos, P. (1956), *Lectures on Ergodic Theory*, Chelsea Publishing Company, New York.

Halmos, P. (1961), 'Recent progress in ergodic theory', *Bulletin of the American Mathematical Society* **67**, 70–80.

Haskell, C. (1992), Brownian Motion and Billiards on the Torus, PhD thesis, University of Stanford, Stanford.

Hénon, M. (1976), 'A two dimensional mapping with a strange attractor', *Communications in Mathematical Physics* **50**, 69–77.

Hilborn, R. (2000), *Chaos and Nonlinear Dynamics, an Introduction for Scientists and Engineers*, Oxford University Press, Oxford.

Hoefer, C. (2008), Causal determinism, *in* E. Zalta, ed., 'The Stanford Encyclopaedia of Philosophy (Winter 2008 Edition)', http://plato.stanford.edu/archives/win2008/entries/determinism-causal/, Stanford.

Hopf, E. (1932*a*), 'Complete transitivity and the ergodic principle', *Proceedings of the National Academy of Sciences of the United States of America* **18**, 204–209.

Hopf, E. (1932*b*), 'Proof of Gibbs' hypothesis on the tendency toward statistical equilibrium', *Proceedings of the National Academy of Sciences of the United States of America* **18**, 333–340.

Jacobson, M. (1981), 'Absolutely continuous invariant measures for one-parameter families of one-dimensional maps', *Communications in Mathematical Physics* **81**, 39–88.

Janssen, J. & Limnios, N. (1999), *Semi-Markov Models and Applications*, Kluwer Academic Publishers, Dordrecht, the Netherlands.

Kellert, S. (1993), *In the Wake of Chaos*, University of Chicago Press, Chicago.

Klir, G. (2006), *Uncertainty and Information: Foundations of Generalized Information Theory*, Wiley, Hoboken, New Jersey.

Kolář, M. & Gumbs, G. (1992), 'Theory for the experimental observation of chaos in a rotating waterwheel', *Physical Review A* **45**, 626–637.

Kolmogorov, A. (1933), *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Springer, Berlin.

Kolmogorov, A. (1958), 'A new metric invariant of transitive dynamical systems and automorphisms of Lebesgue spaces', *Dokl. Acad. Nauk SSSR* **119**, 861–864.

Koopman, B. (1931), 'Hamiltonian systems and transformations in Hilbert space', *Proceedings of the National Academy of Sciences of the United States of America* **17**, 315–318.

Koopman, B. & von Neumann, J. (1932), 'Dynamical systems of continuous spectra', *Proceedings of the National Academy of Sciences of the United States of America* **18**, 255–263.

Krieger, W. (1970), 'On entropy and generators of measure-preserving transformations', *Transactions of the American Mathematical Society* **149**, 453–456.

Lakatos, I. (1961), Essays in the Logic of Mathematical Discovery, PhD thesis, University of Cambridge, Cambridge.

Lakatos, I. (1976), *Proofs and Refutations, The Logic of Mathematical Discovery*, edited by John Worrall and Elie Zahar, Cambridge University Press, Cambridge.

Lakatos, I. (1978), *Mathematics, Science and Epistemology*, Philosophical Papers Volume 2, edited by John Worrall and Elie Zahar, Cambridge University Press, Cambridge.

Landsman, K. (2007), Between classical and quantum, *in* J. Butterfield & J. Earman, eds, 'Philosophy of Physics (Handbooks of the Philosophy of Science A)', North-Holland, Amsterdam, pp. 417–553.

Larvor, B. (1998), *Lakatos: an Introduction*, Routledge, London and New York.

Larvor, B. (2001), 'What is dialectical philosophy of mathematics?', *Philosophia Mathematica* **9**, 212–229.

Laskar, J. (1994), 'Letter to the editor: Large-scale chaos in the solar system', *Astronomy and Astrophysics* **287**, 9–12.

Lavis, D. (2010), An objectivist account of probabilities in statistical physics, *in* C. Beisbart & S. Hartmann, eds, 'Probabilities in Physics', forthcoming, Oxford University Press, Oxford.

Lebowitz, J. (1993), 'Macroscopic laws, microscopic dynamics, time's arrow and Boltzmann's entropy', *Physica A* **194**, 1–27.

Leiber, T. (1998), 'On the actual impact of deterministic chaos', *Synthese* **113**, 357–379.

Leng, M. (2002), 'Phenomenology and mathematical practice', *Philosophia Mathematica* **10**, 3–25.

