Are Deterministic Descriptions And Indeterministic Descriptions Observationally Equivalent?

Charlotte Werndl, Faculty of Philosophy, Cambridge University

Forthcoming in: Studies in History and Philosophy of Modern Physics

June 16, 2009

Abstract

The central question of this paper is: are deterministic and indeterministic descriptions observationally equivalent in the sense that they give the same predictions? I tackle this question for measure-theoretic deterministic systems and stochastic processes, both of which are ubiquitous in science. I first show that for many measure-theoretic deterministic systems there is a stochastic process which is observationally equivalent to the deterministic system. Conversely, I show 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 measure-theoretic deterministic systems which are observationally equivalent to stochastic processes used in science do not include any deterministic systems used in science. I argue that this is not so because deterministic systems used in science even give rise to Bernoulli 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. By proving results in ergodic theory, I show that also this guess is misguided: there are several deterministic systems used in science which give the same predictions at every observation level as Markov processes. All these results show that measure-theoretic deterministic systems and stochastic processes are observationally equivalent more often than one might perhaps expect. Furthermore, I criticise the claims of the previous philosophy papers Suppes and de Barros (1996), Winnie (1998) and Suppes (1999) on observational equivalence.

Contents

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 and Schurz, 1996, p. 203) that deterministic and indeterministic descriptions are not observationally equivalent. However, the question whether these descriptions are observationally equivalent has hardly been discussed.

This paper aims to contribute to fill this gap. Namely, the central questions of this paper 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 paper are measure-theoretic deterministic systems and stochastic processes, respectively. Both are ubiquitous in science. Because of lack of space, I concentrate on descriptions where the time varies in discrete steps; but I point out that analogous results also hold for a continuous time parameter.

More specifically, when saying that a deterministic system and a stochastic process are *observationally equivalent*, I mean the following: the deterministic system, when observed, gives the same predictions as the stochastic process. In what follows, when I say that a stochastic process can be *replaced* by a deterministic system, or conversely, I mean that it can be replaced by such a system in the sense that they are observationally equivalent.

This paper proceeds as follows. In section 2 I will introduce stochastic processes and measure-theoretic deterministic systems. In section 3 I will show that measure-theoretic deterministic systems and stochastic processes can often be replaced by each other. Given this, one might still guess that it is impossible to replace stochastic processes of the kinds in fact used in science by measure-theoretic deterministic systems that are used in science. One might also guess that it is impossible to replace 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 4 that these two guesses are wrong. Therefore, stochastic processes and deterministic systems can be replaced by each other more often than one might perhaps expect. Furthermore, I will criticise the claims of the previous philosophical papers Suppes and de Barros (1996), Winnie (1998) and Suppes (1999) on observational equivalence.

2 Stochastic processes and deterministic systems

The indeterministic and deterministic descriptions I deal with are stochastic processes and measure-theoretic deterministic systems, respectively. There are two types of them: either the time parameter is discrete (discrete processes and systems) or there is a continuous time parameter (continuous processes and systems). I consider only discrete descriptions, but analogous results hold for continuous descriptions, and these results are discussed in Werndl (2009c).

2.1 Stochastic processes

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.[1]

Let me formally define stochastic processes.[2] A *random variable* is a measurable function $Z : \Omega \to \bar{M}$ from a probability space, i.e. a measure space $(\Omega, \Sigma_\Omega, \nu)$ with $\nu(\Omega) = 1$, to a measurable space $(\bar{M}, \Sigma_{\bar{M}})$ where $\Sigma_{\bar{M}}$ denotes a σ-algebra on $\bar{M}$.[3] 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 1 *A* stochastic process $\{Z_t;\, t \in \mathbb{Z}\}$ *is a one-parameter family of random variables* Z_t, $t \in \mathbb{Z}$, *defined on the same probability space* $(\Omega, \Sigma_\Omega, \nu)$ *and taking values in the same measurable space* $(\bar{M}, \Sigma_{\bar{M}})$.

[1] 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).

[2] I assume basic knowledge about measure theory and modern probability theory. For more details, see Doob (1953), Cornfeld et al. (1982) and Petersen (1983).

[3] For simplicity, I assume that any measure space is complete, i.e. every subset of a measurable set of measure zero is measurable.

The set $\bar{M}$ is called the *outcome space* of the stochastic process. The bi-infinite sequence $r(\omega) := (\dots Z_{-1}(\omega), Z_0(\omega), Z_1(\omega) \dots)$ for $\omega \in \Omega$ 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 $r(\omega)$ represents the description of that history by giving the score at each t.

I will often be concerned with stationary stochastic processes. These are processes whose probabilistic laws do not change with time:

Definition 2 *A stochastic process $\{Z_t;\, t \in \mathbb{Z}\}$ is* stationary *if and only if the distributions 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 $t_1, \dots, t_n \in \mathbb{Z}$, $n \in \mathbb{N}$, and all $h \in \mathbb{Z}$ (Doob, 1953, p. 94).*

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

The following stochastic processes will accompany us throughout the paper. They are probably the most widely known.

Example 1: 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. The mathematical definition proceeds as follows. The random variables $X_1, \dots, X_n$, $n \in \mathbb{N}$, are 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 independent if and only if any finite number of them is independent.

Definition 3 *$\{Z_t;\, t \in \mathbb{Z}\}$ 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) $P\{Z_t = s_k\} = p_k$ for all $t \in \mathbb{Z}$ and all $k, 1 \leq k \leq N$; and (iii) $\{Z_t;\, t \in \mathbb{Z}\}$ are 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 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 sequences $\omega = (\dots \omega_{-1}, \omega_0, \omega_1 \dots)$ 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 generated by the cylinder-sets

$$C^{A_1\dots A_n}_{i_1\dots i_n}=\{\omega\in\Omega\,|\,\omega_{i_1}\in A_1,\dots,\omega_{i_n}\in A_n,\,A_j\in\Sigma_{\bar{M}},\ i_j\in\mathbb{Z},\ i_1<\dots<i_n,\ 1\leq j\leq n\}. \tag{1}$$

Since the outcomes are independent, these sets have probability $\bar{\nu}(C^{A_1\dots A_n}_{i_1\dots i_n}) := P\{Z_{i_1}\in A_1\}\dots P\{Z_{i_n}\in A_n\}$. 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.

2.2 Deterministic systems

According to the canonical definition, a description is *deterministic* exactly if any two solutions that agree at one time agree at all times (Butterfield, 2005). I call a sequence which describes the evolution of a deterministic description over time a solution.

This paper is concerned with measure-theoretic deterministic descriptions, in short deterministic systems:

Definition 4 *A* deterministic system *is a quadruple* (M,Σ_M,μ,T) *consisting of a probability space* (M,Σ_M,μ) *and a bijective measurable function* $T\colon M\to M$.

The *solution* through m, $m\in M$, is the sequence $(T^t(m))_{t\in\mathbb{Z}}$. M is the set of all possible states called the *phase space*; and T, which describes how solutions evolve, is called the *evolution function*. Clearly, Definition 4 defines systems which are deterministic according to the above canonical definition.

When observing a deterministic system, 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 the measurable space (M_O,Σ_{M_O}) (cf. Ornstein and Weiss, 1991, p. 16).

