[Charlotte Werndl](#)

On the observational equivalence of continuous-time deterministic and indeterministic descriptions

Article (Accepted version)
(Refereed)

On the Observational Equivalence of Continuous-Time Deterministic and Indeterministic Descriptions

Charlotte Werndl; c.s.werndl@lse.ac.uk

Department of Philosophy, Logic and Scientific Method, London School of Economics

Abstract. This paper presents and philosophically assesses three types of results on the observational equivalence of continuous-time measure-theoretic deterministic and indeterministic descriptions. The first results establish observational equivalence to abstract mathematical descriptions. The second results are stronger because they show observational equivalence between deterministic and indeterministic descriptions found in science. Here I also discuss Kolmogorov's contribution. For the third results I introduce two new meanings of 'observational equivalence at every observation level'. Then I show the even stronger result of observational equivalence at every (and not just some) observation level between deterministic and indeterministic descriptions found in science. These results imply the following. Suppose one wants to find out whether a phenomenon is best modeled as deterministic or indeterministic. Then one cannot appeal to differences in the probability distributions of deterministic and indeterministic descriptions found in science to argue that one of the descriptions is preferable because there is no such difference. Finally, I criticise the extant claims of philosophers and mathematicians on observational equivalence.

Contents

1 Introduction

Determinism and indeterminism and whether one can know that phenomena are governed by deterministic or indeterministic laws are crucial philosophical themes. Hence it is an important question whether there is observational equivalence between deterministic and indeterministic descriptions. However, this question has hardly been discussed. This paper contributes to filling this gap by *presenting and philosophically assessing three types of results of increasing strength on the observational equivalence of deterministic and indeterministic descriptions*. When saying that a deterministic and an indeterministic description are observationally equivalent, I mean that the deterministic description, when observed, gives the same predictions as the indeterministic description. The deterministic and indeterministic descriptions of concern are continuous-time measure-theoretic deterministic systems and stochastic processes.

More specifically, the first results are about observational equivalence to abstract mathematical descriptions. I present a method of constructing, given a deterministic system, an observationally equivalent stochastic process, and conversely. Compared to the first results, the second results are stronger because they show not only observational equivalence to *abstract mathematical descriptions* but observational equivalence between deterministic and stochastic descriptions *of the types antecedently found in science*. Compared to the second results, the third results are even stronger because they show that there is observational equivalence *at every (not just at some) observation level* between deterministic and stochastic descriptions found in science. Notice that the increase in strength is very different for the move from the first to the second results (observational equivalence between descriptions found in science and not just to abstract mathematical descriptions) and for the move from the second to the third results (observational equivalence between descriptions in science at every, and not just at some, observation level).

I argue that a philosophical consequence of these results is the following. Suppose one wants to find out whether a phenomenon is best modeled as deterministic or stochastic. Then one might think of arguing that there is evidence for a deterministic or stochastic description by appealing to a general difference between the probability distributions of stochastic processes found in science and the probability distributions of (possibly fine-enough) observations of deterministic systems found in science. The second and third results of this paper show that these arguments are untenable because there is no such general difference between the respective probability distributions.

Finally, I criticise the previous philosophical discussion. The main philosophy literature on this topic is Suppes and de Barros (1996), Suppes (1999) and Winnie (1998). They claim that the significance of the third results is to provide a choice between deterministic and stochastic descriptions. I argue that while there is indeed such a choice, this is already shown by more basic results, and the third results show something stronger. Also, I criticise the claims of the mathematicians Ornstein and Weiss (1991) on observational equivalence.

As mentioned above, this paper is about *continuous-time* measure-theoretic deterministic systems and stochastic processes (here the time parameter varies continually). There are also *discrete-time* measure-theoretic deterministic systems and stochastic processes (here the time parameter varies in discrete steps). Werndl (2009a) discusses results about discrete-time descriptions which answer the question (for discrete time) of whether deterministic and indeterministic descriptions can be observationally equivalent. One of the

contributions of my paper is to answer this question for continuous-time deterministic systems and stochastic processes. This is important because, first, the discrete-time results leave open the answers for continuous time, and the answers for continuous time do not follow automatically from the discrete-time results. Indeed, as I will explain, the results differ for continuous-time and discrete-time descriptions; also, the proofs for continuous-time descriptions are harder and involve different techniques. Second, continuous-time descriptions are more widespread in science than discrete-time descriptions. Hence the issue of observational equivalence is more pressing for continuous-time descriptions.