Lichtenberg, A. J. & Lieberman, M. A. (1992), *Regular and Chaotic Dynamics*, Springer, Berlin, New York.

Lighthill, J. (1986), 'The recently recognized failure of predictability in newtonian dynamics', *Proceedings of the Royal Society of London, Series A* **407**, 35–50.

Lind, D. (1975), 'A counterexample to a conjecture of Hopf', *Duke Mathematical Journal* **42**, 755–757.

Lissauer, J. (1999), 'Chaotic motion in the solar system', *Reviews of Modern Physics* **71**, 835–845.

Lorenz, E. (1963), 'Deterministic nonperiodic flow', *Journal of the Athmospheric Sciences* **20**, 130–141.

Lorenz, E. (1964), 'The problem of deducing the climate from the governing equations', *Tellus* **XVI**, 1–11.

Luzzatto, S., Melbourne, I. & Paccaut, F. (2005), 'The Lorenz attractor is mixing', *Communications in Mathematical Physics* **260**, 393–401.

Lyubich, M. (2002), 'Almost every regular quadratic map is either regular or stochastic', *Annals of Mathematics* **156**, 1–78.

Mañé, R. (1987), *Ergodic Theory and Differentiable Dynamics*, Springer, Berlin.

Mackey, G. (1974), 'Ergodic theory and its significance for statistical mechanics and probability theory', *Advances in Mathematics* **12**, 178–268.

Mancosu, P. (2008), *The Philosophy of Mathematical Practice*, Oxford University Press, Oxford.

Marsden, J. & Hoffman, M. (1974), *Elementary Classical Analysis*, W.H. Freeman and Company, New York.

Martinelli, M., Dang, M. & Seph, T. (1998), 'Defining chaos', *Mathematics Magazine* **71**, 112–122.

May, R. (1976), 'Simple mathematical models with very complicated dynamics', *Nature* **261**, 459–467.

Mayer, D. & Roepstorff, G. (1983), 'Strange attractors and asymptotic measures of discrete-time dissipative systems', *Journal of Statistical Physics* **31**, 309–326.

Miller, D. (1996), The status of determinism in an uncontrollable world, *in* P. Weingartner & G. Schurz, eds, 'Law and Prediction in the Light of Chaos Research', Springer, Berlin, pp. 103–114.

Montague, R. (1962), Deterministic theories, *in* D. Wilner, ed., 'Decisions, Values and Groups', Pergamon Press, New York, pp. 325–370.

Moser, J. (1973), *Stable and Random Motions in Dynamical Systems*, Yale University Press, New Haven.

Nillsen, R. (1999), 'Chaos and one-to-oneness', *Mathematics Magazine* **72**, 14–21.

Norton, J. (2003), 'Causation as folk science', *Philosopher's Imprints* **3**, 1–22.

Ornstein, D. (1970*a*), 'Bernoulli-shifts with the same entropy are isomorphic', *Advances in Mathematics* **4**, 337–352.

Ornstein, D. (1970*b*), Imbedding Bernoulli shifts in flows, *in* A. Dold & B. Eckmann, eds, 'Contributions to Ergodic Theory and Probability, Proceedings of the First Midwestern Conference on Ergodic Theory held at the Ohio State University, March 27-30', Lecture Notes in Mathematics, vol. 160, Springer, Berlin, pp. 178–218.

Ornstein, D. (1971), 'Some new results in the Kolmogorov-Sinai theory of entropy and ergodic theory', *Bulletin of the American Mathematical Society* **77**, 878–890.

Ornstein, D. (1973*a*), 'An application of ergodic theory to probability theory', *The Annals of Probability* **1**, 43–58.

Ornstein, D. (1973*b*), 'The isomorphism theorem for Bernoulli flows', *Advances in Mathematics* **10**, 124–142.

Ornstein, D. (1974), *Ergodic Theory, Randomness, and Dynamical Systems*, Yale University Press, New Haven and London.

Ornstein, D. (1989), 'Ergodic theory, randomness and "chaos"', *Science* **243**, 182–187.

Ornstein, D. & Galavotti, G. (1974), 'Billiards and Bernoulli schemes', *Communications in Mathematical Physics* **38**, 83–101.

Ornstein, D. & Weiss, B. (1991), 'Statistical properties of chaotic systems', *Bulletin of the American Mathematical Society* **24**, 11–116.