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

Definition 5 *A* measure-preserving deterministic system *is a deterministic system* (M,Σ_M,μ,T) *where the measure* μ *is invariant, i.e. for all* $A\in\Sigma_M$

$$\mu(T(A))=\mu(A). \tag{2}$$

Measure-preserving deterministic systems are important models in physics but are also important in other sciences such as biology, geology etc. This is so because condition (2) is not very restrictive. For first, all deterministic Hamiltonian systems and 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). Second, an invariant measure need not be the Lebesgue measure, i.e. measure-preserving deterministic systems need not be volume-preserving. Indeed, systems which are not volume-preserving (called 'dissipative systems') can often be modeled as measure-preserving systems. For instance, the long-term behaviour of a large class of deterministic systems can be modeled by measure-preserving systems (Eckmann and Ruelle, 1985). More generally, the potential scope of measure-preserving deterministic systems is quite wide: although some evolution functions do not have invariant measures, for very wide classes of evolution functions invariant measures are proven to exist. For instance, if T is a continuous function on a compact metric space, there exists at least one invariant measure (Mañé, 1987, p. 52).

I adopt the common assumption that invariant measures can be interpreted as probability measures. This deep issue has been discussed in statistical mechanics but is not the focus of this paper. I only mention two interpretations that naturally suggest interpreting measures as probability. According to the time-average interpretation, the measure of a set A is the long-run average of the time that a solution spends in A. According to the ensemble interpretation, the measure of a set A at t corresponds to the fraction of solutions starting from some set of initial conditions that are in A at time t [cf. Eckmann and Ruelle, 1985, pp. 625-627; Lavis, forthcoming].

The following deterministic system will accompany us.

INSERT THE FIGURE HERE WITH CAPTION: 'The baker's system on $0 \leq y \leq 1/2$'.

Example 2: The baker's system.
On the unit square $M := [0,1] \times [0,1]$ consider

$$T(x,y) = (2x, \frac{y}{2}) \text{ if } 0 \leq x < \frac{1}{2};\ (2x-1, \frac{y+1}{2}) \text{ if } \frac{1}{2} \leq x \leq 1. \tag{3}$$

Figure 1 illustrates that the baker's system first stretches the unit square to twice its length and half its width; then it cuts the rectangle obtained in half and places the right half on top of the left. For the Lebesgue measure μ and the Lebesgue σ-algebra Σ_M one obtains the measure-preserving deterministic system (M, Σ_M, μ, T). This system also has physical meaning: e.g. it describes the movement of a particle with initial position (x, y) in the unit square. The particle moves with constant speed perpendicular to the unit square. It bounces on several mirrors, causing it to return to the unit square at $T(x, y)$ (Pitowsky, 1995, p. 166).

3 Basic observational equivalence

Let me turn to some results about observational equivalence which are basic in the sense that they are about the question whether, given a deterministic system, it is possible to find *any* stochastic process which is observationally equivalent to the deterministic system, and conversely.

How can a stochastic process and a deterministic system yield the same predictions? When a 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 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 that perhaps 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 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 deterministic system are observationally equivalent'* is: *(i) the set of possible outcomes of the stochastic process is identical to the the set of possible observed values of the deterministic system, 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 replaced by stochastic processes. Then I will investigate when stochastic processes can be replaced by deterministic systems.

3.1 Deterministic systems replaced by stochastic processes

Let (M, Σ_M, μ, T) be a deterministic system. According to the canonical Definition 1, $Z_t(x) := T^t(x)$ is a stochastic process with exactly the same predictions as the deterministic system. However, this process is evidently equivalent to the original deterministic system, and the probabilities that one value leads to another one are trivial (0 or 1). Hence it is still "really" a deterministic system.

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 deterministic system (M, Σ_M, μ, T) is observed with $\Phi : M \rightarrow M_O$. According to Definition 1, $\{Z_t := \Phi(T^t);\ t \in \mathbb{Z}\}$ is a stochastic process. It is constructed by applying Φ to the deterministic system. Hence the outcomes of the stochastic process are the observed values of the deterministic system; and the realisations of the process and the solutions of the 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* $\{\Phi(T^t);\ t \in \mathbb{Z}\}$. But the important question is whether $\{\Phi(T^t);\ t \in \mathbb{Z}\}$ is nontrivial. Indeed, *the stochastic process* $\{\Phi(T^t);\ t \in \mathbb{Z}\}$ *is often nontrivial.* I show now one result in this direction; besides, several other results also indicate this (cf. Cornfeld et al., 1982, pp. 178-179).[4]

Before I can proceed, the following definitions are needed:

Definition 6 *A measure-preserving deterministic system* (M, Σ_M, μ, T) *is* ergodic *if and only if for all* $A, B \in \Sigma_M$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \left(\mu(T^n(A) \cap B) - \mu(A)\mu(B)\right) = 0. \tag{4}$$

[4] For instance, if K-systems are observed with a finite-valued observation function, one obtains nontrivial stochastic processes because for K-systems the entropy of any finite partition is positive (cf. Petersen, 1983, p. 63).

A *partition* of a measure space (M, Σ_M, μ) is a set $\alpha = \{\alpha_1, \dots, \alpha_n\}$ with $\alpha_i \in \Sigma_M,\ n \in \mathbb{N}$, such that $\bigcup_{i=1}^{n} \alpha_i = M,\ \mu(\alpha_i) > 0$, and $\alpha_i \cap \alpha_j = \emptyset$ for $i \neq j,\ 0 \leq i, j \leq n$. A partition is *nontrivial* if and only if it has more than one element. 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. A finite-valued observation function is called *nontrivial* if and only if its corresponding partition is nontrivial (cf. Cornfeld et al., 1982, p. 179).

The following proposition shows that for ergodic deterministic systems for which there is no nontrivial set which is eventually mapped onto itself, and *every* finite-valued observation function, the stochastic process $\{\Phi(T^t);\ t \in \mathbb{Z}\}$ is nontrivial. That is, 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 from o_i to o_j is smaller than 1. Hence there are two or more observed values that can follow o_i; and the probability that o_i moves to any of these observed values is between 0 and 1. This is a strong result because irrespective of how detailed one looks at the deterministic system, one always obtains a nontrivial stochastic process.

Proposition 1 *Assume that the deterministic system (M, Σ_M, μ, T) is ergodic and that 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 for every nontrivial finite-valued observation function $\Phi : M \to M_O$ and the stochastic process $\{Z_t := \Phi(T^t);\ t \in \mathbb{Z}\}$ the following holds: there is an $o_i \in M_O$ such that for all $o_j \in M_O$, $P\{Z_{t+1} = o_j \,|\, Z_t = o_i\} < 1$.*[5]

For a proof, see subsection 6.1. For instance, the baker's system (Example 2) is weakly mixing, and thus any finite-valued observation function gives rise to a nontrivial stochastic process.

Measure-preserving systems are typically what is called 'weakly mixing'[6]

[5]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 $A, B \subseteq \Sigma_{\bar{M}}$ with $P\{Z \in B\} > 0$.

[6](M, Σ_M, μ, T) is weakly mixing if and only if for all $A, B \in \Sigma_M$

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |\mu(T^n(A) \cap B) - \mu(A)\mu(B)| = 0. \tag{5}$$

(Halmos, 1944). It is easy to see that any weakly mixing system satisfies the assumption of Proposition 1 (weakly mixing is stronger than this assumption).[7] Hence Proposition 1 shows that for typical measure-preserving deterministic systems any finite-valued observation function yields a nontrivial stochastic process.