Furthermore, there are several novel contributions of this paper which are not discussed in Werndl (2009a) or the other extant literature. In particular, I show that there are results on observational equivalence which increase in strength in a certain sense (namely, the focus on descriptions found in science and on every observation level). For this I introduce two new notions of observational equivalence at every observation level, I derive results for these notions, and I discuss Kolmogorov's contribution to observational equivalence. Moreover, I assess the significance of these results and point at their philosophical consequences. Also, the focus on continuous-time enables me to criticise the philosophical reflections of mathematicians, namely Ornstein and Weiss (1991). And I criticise in detail the claims of philosophers such as Suppes and de Barros (1996), Suppes (1999) and Winnie (1998) on observational equivalence.

This paper proceeds as follows. In Section 2 I introduce continuous-time measure-theoretic deterministic descriptions and stochastic processes. In Section 3 I present and assess the first results, in Section 4 the second results, and in Section 5 the third results on observational equivalence. Finally, in Section 6 I criticise the previous philosophical discussion on observational equivalence by philosophers and mathematicians.

2 Continuous-Time Deterministic Systems and Stochastic Processes

For what follows, I need to introduce a few basic measure-theoretic notions. Intuitively speaking, a probability space (M, Σ_M, μ) consists of a set M, a set Σ_M of subsets of M to which a probability is assigned, called a sigma-algebra of M, and a probability measure μ which assigns a probability to subsets of M. Formally, (M, Σ_M, μ) is a *probability space* if, and only if (iff), M is a set; Σ_M is a *sigma-algebra* of M, i.e., a set of subsets of M with (i) $\emptyset \in \Sigma_M$, (ii) for all $A \in \Sigma_M$, $M \setminus A \in \Sigma_M$, and (iii) for any $A_n \in \Sigma_M$, $n \geq 1$, $\bigcup_n A_n \in \Sigma_M$; and μ is a *probability measure*, i.e., a function $\mu : \Sigma_M \to [0, 1]$ with (i) $\mu(A) \geq 0$ for all $A \in \Sigma_M$, (ii) $\mu(\emptyset) = 0$ and $\mu(M) = 1$, and (iii) $\sum_{n \geq 1} \mu(A_n) = \mu(\bigcup_{n \geq 1} A_n)$ for any $A_n \in \Sigma_M$. A pair (M, Σ_M) where Σ_M is a sigma-algebra of M is called a *measurable space*. Finally, mathematically treatable functions in measure theory are called measurable functions; the functions encountered in science are all measurable. Technically, a function $T : M \to N$, where (M, Σ_M) and (N, Σ_N) are measurable spaces, is *measurable* iff $T^{-1}(A) \in \Sigma_M$ for all $A \in \Sigma_N$.[1]

---

[1] For more details on measure theory, see Cornfeld et al. (1982), Doob (1953) and Petersen (1983).

2.1 *Deterministic Systems*

This paper is about continuous-time measure-theoretic deterministic systems, in short deterministic systems, which are widespread in science. The three main elements of a deterministic system are the *phase space*, i.e., a set M of all possible states, a probability measure μ which assigns a probability to regions of phase space[2], and the *evolution functions*, i.e., a family of functions $T_t : M \to M$ where $T_t(m)$ represents the state of the system after t time units that started in initial state m. Formally:

Definition 1 *A* deterministic system *is a quadruple* (M, Σ_M, μ, T_t) *where* (M, Σ_M, μ) *is a probability space (M is the phase space) and* $T_t : M \to M$ *(the evolution functions), $t \in \mathbb{R}$, are measurable functions such that* $T_{t_1+t_2}(m) = T_{t_2}(T_{t_1}(m))$ *for all $m \in M$ and all $t_1, t_2 \in \mathbb{R}$, and* $T(t,m) := T_t(m)$ *is measurable in (t,m).*

A solution is a possible path of the deterministic system. Formally, the *solution* through m, $m \in M$, is the function $s_m : \mathbb{R} \to M$, $s_m(t) = T(t,m)$. All deterministic systems are *deterministic* according to the canonical definition: any solutions that agree at one instant of time agree at all times (Butterfield 2005).

I will often deal with measure-preserving deterministic systems where, intuitively speaking, the probability of any region remains the same when the region is evolved.