Oseledec, V. (1968), 'Multiplicative ergodic theorem, Lyapunov characteristic numbers for dynamical systems', *Transactions of the Moscow Mathematical Society* **19**, 197–221.

Ott, E. (2002), *Chaos in Dynamical Systems*, Cambridge University Press, Cambridge.

Park, K. (1982), 'A special family of ergodic flows and their $\bar{d}$-limits', *Israel Journal of Mathematics* **42**, 343–353.

Peitgen, H.-O., Jürgens, H. & Saupe, D. (1992), *Chaos and Fractals, New Frontiers of Science*, Springer, New York.

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

Pitowsky, I. (1995), 'Laplace's demon consults an oracle: The computational complexity of prediction', *Studies in History and Philosophy of Modern Physics* **27**, 161–180.

Polya, G. (1949), 'With or without motivation', *The American Mathematical Monthly* **56**, 684–691.

Polya, G. (1954), *Patterns of Plausible Inference*, Volume II of Mathematics and Plausible Reasoning, Princeton University Press, Princeton.

Radunskaya, A. (1992), Statistical Properties of Deterministic Bernoulli Flows, PhD thesis, University of Stanford, Stanford.

Robinson, C. (1995), *Dynamical Systems, Stability, Symbol Dynamics and Chaos*, CRC Press, Tokyo.

Rohlin, V. (1960), 'New progress in the theory of transformations with invariant measure', *Russian Mathematical Surveys* **15**, 1–22.

Rudolph, D. (1976), 'A two-valued step coding for ergodic flows', *Mathematische Zeitschrift* **150**, 201–220.

Rudolph, D. (1990), *Fundamentals of Measurable Dynamics, Ergodic Theory on Lebesgue spaces*, Oxford University Press, Oxford.

Ruelle, D. (1997), 'Chaos, predictability, and idealizations in physics', *Complexity* **3**, 26–28.

Ruelle, D. & Takens, F. (1971), 'On the nature of turbulence', *Communications in Mathematical Physics* **20**, 167–192.

Schurz, G. (1996), Kinds of unpredictability in deterministic systems, *in* P. Weingartner & G. Schurz, eds, 'Law and Prediction in the Light of Chaos Research', Springer, Berlin, pp. 123–141.

Schuster, G. & Just, W. (2005), *Deterministic Chaos: an Introduction*, Wiley-VCH Verlag, Weinheim.

Scott, S. (1991), *Chemical Chaos*, Clarendon Press, Oxford.

Shields, P. (1973), *The Theory of Bernoulli-shifts*, University of Chicago Press, Chicago.

Shiryaev, A. (1989), 'Kolmogorov: life and creative activities', *The Annals of Probability* **17**, 866–944.

Sinai, Y. (1959), 'On the concept of entropy for dynamical systems', *Dokl. Acad. Nauk SSSR* **124**, 768–771.

Sinai, Y. (1963), 'Probabilistic ideas in ergodic theory', *American Mathematical Society Translations* **31**, 62–84.

Sinai, Y. (1989), 'Kolmogorov's work on ergodic theory', *The Annals of Probability* **17**, 833–839.

Sinai, Y. (2000), *Dynamical Systems, Ergodic theory and Applications*, Springer, Berlin.

Sinai, Y. (2007), 'Kolmogorov-Sinai entropy'. Scholarpedia, Retrieved from the World Wide Web on January 24, 2008: www.scholarpedia.org/article/Kolmogorov-Sinai_entropy.

Skinner, J., Goldberger, A., Mayer-Kress, G. & Ideker, R. (1997), 'Chaos in the heart: Implications for clinical cardiology', *Bio/Technology* **8**, 1018–1024.

Sklar, L. (1993), *Physics and Chance: Philosophical Issues in the Foundations of Statistical Mechanics*, Cambridge University Press, Cambridge.

Smith, L. (2007), *A Very Short Introduction to Chaos*, Oxford University Press, Oxford.

Smith, L., Ziehmann, C. & Fraedrich, K. (1999), 'Uncertainty dynamics and predictability in chaotic systems', *Quarterly Journal of the Royal Meteorological Society* **125**, 2855–2886.

Smith, P. (1998), *Explaining Chaos*, Cambridge University Press, Cambridge.

Stone, M. (1989), 'Chaos, prediction and Laplacian determinism', *American Philosophical Quarterly* **26**, 123–131.

Strogatz, S. (1994), *Nonlinear Dynamics and Chaos, with Applications to Physics, Biology, Chemistry, and Engineering*, Addison Wesley, New York.