Yet Proposition 1 does not say much about whether the measure-preserving deterministic systems encountered in science fulfill the assumption of Proposition 1 because those systems constitute a small class of all measure-preserving systems. Indeed, the KAM theorem says that the phase space of integrable Hamiltonian systems which are perturbed by a small nonintegrable perturbation breaks up into stable and unstable regions. With increasing perturbation the unstable regions become larger and often eventually cover nearly the entire hypersurface of constant energy. Because a solution is confined to a region, the KAM theorem implies that the discrete versions of infinitely differentiable Hamiltonian systems are typically not ergodic (Berkovitz et al., 2006, section 4). (I call the discrete-time systems obtained by looking at a continuous-time system S at points of time nt_0, $n \in \mathbb{N}$, $t_0 \in \mathbb{R}$ arbitrary, $t_0 \neq 0$, the *discrete versions of* S).[8]

Despite this, Proposition 1 applies to several systems encountered in science. First, a motion is chaotic just in case it is deterministic yet also unstable because nearby initial conditions eventually lead to very different outcomes. I will not need a more exact definition; but I will call a system *chaotic* if the motion is chaotic on the entire phase space and *locally chaotic* if the motion is chaotic on a region of phase space. Chaotic systems are usually regarded as weakly mixing (Berkovitz et al., 2006, p. 688; Werndl, 2009a, section 3). And as will be argued in subsection 4.1, there are several physically relevant chaotic and weakly mixing systems. Moreover, in subsection 4.2.2 it will be shown that there are even systems which are neither chaotic nor locally chaotic but which satisfy Proposition 1. Second, even

[7] First, it is clear that weakly mixing systems are ergodic. Second, assume that for a weakly mixing 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 (5) cannot hold for $A := C$ and $B := C$. In subsection 4.2.2 I will show that the irrational rotation on the circle satisfies the assumption of Proposition 1 but is not weakly mixing.

[8] Alternatively, continuous-time deterministic systems can be discretised by considering the successive hits of a trajectory on a suitable Poincaré section. All I say about discrete versions of continuous systems also holds true for discrete-time systems arising in this way, except that the latter are more often ergodic (Berkovitz et al., 2006, pp. 680–685).

if the whole system does not satisfy the assumption of Proposition 1, the motion of the system restricted to some regions of phase space might well satisfy this assumption. In fact, Proposition 1 immediately implies the following result. Assume that for a measure-preserving system (M, Σ_M, μ, T) there is a $A \in \Sigma_M, \mu(A) > 0$, such that the system restricted to A[9] fulfills the assumption of Proposition 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 $k, l,\ k \neq l$, such that $\mu(A \cap \alpha_k) \neq 0$ and $\mu(A \cap \alpha_l) \neq 0$, there is an outcome $o_i \in M_O$ such that for all outcomes $o_j \in M_O$ it holds that $P\{Z_{t+1} = o_j \,|\, Z_t = o_i\} < 1$. In particular, although mathematically little is known, it is conjectured that the motion restricted to unstable regions of KAM-type systems is weakly mixing (Berkovitz et al., 2006, section 4; Werndl, 2009a, section 3). If this is true, then my argument shows that for many observation functions of KAM-type systems one obtains nontrivial stochastic processes.

3.2 Stochastic processes replaced by deterministic systems

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

The following idea of how to replace stochastic processes by deterministic systems is well known in the technical literature (Petersen, 1983, pp. 6–7)[10] and known to philosophers (Butterfield, 2005); I also need to discuss it for what follows later. The underlying thought is that for each realisation $r(\omega)$, one sets up a deterministic system with phase space $\{r(\omega)\}$. So consider a 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 be the σ-algebra generated by the cylinder-sets as defined in (1) at the end of subsection 2.1. $\{Z_t; t \in \mathbb{Z}\}$ 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

[9] That is, the system $(A, \Sigma_{M \cap A}, \mu_A, T_A)$, where $\Sigma_{M \cap A} := \{B \cap A |\, B \in \Sigma_M\}$, $\mu_A(X) := \frac{\mu(X)}{\mu(A)}$, and T_A denotes T restricted to A.

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

this pre-measure to a measure on Σ_M. Let $T : M \to M$ be the left shift, i.e. $T((\ldots m_{-1}m_0m_1\ldots)) := (\ldots m_0m_1m_2\ldots)$. T is bijective and measurable, and so 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 $\Phi_0 : M \to \bar{M}, \Phi_0(m) = m_0$. I now define:

Definition 7 $(M, \Sigma_M, \mu, T, \Phi_0)$ *as constructed above is the* deterministic representation *of the process* $\{Z_t;\, t \in \mathbb{Z}\}$.

For the deterministic representation $(M, \Sigma_M, \mu, T, \Phi_0)$ of $\{Z_t; t \in \mathbb{Z}\}$ it is assumed that the 0-th coordinate is observed. Consequently, the possible outcomes of $\{Z_t; t \in \mathbb{Z}\}$ are the possible observed values of $(M, \Sigma_M, \mu, T, \Phi_0)$. Clearly, any realisation $r(\omega)$ of the process is contained in M, and observing the solution $(T^t(r(\omega)))_{t\in\mathbb{Z}}$ with Φ_0 exactly gives $r(\omega)$. Furthermore, the measure μ is defined by the probabilities which are assigned by $\{Z_t; t \in \mathbb{Z}\}$ to each cylinder set. Hence the probability distribution over the realisations of $\{Z_t; t \in \mathbb{Z}\}$ is the same as the one over the sequences of observed values of $(M, \Sigma_M, \mu, T, \Phi_0)$. 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 replaced by at least one deterministic system.* (When there is no risk of confusion, I also refer to the system (M, Σ_M, μ, T) of the deterministic representation $(M, \Sigma_M, \mu, T, \Phi_0)$ as the deterministic representation.)

For Bernoulli processes (Example 1) the deterministic representation is the following. (M, Σ_M, μ) is the measure space $(\Omega, \Sigma_\Omega, \nu)$ as defined at the end of subsection 2.1. $T((\ldots \omega_{-1}\omega_0\omega_1\ldots)) := (\ldots \omega_0\omega_1\omega_2\ldots)$ for $\omega \in \Omega$ and $\Phi_0(\omega) = \omega_0$.

Definition 8 *The deterministic representation* (M, Σ_M, μ, T) *of the Bernoulli process with probabilities* $p_1, \ldots, p_N,\ N \in \mathbb{N}$, *is called the* Bernoulli shift *with probabilities* $(p_1, \ldots, p_N)$.

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 replace a stochastic process. I will turn to this question in section 4, where I show that for some stochastic processes this is indeed the case. To my knowledge, it is unknown whether *every* stochastic process can be thus replaced.

3.3 A mathematical definition of observational equivalence

Let me now mathematically define what it means for a stochastic process and a deterministic system to be observationally equivalent. The notion of isomorphism captures the idea that 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.