Definition 2 *A* measure-preserving deterministic system *is a deterministic system* (M, Σ_M, μ, T_t) *where μ is invariant, i.e.,* $\mu(T_t(A)) = \mu(A)$ *for all $A \in \Sigma_M$ and $t \in \mathbb{R}$.*

Measure-preserving deterministic systems are important in all the sciences, especially in physics (cf. Eckmann and Ruelle 1985). All deterministic Hamiltonian systems and deterministic systems in statistical mechanics are measure-preserving, and their invariant probability measure is the normalised Lebesgue measure or a similar measure (Cornfeld et al. 1982, 4–10). A measure-preserving deterministic system is *volume-preserving* iff the normalised Lebesgue measure – the standard measure of the volume – is the invariant probability measure. A measure-preserving system which fails to be volume-preserving is *dissipative*. Dissipative systems can also often be modeled as measure-preserving systems, e.g., the Lorenz system, which was introduced by Lorenz (1963) to model weather phenomena (cf. Luzatto et al. 2005).

When observing a deterministic system (M, Σ_M, μ, T_t), one observes a value $\Phi(m)$ dependent on the actual state m. Thus, technically, observations are *observation functions*, i.e., measurable functions $\Phi : M \to M_O$ where (M_O, Σ_{M_O}) is a measurable space (cf. Ornstein and Weiss 1991, 16). An observation function Φ is *finite-valued* iff it takes only finitely many values $\{o_1, \ldots, o_k\}$, $k \in \mathbb{N}$, and every o_i can be "seen", i.e., $\mu(\{m \in M \mid \Phi(m) = o_i\}) > 0$ for all i, $1 \le i \le k$. A finite-valued observation function is *nontrivial* iff $k \ge 2$. In practice observations are finite-valued.

The following deterministic system will accompany us.

Example 1: A billiard system with convex obstacles.
A billiard system with convex obstacles is a system where a ball moves with constant

---

[2]The question of how to interpret this probability is not one of the main topics here. I just mention a popular interpretation. According to the *time-average interpretation*, the probability of a set A is the long-run fraction of the proportion of time the system spends in A (cf. Eckmann and Ruelle 1985; Werndl 2009b).
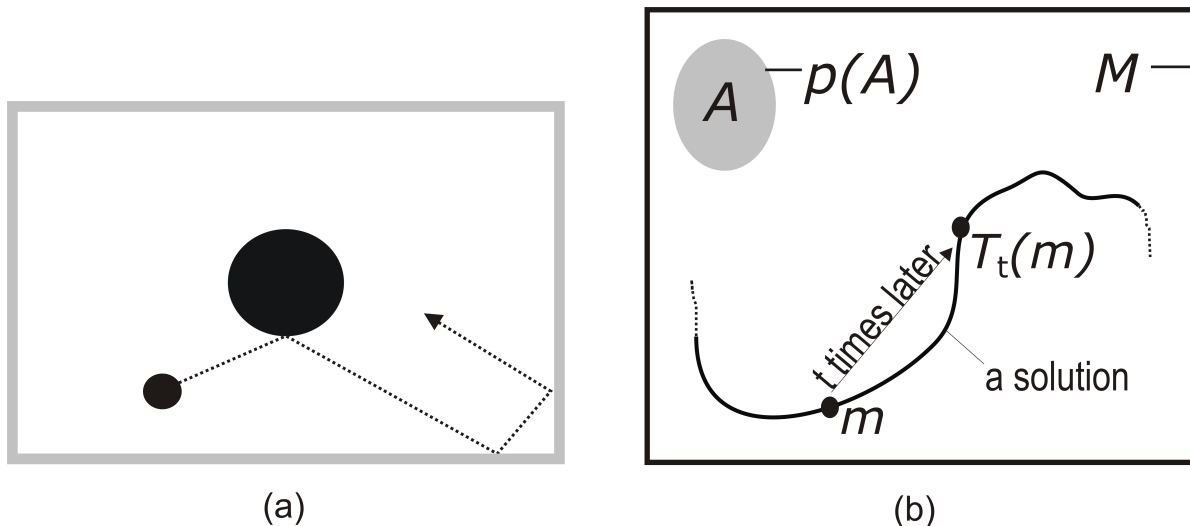
Figure 1: A billiard system with a convex obstacle; (a) a specific state of the billiard (b) the mathematical representation of the billiard