Suppes, P. (1993), 'The transcendental character of determinism', *Midwest Studies in Philosophy* **18**, 242–257.

Suppes, P. (1999), 'The noninvariance of deterministic causal models', *Synthese* **121**, 181–198.

Suppes, P. & de Barros, A. (1996), Photons, billiards and chaos, *in* P. Weingartner & G. Schurz, eds, 'Law and Prediction in the Light of Chaos Research', Springer, Berlin, pp. 189–201.

Sźasz, D. (2000), *Hard Ball systems and the Lorentz gas*, Encyclopaedia of Mathematical Sciences 101, Springer, Berlin.

Tappenden, J. (2008*a*), Mathematical concepts and definitions, *in* P. Mancosu, ed., 'The Philosophy of Mathematical Practice', Oxford, Oxford University Press, pp. 256–275.

Tappenden, J. (2008*b*), Mathematical concepts: fruitfulness and naturalness, *in* P. Mancosu, ed., 'The Philosophy of Mathematical Practice', Oxford, Oxford University Press, pp. 276–301.

Uffink, J. (2007), Compendium to the foundations of classical statistical physics, *in* J. Butterfield & J. Earman, eds, 'Philosophy of Physics (Handbooks of the Philosophy of Science B)', North-Holland, Amsterdam, pp. 923–1074.

von Neumann, J. (1932*a*), 'Proof of the quasi-ergodic hypothesis', *Proceedings of the National Academy of Sciences of the United States of America* **18**, 70–82.

von Neumann, J. (1932*b*), 'Zur Operatorenmethode in der klassischen Mechanik', *The Annals of Mathematics* **33**, 587–642.

von Plato, J. (1994), *Creating Modern Probability*, Cambridge University Press, Cambridge.

Walters, P. (1982), *An Introduction to Ergodic Theory*, Springer, New York.

Weingartner, P. (1996), Under what transformations are laws invariant?, *in* P. Weingartner & G. Schurz, eds, 'Law and Prediction in the Light of Chaos Research', Springer, Berlin, pp. 47–88.

Weingartner, P. & Schurz, G. (1996), *Law and Prediction in the Light of Chaos Research*, Springer, Berlin.

Werndl, C. (2009*a*), 'Are deterministic descriptions and indeterministic descriptions observationally equivalent?', *Studies in History and Philosophy of Modern Physics* **40**, 232–242.

Werndl, C. (2009*b*), Deterministic versus indeterministic descriptions: Not that different after all?, *in* A. Hieke & H. Leitgeb, eds, 'Reduction, Abstraction, Analysis, Proceedings of the 31st International Ludwig Wittgenstein-Symposium', Ontos, Frankfurt, pp. 63–78.

Werndl, C. (2009*c*), 'Justifying definitions in matemathics—going beyond Lakatos', *Philosophia Mathematica* **doi:10.1093/philmat/nkp006**.

Werndl, C. (2009*d*), On the justification and formulation of mathematical definitions: the case of deterministic chaos, *in* M. Dorato, M. Rédei &

M. Suárez, eds, ‘Proceedings of the First Conference of the European Philosophy of Science Association’, forthcoming, Springer, Berlin.

Werndl, C. (2009*e*), ‘What are the new implications of chaos for unpredictability?’, *The British Journal for the Philosophy of Science* **60**, 195–220.

Wiggins, S. (1990), *Introduction to Applied Nonlinear Dynamical Systems and Chaos*, Text in Applied Mathematics 2, Springer.

Winnie, J. (1998), Deterministic chaos and the nature of chance, *in* J. Earman & J. Norton, eds, ‘The Cosmos of Science – Essays of Exploration’, Pittsburgh University Press, Pittsburgh, pp. 299–324.

Young, L.-S. (1997), ‘Ergodic theory and chaotic dynamical systems, XII-th international congress of mathematical physics’, *Reviews of Modern Physics* **57**, 617–654.

Young, L.-S. (2002), ‘What are SRB measures, and which dynamical systems have them?’, *Journal of Statistical Physics* **108**, 733–754.

Zaslavsky, G. (2005), *Hamiltonian Chaos and Fractional Dynamics*, Oxford University Press, Oxford.

Ziehmann, C., Smith, L. & Kurths, J. (1986), ‘Localized Lyapunov exponents and the prediction of predictability’, *Physics Letters A* **271**, 1–15.