Definition 9 *$(M_1, \Sigma_{M_1}, \mu_1, T_1)$ is* isomorphic *to $(M_2, \Sigma_{M_2}, \mu_2, T_2)$ (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 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 $m \in \hat{M}_1$ (cf. Petersen, 1983, p. 4).*

One easily sees that 'being isomorphic' is an equivalence relation. Isomorphic systems may have different phase spaces. If identical sets $\hat{M}_1$ and $\hat{M}_2$ can be found, then the 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 deterministic system (M, Σ_M, μ, T), observed with Φ, gives the same predictions as $\{Z_t | \, t \in \mathbb{Z}\}$ exactly if (i) the outcomes of $\{Z_t | \, t \in \mathbb{Z}\}$ are the observed values of (M, Σ_M, μ, T), and (ii) the deterministic representation of $\{\Phi(T^t);\, t \in \mathbb{Z}\}$ is probabilistically equivalent to the deterministic representation of $\{Z_t | \, t \in \mathbb{Z}\}$. Hence one arrives at the following definition of 'observational equivalence'; (for what follows, a definition for measure-preserving systems and, correspondingly, stationary stochastic processes will suffice):[11]

[11]For a measure-preserving system (M, Σ_M, μ, T) the 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, $\{x \in M \,|\, \Phi(T^{t_1+h}(x)) \in A_1, \ldots, \Phi(T^{t_n+h}(x)) \in A_n\}$ is $B := T^{-(t_1+h)}(\Phi^{-1}(A_1) \cap \ldots \cap T^{t_1-t_n}\Phi^{-1}(A_n))$. Because the system is measure-preserving, $\mu(A) = \mu(B)$, implying that $\{\Phi(T^t);\, t \in \mathbb{Z}\}$ is stationary. And if $\{Z_t | \, t \in \mathbb{Z}\}$ is stationary, its deterministic representation (M, Σ_M, μ, T) is measure-preserving. For stationarity implies that $\mu(T(A)) = \mu(A)$ for any cylinder set A, and hence that $\mu(T(A)) = \mu(A)$ for all $A \in \Sigma_M$ (cf. Cornfeld et al., 1982, p. 178).

Definition 10 *The stationary stochastic process* $\{Z_t;\ t \in \mathbb{Z}\}$ *and the measure-preserving deterministic system* (M, Σ_M, μ, T)*, observed 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* $\{Z_t;\ t \in \mathbb{Z}\}$*.*

All the cases of observational equivalence already discussed are cases of observational equivalence in the sense of Definition 10. First, I claimed in subsection 3.1 that (M, Σ_M, μ, T) observed with Φ is observationally equivalent to the stochastic process $\{\Phi(T^t);\ t \in \mathbb{Z}\}$. This is true because every system is manifestly isomorphic to itself. Second, I claimed in subsection 3.2 that the deterministic representation $(M, \Sigma_M, \mu, T, \Phi_0)$ of $\{Z_t;\ t \in \mathbb{Z}\}$ is observationally equivalent to $\{Z_t;\ t \in \mathbb{Z}\}$. This is true because the deterministic representation of $\{\Phi_0(T^t);\ t \in \mathbb{Z}\}$ is $(M, \Sigma_M, \mu, T, \Phi_0)$.

One final point: assume that (M, Σ_M, μ, T) is isomorphic via $\phi : \hat{M} \to \hat{M}_2$ to the deterministic representation $(M_2, \Sigma_{M_2}, \mu_2, T_2, \Phi_0)$ of $\{Z_t;\ t \in \mathbb{Z}\}$. 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) restricted to $\hat{M}$ and observed with $\Phi_0(\phi(m))$ 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 following definition will be important later:

Definition 11 (X, Σ, μ, T) *is a* Bernoulli system *if and only if it is isomorphic to a Bernoulli shift.*

The meaning of Bernoulli systems is clear, viz. the solutions of a Bernoulli system can put into one-to-one correspondence with the realisations of a Bernoulli process. Thus a Bernoulli system, observed with $\Phi_0(\phi)$, produces a Bernoulli process. Finally, I note the important result that two Bernoulli shifts (and hence two 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 $\sum_{i=1}^{n} -p_i \log_2 p_i$ (Ornstein, 1974, pp. 5–3; Werndl, 2009b).

4 Advanced observational equivalence

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

4.1 Deterministic system in science which replace stochastic processes in science

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

Bernoulli processes (Example 1) are often regarded as the most random discrete-time stochastic processes because their outcomes are independent (cf. Ornstein, 1989). Are there deterministic systems in science which, when observed, are observationally equivalent to Bernoulli processes? And are there even deterministic systems in science which are Bernoulli systems? Historically, it was long thought that the answer to these questions is negative (cf. Sinai, 1989, p. 834). So it was a big surprise when it was discovered from the 1960s onwards that *there are several deterministic systems in science which are Bernoulli systems* (among them systems producing Bernoulli processes with equiprobable outcomes). Let me mention some of the most important examples, which are also some of the most important examples of chaotic systems.[12]

To start with, there are systems in Newtonian mechanics, some of which are simple models of statistical mechanical systems, whose discrete versions are proven to be Bernoulli systems. The most prominent examples are: first, some hard sphere systems, which describe the motion of a number of hard

[12]Bernoulli systems are regarded as strongly chaotic.

spheres undergoing elastic reflections 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 $m \geq N$; second, billiard systems with convex obstacles; and third, geodesic flows of negative curvature, i.e. frictionless motion of a particle moving with unit speed on a compact manifold with everywhere negative curvature. It is usually very hard to prove that systems are Bernoulli. Therefore, for many systems it is only conjectured that their discrete versions are Bernoulli, e.g., for *all* hard sphere systems and the motion of KAM-type systems restricted to some regions of phase space (Ornstein and Weiss, 1991, section 4; Young, 1997; Berkovitz et al., 2006, p. 679–680).

Furthermore, there are dissipative systems which are Bernoulli systems: such as the logistic map and generalised versions thereof, the Hénon map and generalised versions thereof, and the discrete versions of the Lorenz system and generalised versions thereof. Some of these systems give relatively accurate predictions, e.g. the Lorenz system as a model for water-wheels. Yet often these systems are motivated as simple models which help us to understand, and not so much to predict, phenomena: e.g. the logistic map for population and climate dynamics, and the Hénon map and the Lorenz system for weather dynamics (Lorenz, 1964; May, 1976; Jacobson, 1981; Benedicks and Young, 1993; Smith, 1998, chapter 8; Lyubich, 2002; Luzzatto, 2005).

Also the baker's system (M, Σ_M, μ, T) (Example 2), a somewhat artificial example of deterministic motion, is Bernoulli. 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

$$x = 0.\omega_0\omega_1 \ldots = \sum_{i=1}^{\infty} \frac{\omega_{i-1}}{2^i}; \quad y = 0.\omega_{-1}\omega_{-2} \ldots = \sum_{i=1}^{\infty} \frac{\omega_{-i}}{2^i}, \tag{6}$$

where the representation of numbers of the form $j2^{-n}, n \in \mathbb{N}$, $0 \leq j \leq 2^n$, is assumed to end with an infinite sequences of zeros. Consider the Bernoulli shift $(M_2, \Sigma_{M_2}, \mu_2, T_2)$ with states 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 ending with an infinite sequence of 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 $\alpha = \{\alpha_1, \alpha_2\} := \{[0, \frac{1}{2}) \times [0, 1], [\frac{1}{2}, 1] \times [0, 1]\}$ yields the Bernoulli process with states s_1, s_2 and probabilities $(\frac{1}{2}, \frac{1}{2})$.

A Bernoulli system is weakly mixing (Petersen, 1983, p. 58). Hence, provided it is observed with a finite-valued observation function, one always obtains a nontrivial stochastic process (Proposition 1).

What is the significance of the these results? They show that the conjecture advanced at the beginning of this subsection is wrong: *it is possible to replace stochastic processes in science by deterministic systems in science.*[13]

Of course, the question arises whether for deterministic systems in science which are observationally equivalent to stochastic processes 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 and the phenomenon under consideration. For some systems the observation function is very involved and thus no 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. 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.