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 and Galavotti 1974). Figure 1(a) shows a specific state of the billiard system, where the dashed line indicates the path the ball will take. Figure 1(b) shows the mathematical representation of the system. M is the set of all possible positions and directions of the ball. Hence the state of the billiard shown in Figure 1(a) corresponds to exactly one point $m \in M$ shown in Figure 1(b). The normalised Lebesgue measure μ assigns the probability $\mu(A)$ to the event that the billiard is in one of the states represented by A for any region A in Σ_M (intuitively speaking, Σ_M is the set of all regions; formally, Σ_M is the Lebesgue σ-algebra). $T_t(m)$, where $m = (q, p)$, gives the position and the direction after t time units of the ball that starts out in initial position q and initial direction p. And a solution is a possible path of the billiard over time.

2.2 *Stochastic Processes*

The indeterministic descriptions of concern in this paper are continuous-time stochastic processes, in short stochastic processes. Stochastic processes are governed by probabilistic laws. They are the main indeterministic descriptions used in science.

A random variable Z gives the outcome of a probabilistic experiment, where the distribution $P\{Z \in A\}$ tells one the probability that the outcome will be in A. Formally, a *random variable* is a measurable function $Z : \Omega \to \bar{M}$ from a probability space $(\Omega, \Sigma_\Omega, \nu)$ to a measurable space $(\bar{M}, \Sigma_{\bar{M}})$. $P\{Z \in A\} = \nu(Z^{-1}(A))$ for all $A \in \Sigma_{\bar{M}}$ is the *distribution* of Z.

A stochastic process $\{Z_t; \ t \in \mathbb{R}\}$ consists of a probabilistic experiment (i.e., a random variable) for each time t. So $Z_t(\omega)$ gives the outcome of the process at time t (where ω represents a possible history in all its details). Formally:

Definition 3 *A stochastic process $\{Z_t; \ t \in \mathbb{R}\}$ is a 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*
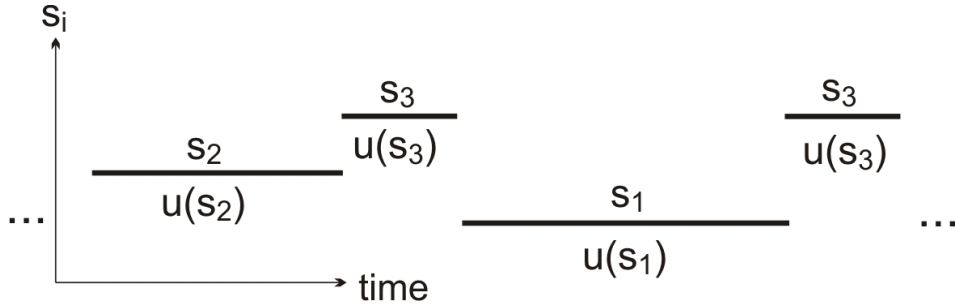
Figure 2: A realisation of a semi-Markov process

measurable space $(\bar{M}, \Sigma_{\bar{M}})$ such that $Z(t, \omega) = Z_t(\omega)$ is jointly measurable in (t, ω).

$\bar{M}$ is the set of possible outcomes of the process, called the *outcome space*. A *realisation* is a possible path of the stochastic process. Formally, it is function $r_\omega : \mathbb{R} \to \bar{M}$, $r_\omega(t) = Z(t, \omega)$, for $\omega \in \Omega$ arbitrary (cf. Doob 1953, 4–46).

I will often deal with stationary processes, the probability distributions of which do not change with time. $\{Z_t;\ t \in \mathbb{R}\}$ is *stationary* iff the distribution of $(Z_{t_1+h}, \ldots, Z_{t_n+h})$ is the same as the one of $(Z_{t_1}, \ldots, Z_{t_n})$ for all $t_1, \ldots, t_n \in \mathbb{R}$, $n \in \mathbb{N}$ and $h \in \mathbb{R}$.

When observing a stochastic process at time t, one observes a value $\Gamma(Z_t)$ dependent on the outcome Z_t. Hence, technically, observations are modeled by *observation functions*, i.e., measurable functions $\Gamma : \bar{M} \to \bar{M}_O$, where $(\bar{M}_O, \Sigma_{\bar{M}_O})$ is a measurable space. Clearly, observing the stochastic process $\{Z_t;\ t \in \mathbb{R}\}$ with Γ yields the process $\{\Gamma(Z_t);\ t \in \mathbb{R}\}$.

The following stochastic processes will accompany us.

Example 2: Semi-Markov processes.
A semi-Markov process has 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. Figure 2 shows a possible realisation of a semi-Markov process. For a formal definition, see Subsection 8.1. Semi-Markov processes are widespread in the sciences, from physics and biology to the social sciences. They are particularly important in queueing theory (cf. Janssen and Limnios 1999).

Example 3: n-step semi-Markov processes.
n-step semi-Markov processes generalise semi-Markov processes. An n-step semi-Markov process, $n \in \mathbb{N}$, has 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 the past n outcomes (hence semi-Markov processes are 1-step semi-Markov processes). For a formal definition, see Subsection 8.1. Again, n-step semi-Markov processes are widespread in science (cf. Janssen and Limnios 1999).

A final comment: the descriptions introduced in this section are classical. What results hold for quantum-mechanical descriptions and how similarities and differences in the results bear on the relation between classical and quantum mechanics are interesting questions; but they are beyond the scope of this paper.
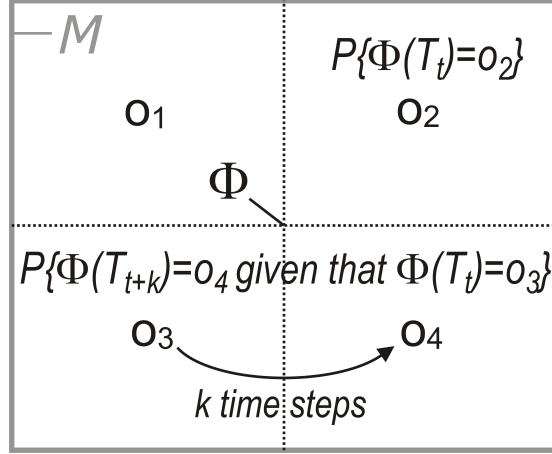
Figure 3: The billiard observed with Φ is observationally equivalent to $\{\Phi(T_t);\ t \in \mathbb{R}\}$.

3 Observational Equivalence: Results I

This section is about observational equivalence to mathematical descriptions. I show how, when starting with deterministic systems, one finds observationally equivalent stochastic processes (Subsection 3.1), and how, when starting with stochastic processes, one finds observationally equivalent deterministic systems (Subsection 3.2).

I speak of observational equivalence if *the deterministic system, when observed, and the stochastic process give the same predictions.* More specifically, deterministic systems are endowed with a probability measure. Hence when observing a deterministic system, the resulting predictions are the probability distributions over sequences of possible observations. The predictions obtained from a stochastic process are the probability distributions over the realisations. Consequently, a deterministic system, when observed, and a stochastic process give the same predictions iff (i) the possible outcomes of the process and the possible observed values of the system coincide, and (ii) the realisations of the process and the solutions of the system coarse-grained by the observation function have the same probability distribution.

3.1 *Starting With Deterministic Systems*

Assume (M, Σ_M, μ, T_t) is observed with $\Phi : M \to M_O$. Then $\{\Phi(T_t);\ t \in \mathbb{R}\}$ is a stochastic process, which arises by applying Φ to the deterministic system. Thus the possible outcomes of the process and the possible observed values of the deterministic system coincide, and the probability distributions over the realisations of the process and over the solutions of the system coarse-grained by the observation function are the same. Therefore, (M, Σ_M, μ, T_t) *observed with* Φ *is observationally equivalent to* $\{\Phi(T_t);\ t \in \mathbb{R}\}$.

Let me illustrate this with the billiard system with convex obstacles. Figure 3 shows the phase space M of the billiard system (cf. Figure 1(b)) and an observation function Φ with values o_1, o_2, o_3, o_4: for all states m in the top left box the observed value is o_1, etc. Because a probability measure μ is defined on M, one obtains probabilities such as $P\{\Phi(T_t) = o_2\}$ and $P\{\Phi(T_{t+k}) = o_4$ given that $\Phi(T_t) = o_3\}$. $\{\Phi(T_t);\ t \in \mathbb{R}\}$ has the outcomes o_1, o_2, o_3, o_4

and its probability distributions are obtained by observing the billiard system with Φ. Thus the billiard system observed with Φ is observationally equivalent to $\{\Phi(T_t); t \in \mathbb{R}\}$.