4.2 Stochastic processes which replace deterministic systems in science at every observation level

4.2.1 ε-congruence and replacement by Markov processes

The previous discussion showed that for several deterministic systems in science, regardless which finite-valued observation function one applies, one always obtains a stochastic process. But to obtain systems 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 deterministic systems in science one still obtains stochastic processes in science. In particular, it is hard to imagine that one still obtains random stochastic processes. Therefore, one might conjecture that *it is impossible to replace deterministic systems in science at every observation level by stochastic processes in science.*

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

Let me introduce one of the most natural ways of understanding the phrase ‘at any observation level’, i.e. the notion that stochastic processes of a certain type replace a deterministic system at any observation level. I first explain what it means for a 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 a deterministic system and a stochastic process give the same predictions at observation level ε if the following holds: the solutions of the 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.

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 the possible outcomes of the stochastic process are a subset of the phase space of the deterministic system. For what follows, it suffices to consider measure-preserving deterministic systems and, correspondingly, stationary processes. Now recall Definition 7 of the deterministic representation and Definition 9 of being isomorphic. So finally, I can define:

Definition 12 *Let (M, Σ_M, μ, T) be a 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 $\{Z_t;\, t \in \mathbb{Z}\}$, which takes values in (M, d_M), i.e. $\Phi_0 : M_2 \to M$. (M, Σ_M, μ, T) is ε-*congruent *to $(M_2, \Sigma_{M_2}, \mu_2, T_2, \Phi_0)$ 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 (cf. Ornstein and Weiss, 1991, pp. 22–23).*

Note that ε-congruence does not assume that the deterministic system is observed with an observation function. Of course, observation functions can be introduced. Assume one observes a deterministic system with an observation function. Then there is a stochastic process which is ε-congruent to the deterministic system such that the probabilistic predictions resulting from the observation function differ at most by ε from the probabilistic predictions obtained by applying the observation function to the ε-congruent stochastic process.

By generalising over ε, one obtains a natural meaning of the notion that stochastic processes of a certain type replace a measure-preserving deterministic system at any 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 deterministic system.*

For Bernoulli processes the next outcome of the process is independent of its previous outcomes. So, intuitively, it seems clear that deterministic systems in science, for which the next state of the system is constrained by its previous states (because of the underlying determinism at the level of states), cannot be replaced 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 theorem shows that, for our notion of replacement at every observation level, this idea is indeed correct under very mild assumptions, which hold for deterministic systems in science.[14] Hence this theorem shows a limitation on the observational equivalence of deterministic systems and stochastic processes.

Theorem 1 *Let (M, Σ_M, μ, T) be a measure-preserving deterministic system where Σ_M contains all open balls of the metric space (M, d_M), 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 $d(T(x), D) := \inf\{d(T(x), m) \mid m \in D\} > 0$. Then there is some $\varepsilon > 0$ for which there is no Bernoulli process to which (M, Σ_M, μ, T) is ε-congruent.*

For a proof, see subsection 6.2.[15]

[14] This theorem also holds for *generalised Bernoulli processes*—stochastic processes consisting of independent and identically distributed random variables whose outcome space need not be finite. That is, under the assumption of Theorem 1, there is an $\varepsilon > 0$ for which there is no generalised Bernoulli process to which (M, Σ_M, μ, T) is ε-congruent (cf. Remark 1 at the end of subsection 6.2).

[15] A deterministic system which is replaced at every observation level by a Bernoulli

Given this result, it is natural to ask (which, incidentally, Smith (1998) does not do) whether deterministic systems in science can be replaced at every observation level by other stochastic processes in science. *The answer is 'yes'. Besides, all one needs are irreducible and aperiodic Markov processes, which are widely used in science.* These Markov processes are often regarded as random; in particular, Bernoulli processes are regarded as the most random processes and Markov processes as the next most random (Ornstein and Weiss, 1991, p. 38 and p. 66).

For Markov processes the next outcome depends only on the previous outcome.

Definition 13 $\{Z_t;\, t \in \mathbb{Z}\}$ *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)* $P\{Z_{t+1} = s_j \,|\, Z_t, Z_{t-1} \ldots, Z_k\} = P\{Z_{t+1} = s_j \,|\, Z_t\}$ *for any* t*, any* $k \in \mathbb{Z},\ k \leq t$*, and any* $s_j \in \bar{M}$*; and (iii)* $P\{X_{t+1} = s_j \,|\, X_t = s_i\}$ *for any* $s_i, s_j \in \bar{M}$ *is independent of* $t,\ t \in \mathbb{Z}$*.*

Clearly, such a process is stationary.

Define $P^k(s_i, s_j) := P\{X_{n+k} = s_i \,|\, X_n = s_j\}$ 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 exactly if for every possible outcome there is no periodic pattern in which the process can visit that outcome. Mathematically, the *period* d_{s_i} of an outcome $s_i \in \bar{M},\ 1 \leq i \leq N$, is defined by $d_i = gcd\{n \geq 1 \,|\, P^n(s_i, s_i) > 0\}$ 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.

The following theorem shows that Bernoulli systems (cf. Definition 11) can be replaced at every observation level by irreducible and aperiodic Markov processes.

Theorem 2 *Let* (M, Σ_M, μ, T) *be a Bernoulli system where the metric space* (M, d_M) *is separable*[16] *and* 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.*

---

process will be measure-preserving. Hence it suffices to concentrate on measure-preserving systems.

[16] (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, see subsection 6.3. The assumptions in this theorem are fulfilled by all Bernoulli systems in science.

The following theorem shows that, for our notion of replacement at every observation level, also *only* Bernoulli systems can be replaced by irreducible and aperiodic Markov processes.

Theorem 3 *The deterministic representation of any irreducible and aperiodic Markov process is a Bernoulli system.*

For a proof of this deep theorem, see Ornstein (1974, pp. 45–47).

For example, consider the baker's system (M, Σ_M, μ, T) (Example 2), where d_M is the Euclidean metric. It is a 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. Let me explain this. For an arbitrary $\varepsilon > 0$ choose $n \in \mathbb{N}$ such that $\frac{\sqrt{2}}{2^n} < \varepsilon$. Consider the partition $\alpha_n=$

$$\{\alpha_1, \alpha_2, \ldots, \alpha_{2^{2n}}\} := \{[0, \frac{1}{2^n})\times[0, \frac{1}{2^n}), [0, \frac{1}{2^n})\times[\frac{1}{2^n}, \frac{2}{2^n}), \ldots, [\frac{2^n-1}{2^n}, 1]\times[\frac{2^n-1}{2^n}, 1]\}. \tag{7}$$

Now let $\Phi_{\alpha_n}(m) := \sum_{i=1}^{2^{2n}} o_{\alpha_i} \chi\alpha_i(m)$, where

$$o_{\alpha_1} = (\frac{1}{2^{n+1}}, \frac{1}{2^{n+1}}), o_{\alpha_2} = (\frac{1}{2^{n+1}}, \frac{3}{2^{n+1}}), \ldots, o_{\alpha_{2^{2n}}} = (\frac{2^{n+1}-1}{2^{n+1}}, \frac{2^{n+1}-1}{2^{n+1}}). \tag{8}$$

It is not hard to see that $\{\Phi_{\alpha_n}(T^t);\ t \in \mathbb{Z}\}$ is an irreducible and aperiodic Markov process whose deterministic representation is isomorphic to (M, Σ, μ, T). (M, Σ, μ, T) is ε-congruent to this Markov process since

$$d_M(m, \Phi_{\alpha_n}(m)) \leq \frac{\sqrt{2}}{2^n} < \varepsilon \ \text{ for all } m \in M. \tag{9}$$