However, we want to know whether $\{\Phi(T_t); t \in \mathbb{R}\}$ is nontrivial. For let $\Phi(m) = m$ be the identity function. $\{\Phi(T_t(m)); t \in \mathbb{R}\} = \{T_t(m); t \in \mathbb{R}\}$ is formally a stochastic process which is observationally equivalent to the deterministic system. Yet it is just the original deterministic system. This result is the formalisation of the old idea that a deterministic system is the special case of a stochastic process where all probabilities are 0 or 1 (cf. Butterfield 2005). It illustrates that if one wants to arrive at nontrivial stochastic processes, one has to apply observation functions which are many-to-one.

Yet $\{\Phi(T_t); t \in \mathbb{R}\}$ *is often nontrivial.* Several results show this. To my knowledge, the following theorem has never been discussed before. This theorem characterises a class of deterministic systems as systems where, regardless which finite-valued observation function Φ is applied, $\{\Phi(T_t); t \in \mathbb{R}\}$ is nontrivial in the following sense: for any $k \in \mathbb{R}^+$ there are outcomes $o_i, o_j \in M_O$ such that the probability of moving in k time steps from o_i to o_j is between 0 and 1. This result is strong because there are nontrivial probability distributions for *any* finite-valued observation function and *all* time steps.

Theorem 1 *Iff for a 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*[3], $T_n(C) = C$, *then the following holds: for every nontrivial finite-valued observation function* $\Phi : M \to M_O$, *every* $k \in \mathbb{R}^+$ *and* $\{Z_t; t \in \mathbb{R}\} := \{\Phi(T_t); t \in \mathbb{R}\}$ *there are* $o_i, o_j \in M_O$ *with* $0 < P\{Z_{t+k} = o_j \mid Z_t = o_i\} < 1$.

For a proof, see Subsection 8.2.[4]

The assumption of Theorem 1 is equivalent to weak mixing (Hopf 1932). This indicates a difference between the results for discrete and continuous time. As shown in Werndl (2009a), for discrete time the condition needed to arrive at the results that, regardless which finite-valued observation function and time step is considered, one always obtains nontrivial probability distributions is *weaker* than weak mixing. In particular, Werndl (2009a, Section 4.2.2) shows that this result holds even for a class of *stable* deterministic systems (i.e., where solutions which start closely stay close all the time). This is not the case for continuous time because if a deterministic system is weak mixing, there is a sense in which solutions which start closely eventually separate (cf. Werndl 2009a). Yet Theorem 1 is just one of the possible theorems one can prove. Dependent on one's purpose, one can weaken (or strengthen) the conditions to prove weaker (or stronger[5]) results on observational equivalence. And weaker results than Theorem 1 show that one obtains observational equivalence to nontrivial stochastic processes also for stable deterministic systems and thus also for non-chaotic systems (*chaotic* systems show deterministic yet unstable behaviour). For instance, consider the deterministic system of a rotation on a circle, which is stable and non-chaotic; $M = [0, 1)$ represents the circle of unit radius, Σ_M is the Lebesgue σ-algebra, μ the Lebesgue measure and $T_t(m) = \alpha t m \pmod 1$, $\alpha \in \mathbb{R}$. It

---

[3]That is, except for Q with $\mu(Q) = 0$.

[4]If (M, Σ_M, μ, T_t) is measure-preserving, $\{\Phi(T_t); t \in \mathbb{R}\}$ is stationary (the proof in Werndl 2009a, Section 3.3, carries over to continuous time).

[5]For instance, a stronger result is that for Kolmogorov systems and any finite-valued observation function the following holds: even if you know the entire infinite history of the process, you will not be able to predict with certainty which outcome follows next (Uffink 2006, 1012–1014).

is not hard to see that for any finite-valued observation function Φ the stochastic process $\{Z_t; t \in \mathbb{R}\} = \{\Phi(T_t); t \in \mathbb{R}\}$ is nontrivial in the sense that for almost all (but not all) time steps there are nontrivial probability distributions. Technically: for any finite-valued Φ and all, except for a set of measure zero, $k \in \mathbb{R}^+$, there are $o_i, o_j \in M_O$ with $0 < P\{Z_{t+k} = o_j \mid Z_t = o_i\} < 1$, where $\{Z_t\} = \{\Phi(T_t); t \in \mathbb{R}\}$.