Recall that irreducible and aperiodic Markov processes are widely used in science, and they are even regarded as being second most random. Also recall that several deterministic systems in science are Bernoulli systems (subsection 4.1). Hence Theorem 1 and Theorem 2 show that irreducible and aperiodic Markov processes are the most random stochastic processes which are needed in order to replace deterministic systems at every observation level. This implies that the conjecture advanced at the beginning of this subsection is wrong: *it is possible to replace measure-theoretic deterministic systems in science at every observation level by stochastic processes in science.*

4.2.2 Previous philosophical discussion

Let me discuss the previous philosophical papers about the topic of this section. Suppes and de Barros (1996) and Suppes (1999) discuss an instance of Theorem 2, namely that for discrete versions of billiards with convex obstacles and for any $\varepsilon > 0$ 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 section) and Winnie (1998) discuss the theorem that some continuous-time deterministic systems can be replaced at every observation level by semi-Markov processes.

Suppes and de Barros (1996, p. 196), Winnie (1998, p. 317) and Suppes (1999, p. 181–182) claim that the philosophical significance of these results is that for chaotic motion and every observation level one can choose between a deterministic description in science and a stochastic description. For instance, Suppes and 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 the ε-congruence results show more. As argued in section 3.1, the basic results on observational equivalence already show that for many deterministic systems, including many deterministic systems in science, the following holds: for any finite-valued observation function one can choose between a nontrivial stochastic or a deterministic description. This implies that, in a way, many deterministic systems can be replaced at every observation level by nontrivial stochastic processes. And as one would expect, given a deterministic system in science which satisfies the assumption of Proposition 1, for every $\varepsilon > 0$ there is a nontrivial stochastic process which is ε-congruent to the system. For basically all ergodic deterministic systems in science have a generating partition (Definition 14).[17] Besides, one easily sees the following: assume that (M, Σ_M, μ, T) is an ergodic measure-preserving deterministic system with a generating partition, and that (M, d_M) is separable and Σ_M contains all

[17]Basically all deterministic systems in science have finite Kolmogorov-Sinai entropy; and ergodic systems with finite Kolmogorov-Sinai entropy have a generating partition (cf. Petersen, 1983, p. 244; Ornstein and Weiss, 1991, p. 19).

open balls of (M, d_M). Then for every $\varepsilon > 0$ there is a stochastic process $\{\Phi(T^t);\, t \in \mathbb{Z}\}$, where $\Phi : M \to M$ is finite-valued, which is ε-congruent to the system.[18] And similar results for chaotic systems were known long before the ε-congruence results were proved (cf. subsection 3.1). *Hence the fact that at every observation level one has a choice between a deterministic description in science and a stochastic process was known long before the ε-congruence results were proved, and so cannot be the philosophical significance of these results as claimed by these authors.*[19] As I have argued in subsection 4.2.1, the significance of the ε-congruence results is something stronger: namely that it is possible to replace deterministic systems in science at every observation level by stochastic processes in science.

Furthermore, Suppes and de Barros (1996), Winnie (1998) and Suppes (1999) do not seem to be aware that *also for non-chaotic systems there is a choice between a deterministic and a stochastic description.* To show this, I do not have to discuss the hard question of how to define chaos. It will suffice to show that Proposition 1 also applies to systems which are uncontroversially neither chaotic nor locally chaotic. 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, μ is the Lebesgue measure, and T is the rotation $T(m) := m + \alpha \pmod 1$, where $\alpha \in \mathbb{R}$ is irrational. It is uncontroversial that this system is neither chaotic nor locally chaotic because all solutions are stable, i.e. nearby solutions stay

[18] 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 $\mu(M \setminus \cup_{i=1}^{r} B(m_i, \frac{\varepsilon}{2})) < \frac{\varepsilon}{2}$ ($B(m, \varepsilon)$ is the ball of radius ε around m). Let α be a generating partition. Then 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 $\Phi := \sum_{i=1}^{l} o_i \chi_{\beta_i}$ with $o_i \in \beta_i$. Φ is finite-valued and, since β is generating, (M, Σ_M, μ, T) is isomorphic via ϕ to the deterministic representation $(M_2, \Sigma_{M_2}, \mu_2, T_2, \Phi_0)$ of the stochastic process $Z_t := \{\Phi(T^t); t \in \mathbb{Z}\}$ (Petersen, 1983, p. 274). And, by construction, $d_M(x, \Phi_0(\phi(x))) < \varepsilon$ except for a set in M smaller than ε.

[19] The reader should also be warned that there are some technical lacunae in Suppes and de Barros (1996) and Suppes (1999). For instance, according to their definition, any two 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 continuous-time ε-congruence results require the motion to be a Bernoulli flow and so do not generally hold for ergodic systems. And in these papers it is wrongly assumed that the notions of isomorphism and ε-congruence require that the deterministic system is looked at through an observation function (Suppes and de Barros, 1996, p. 195–196, p. 200; p. 198-200; Suppes, 1999, p. 192, p. 195; pp. 189–192).

close for all times. However, one easily sees that it satisfies the assumption of Proposition 1.[20] Consequently, this deterministic system is replaced at every observation level by a nontrivial stochastic process.[21]

There remains the question: if one can choose between a deterministic system and a stochastic process, which description is preferable? Winnie (1998, pp. 317–318) dismisses Suppes's (1993, p. 254) claim that in the case of the ε-congruence results both descriptions are equally good. Winnie argues that the deterministic description is preferable: assume a stochastic process replaces a deterministic system for the current observation level. At some point in the future the observational accuracy may be so fine that another stochastic process will be needed to replace the deterministic system, and so on. Because there is in principle no limitation on the observational accuracy, there is no stochastic process that one can be sure, for practical purposes, will always give the same predictions as the deterministic system. Hence the deterministic description is preferable.

However, I think neither Winnie's (1998) nor Suppes' (1993) view is tenable. In a way, if the phenomenon under consideration is really stochastic, the stochastic description is preferable, even if the stochasticity is at a small scale and thus not observable. Likewise, if the phenomenon is really deterministic, the deterministic description is preferable. Now assume one cannot know for sure whether the phenomenon is deterministic or stochastic. Which description is then preferable in the sense of being preferable relative to our current knowledge and evidence? The answer depends on many factors, such as the kind of phenomenon under consideration, theories about fundamental physics, etc. And it may well be that the stochastic description is preferable if, for instance, a fundamental theory suggests this. To sum up, neither Winnie's nor Suppes' view is tenable, and the question of which description is preferable needs more careful examination.

[20]The system is ergodic (Petersen, 1983, p. 49). And there can be no $n \in \mathbb{N}$ and $C \in \Sigma_M,\ 0 < \mu(C) < 1$ such that $T^n(C) = C$ except for a set of measure zero since this would imply that all solutions are periodic.

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

5 Conclusion

The central question of this paper has been: are deterministic and indeterministic descriptions observationally equivalent in the sense that deterministic descriptions, when observed, and indeterministic descriptions give the same predictions? I have tackled it for discrete-time stochastic processes and measure-theoretic deterministic systems, both of which are ubiquitous in science.

I have demonstrated that every stochastic process is observationally equivalent to a deterministic system, and that many deterministic systems are observationally equivalent to stochastic processes. Still, one might guess that the measure-theoretic deterministic systems which are observationally equivalent to stochastic processes in science do not include any deterministic systems in science. I have shown this to be false because some deterministic systems in science even produce Bernoulli processes. Despite this, one might guess that measure-preserving deterministic systems in science cannot give the same predictions at every observation level as stochastic processes in science. I have shown that there is indeed a limitation on observational equivalence, namely deterministic systems in science cannot give the same predictions at every observation level as Bernoulli processes. However, the guess is still wrong because I have shown (one of the ε-congruence results) that several deterministic systems in science give the same predictions at every observation level as Markov processes. Hence stochastic processes and deterministic systems are more often observationally equivalent than one might perhaps have expected.

Furthermore, I have criticised the previous philosophical literature, namely Suppes and de Barros (1996), Winnie (1998) and Suppes (1999). They argue that the philosophical significance of the ε-congruence results 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 subsection 3.1. The philosophical significance of the ε-congruence result is really something stronger, namely, that there are deterministic systems in science that give the same predictions at every observation level as stochastic processes in science. Furthermore, these authors seem not to be aware that there are also uncontroversially non-chaotic deterministic systems which can be replaced at every observation level by stochastic processes.

6 Appendix: Proofs

6.1 Proof of Proposition 1

Proposition 1 *Assume that the deterministic system* (M, Σ_M, μ, T) *is ergodic and that 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 for every nontrivial finite-valued observation function* $\Phi : M \to M_O$ *and the stochastic process* $\{Z_t := \Phi(T^t);\ t \in \mathbb{Z}\}$ *the following holds: there is an* $o_i \in M_O$ *such that for all* $o_j \in M_O$, $P\{Z_{t+1} = o_j \mid Z_t = o_i\} < 1$.

Proof: I have not found a proof of this result in the literature and thus provide one here. Notice that it suffices to prove the following:

> $(*)$ Assume that (M, Σ_M, μ, T) is ergodic and 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, except for a set of measure zero, $T^n(C) = C$. Then for any nontrivial partition $\alpha = \{\alpha_1, \ldots, \alpha_n\}$ there is an $i \in \{1, \ldots, n\}$ such that for all j, $1 \leq j \leq n$, $\mu(T(\alpha_i) \setminus \alpha_j) > 0$.

For recall that any finite observation function has a corresponding partition (cf. subsection 3.1). Hence the conclusion of $(*)$ implies that for any nontrivial finite observation function $\Phi : M \to M_O$ there is an outcome $o_i \in M_O := \cup_{k=1}^{n} o_k$, $n \in \mathbb{N}$, such that for all possible outcomes $o_j \in M_O$ it follows that $P\{Z_{t+1} = o_j \mid Z_t = o_i\} < 1$, $t \in \mathbb{Z}$.

So assume that the conclusion of $(*)$ does not hold, i.e. there exists a nontrivial partition α such that for each α_i there exists an α_j with, except for a set of measure zero, $T(\alpha_i) \subseteq \alpha_j$.

Case 1: for all i there is a j such that, except for a set of measure zero, $T(\alpha_i) = \alpha_j$. Then ergodicity implies that α_1 is mapped, except for a set of measure zero, onto all α_k, $2 \leq k \leq n$, before being mapped onto itself. 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, except for a set of measure zero, $T^n(C) = C$.

Case 2: for some i there is a j with, except for a set of measure zero, $T(\alpha_i) \subset \alpha_j$ and $\mu(\alpha_i) < \mu(\alpha_j)$. Ergodicity implies that there exists a $k \in \mathbb{N}$ such that, except for a set of measure zero, $T^k(\alpha_j) \subseteq \alpha_i$. Hence it holds that $\mu(\alpha_j) \leq \mu(\alpha_i)$, yielding a contradiction, viz. $\mu(\alpha_i) < \mu(\alpha_j) \leq \mu(\alpha_i)$.

6.2 Proof of Theorem 1

Theorem 1 *Let (M, Σ_M, μ, T) be a 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 $d(T(x), D) := \inf\{d(T(x), m) \,|\, m \in D\} > 0$. Then there is some $\varepsilon > 0$ for which there is no Bernoulli process to which (M, Σ_M, μ, T) is ε-congruent.*

Proof: I have not found a proof of this result in the literature and thus provide one here.

For $m \in M$, $E \subseteq M$ and $\varepsilon > 0$ let the ball of radius ε around m be $B(m, \varepsilon) := \{y \in M \,|\, d(y, m) < \varepsilon\}$ 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 Bernoulli system. Let $\alpha_{\Phi_0} := \{\alpha^1_{\Phi_0} \ldots \alpha^s_{\Phi_0}\}$, $s \in \mathbb{N}$, be the partition of $(\Omega, \Sigma_\Omega, \nu)$ corresponding to the observation function Φ_0 (cf. subsection 3.1). Let $\check{M} := M \setminus \hat{M}$ and $\check{\Omega} := \Omega \setminus \hat{\Omega}$. Clearly, $\phi^{-1}(\alpha_{\Phi_0}) := \{\phi^{-1}(\alpha^1_{\Phi_0} \setminus \check{\Omega}) \cup \check{M}, \phi^{-1}(\alpha^2_{\Phi_0} \setminus \check{\Omega}), \ldots, \phi^{-1}(\alpha^s_{\Phi_0} \setminus \check{\Omega})\}$ 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 $A_1, \ldots A_n$, $n \in \mathbb{N}$, 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 $C_1, \ldots C_k$, $k \in \mathbb{N}$, 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 $(\Omega, \Sigma_\Omega, \nu, S, \Phi_0)$ is a Bernoulli process 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 $\varepsilon < \frac{\rho_1\rho_2}{8}$.

Remark 1. I say that a stochastic process $\{Z_t;\, t \in \mathbb{Z}\}$ is a *generalised Bernoulli process* if and only if (i) it takes values in an arbitrary measurable space $(\bar{M}, \Sigma_{\bar{M}})$; (ii) the random variables Z_t have the same distribution for all t; and (iii) $\{Z_t;\, t \in \mathbb{Z}\}$ are independent. A *generalised Bernoulli shift* is the deterministic representation of a generalised Bernoulli process. A *generalised Bernoulli system* is a deterministic system which is isomorphic to a generalised Bernoulli shift. Now it is important to note that Theorem 1 also holds for generalised Bernoulli processes, i.e.:

Theorem 1* *Let (M, Σ_M, μ, T) be a 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 $d(T(x), D) := \inf\{d(T(x), m) \,|\, m \in D\} > 0$. Then it is not the case that for all $\varepsilon > 0$ there is a generalised Bernoulli process such that (M, Σ_M, μ, T) is ε-congruent to this generalised Bernoulli process.*

Proof: the proof goes through as above when one considers generalised Bernoulli processes instead of Bernoulli processes, generalised Bernoulli shifts instead of Bernoulli shifts and generalised Bernoulli systems instead of Bernoulli systems, and one defines $A := \phi^{-1}(\Phi_0^{-1}(B(x, 3\delta)) \backslash \check{\Omega})$ and $C := \phi^{-1}(\Phi_0^{-1}(B(D, \beta)) \backslash \check{\Omega})$. Clearly, because for a generalised Bernoulli process the random variables are independent, it still holds that $\mu(T(A) \cap C) = \mu(A)\mu(C)$ and the proof goes through as above.

6.3 Proof of Theorem 2

Theorem 2 *Let (M, Σ_M, μ, T) be a 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 irreducible and aperiodic Markov process such that (M, Σ_M, μ, T) is ε-congruent to this Markov process.*

Proof: I have not found a proof of this result in the literature and thus provide one here. I need the following definition.

Definition 14 *A partition α of (M, Σ_M, μ, T) is* generating *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).*

Per assumption, the deterministic system (M, Σ_M, μ, T) is isomorphic via $\phi : \hat{M} \to \hat{\Omega}$ to the deterministic representation $(\Omega, \Sigma_\Omega, \nu, S, \Phi_0)$ of a Bernoulli shift with outcome space $\bar{M}$. Let $\alpha_{\Phi_0} := \{\alpha^1_{\Phi_0} \ldots \alpha^k_{\Phi_0}\}$, $k \in \mathbb{N}$, be the partition of $(\Omega, \Sigma_\Omega, \nu)$ corresponding to the observation function Φ_0 (cf. subsection 3.1). Let $\check{M} := M \setminus \hat{M}$ and $\check{\Omega} := \Omega \setminus \hat{\Omega}$. $\phi^{-1}(\alpha_{\Phi_0}) := \{\phi^{-1}(\alpha^1_{\Phi_0} \setminus \check{\Omega}) \cup \check{M}, \phi^{-1}(\alpha^2_{\Phi_0} \setminus \check{\Omega}), \ldots, \phi^{-1}(\alpha^k_{\Phi_0} \setminus \check{\Omega})\}$ is a partition of (M, Σ_M, μ). For the partitions $\alpha = \{\alpha_1, \ldots, \alpha_n\}$ and $\beta = \{\beta_1, \ldots, \beta_n\}, n \in \mathbb{N}$, $\alpha \vee \beta$ is the partition $\bigcup_{1 \leq i,j \leq n} \alpha_i \cap \beta_j$. Clearly, if α is a partition of M, $T^{-t}\alpha := \{T^{-t}\alpha_1, \ldots, T^{-t}\alpha_n\}$, $t \in \mathbb{Z}$, are partitions.

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 Bernoulli system $\Phi^{-1}(\alpha_{\Phi_0})$ is generating (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

$$d_M(m, \Phi_0^Q(\phi(m))) < \varepsilon \text{ except for a set in } M \text{ of measure } < \varepsilon. \tag{10}$$

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

$$\frac{\nu(\{\omega \in \Omega \,|\, (\omega_{-n} \ldots \omega_n) = A \text{ and } (\omega_{-n+1} \ldots \omega_{n+1}) = B_1\})}{\nu(\{\omega \in \Omega \,|\, (\omega_{-n+1} \ldots \omega_{n+1}) = B_1\})} = \tag{11}$$
$$\frac{\nu(\{\omega \in \Omega | (\omega_{-n} \ldots \omega_n) = A \text{ and} (\omega_{-n+1} \ldots \omega_{n+1}) = B_1, \ldots, (\omega_{-n+k} \ldots \omega_{n+k}) = B_k\})}{\nu(\{\omega \in \Omega | (\omega_{-n+1} \ldots \omega_{n+1}) = B_1, \ldots, (\omega_{-n+k} \ldots \omega_{n+k}) = B_k\})},$$

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 $\psi : \Omega \rightarrow X$, $\psi(\omega) = \ldots \Phi_0^Q(S^{-1}(\omega)), \Phi_0^Q(\omega), \Phi_0^Q(S(\omega)) \ldots$, for $\omega \in \Omega$. Clearly, there is a set $\hat{X} \subseteq X$ with $\lambda(\hat{X}) = 1$ such that $\psi : \Omega \rightarrow \hat{X}$ 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 (10) it holds that

$$d_M(m, \Theta_0(\theta(m))) < \varepsilon \text{ except for a set in } M \text{ of measure } < \varepsilon. \tag{12}$$

Acknowledgments

I am indebted to Jeremy Butterfield for his continued support and, in particular, for getting up at 3 a.m. to read my papers. For valuable comments I also want to thank Cymra Haskell, Franz Huber, Thomas Müller, Donald Ornstein, Amy Radunskaya, Peter Smith, Jos Uffink, two anonymous referees, and the audiences at the Oxford Philosophy of Physics Discussion Group, the Cambridge Moral Sciences Club, and the Wittgenstein Symposium 2008. I am grateful to St John's College Cambridge for financial support.

References

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

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

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

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

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

Doob, J.L. (1953). *Stochastic Processes*. New York: John Wiley & Sons.

Eckmann, J.-P. and Ruelle, D. (1985). Ergodic theory of chaos and strange attractors. *Reviews of Modern Physics*, 57, 617–654.

Halmos, P.R. (1944). In general a measure-preserving transformation is mixing. *The Annals of Mathematics*, 45, 786–792.

Jacobson, M.V. (1981). Absolutely continuous invariant measures for one-parameter families of one-dimensional maps. *Communications in Mathematical Physics*, 81, 39–88.

Lavis, D.A. (forthcoming). An objectivist account of probabilities in statistical physics. C. Beisbart and S. Hartmann (Eds.), *Probabilities in Physics*. Oxford: Oxford University Press.

Lorenz, E.G. (1964). The problem of deducing the climate from the governing equations. *Tellus*, XVI (1), 1–11.

Luzzatto, S., Melbourne, I. and Paccaut, F. (2002). 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*. Berlin: Springer.

May, R.M. (1976). Simple mathematical models with very complicated dynamics. *Nature*, 261, 459-67.

Norton, J.D. (2003). Causation as folk science. *Philosopher's Imprint*, 3, 1–22.

Ornstein, D. (1974). *Ergodic Theory, Randomness, and Dynamical Systems*. New Haven and London: Yale University Press.

Ornstein, D. (1989). Ergodic theory, randomness and "chaos". *Science*, 243, 182–187.

Ornstein, D. and Weiss, B. (1991). Statistical properties of chaotic systems. *Bulletin of the American Mathematical Society*, 24, 11–116.

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

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.

Sinai, Y.G. (1989). Kolmogorov's work on ergodic theory. *The Annals of Probability*, 17, 833–839.

Smith, P. (1998). *Explaining Chaos*. Cambridge: Cambridge University Press.

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. and de Barros, A. (1996). Photons, billiards and chaos. In P. Weingartner and G. Schurz (Eds.), *Law and Prediction in the Light of Chaos Research* (pp. 189–201). Berlin: Springer.

Weingartner, P. and Schurz, G. (1996). *Law and Prediction in the Light of Chaos Research*. Berlin: Springer.

Werndl, C. (2009a). What are the new implications of chaos for unpredictability? *The British Journal for the Philosophy of Science*, 60, 195–220.

Werndl, C. (2009b). Justifying definitions in mathematics—going beyond Lakatos. *Philosophia Mathematica*, doi: 10.1093/philmat/nkp006.

Werndl, C. (2009c). *Philosophical Aspects of Chaos: Definitions in Mathematics, Unpredictability and the Observational Equivalence of Deterministic and Indeterministic Descriptions.* Ph.D. thesis. Cambridge University.

Winnie, J.A. (1998). Deterministic chaos and the nature of chance. In J. Earman and J. Norton (Eds.), *The Cosmos of Science – Essays of Exploration* (pp. 299–324). Pittsburgh: Pittsburgh University Press.

Young, L.-S. (1997). Ergodic theory of chaotic dynamical systems. *XII-th International Congress of Mathematical Physics, Brisbane* (pp. 311–319). Cambridge, MA: International Press.