Measure-preserving systems are typically weakly mixing (Halmos 1944, 1949). Consequently, Theorem 1 typically applies to measure-preserving deterministic systems. Yet this does not tell us whether Theorem 1 applies to the deterministic systems of physical relevance; systems of physical relevance are only a small class of all measure-preserving systems. But Theorem 1 applies to several physically relevant deterministic systems. It applies to systems in Newtonian mechanics: for instance, first, to billiard systems with convex obstacles (Example 1) (Ornstein and Galavotti 1974); second, to many hard-ball systems, which describe the motion of a number of hard balls undergoing elastic reflections at the boundary and collisions amongst each other and are important in statistical mechanics because they model gases; e.g., to two hard balls in a box and the motion of N hard balls on a torus for almost all values $(m_1, \ldots, m_N, r)$, where m_i is the mass of the i-th ball and r is the radius of the balls, $N \geq 2$ (Berkovitz et al. 2006, 679–680; Simányi 2003); third, to geodesic flows of negative curvature, i.e., frictionless motion of a particle moving with unit speed on a compact manifold with everywhere negative curvature (Ornstein and Weiss 1991, Section 4). Also, Theorem 1 applies to dissipative systems, such as Lorenz-type systems, which model weather dynamics and waterwheels (Lorenz 1963; Luzzatto et al. 2005; Strogatz 1994). Because proofs are extremely hard, it is often only conjectured that deterministic systems satisfy Theorem 1, e.g., for any number of hard balls in a box (Berkovitz et al. 2006, 679–680). Note that the deterministic systems listed in this paragraph are chaotic.

There are also many systems to which Theorem 1 does not apply. For instance, according to the KAM theorem, the phase space of integrable Hamiltonian systems which are perturbed by a small non-integrable perturbation breaks up into stable regions and regions with unstable behavior. Because of the break up into regions, Theorem 1 does not apply to these systems (cf. Berkovitz et al. 2006, Section 4). But even if Theorem 1 does not apply to the whole deterministic system, it might well apply to the motion restricted to some regions A of phase space (and this is conjectured for KAM-type systems). Then it follows from Theorem 1 that all finite-valued observations which discriminate between values in A lead to nontrivial stochastic processes.

3.2   *Starting With Stochastic Processes*

The following idea of how, given stochastic processes, one finds observationally equivalent deterministic systems will be needed for what follows. It is well known in mathematics and in philosophy of physics (Butterfield 2005; Doob 1953, 6–7; for discrete time it is discussed in Werndl 2009a, Section 3.2).

The underlying idea is that one constructs a deterministic system whose phase space is the set of all possible realisations $m(\tau)$, whose evolution functions T_t shift the realisation t time steps to the left, and whose observation function Φ_0 gives the value of the realisation at time 0. Because T_t shifts the realisation $m(\tau)$ t time steps to the left, observing $T_t(m(\tau))$ with Φ_0 gives the value of the realisation $m(\tau)$ at time t. Hence successive observations

return the values of the realisation $m(\tau)$ at different times and so simply yield the outcomes of the stochastic process.

Formally, for $\{Z_t; t \in \mathbb{R}\}$ with outcome space $\bar{M}$, let M be the set of all possible realisations, i.e., functions $m(\tau)$ from $\mathbb{R}$ to $\bar{M}$. Let Σ_M be the σ-algebra generated[6] by the cylinder-sets

$$C_{i_1 \ldots i_n}^{A_1 \ldots A_n} = \{m \in M \mid m(i_1) \in A_1, \ldots, m(i_n) \in A_n, A_j \in \Sigma_{\bar{M}}, i_j \in \mathbb{R}, i_1 < \ldots < i_n, 1 \le j \le n\}. \quad (1)$$

Let μ be the measure on M determined by the probability distributions of $\{Z_t; t \in \mathbb{R}\}$. That is, μ is the unique probability measure arising by assigning to each $C_{i_1 \ldots i_n}^{A_1 \ldots A_n}$ the probability $P\{Z_{i_1} \in A_1, \ldots, Z_{i_n} \in A_n\}$. The evolution functions shift a realisation t times to the left, i.e., $T_t(m(\tau)) = m(\tau + t)$. (M, Σ_M, μ, T_t) is a deterministic system called the *deterministic representation* of $\{Z_t; t \in \mathbb{R}\}$ (cf. Doob 1953, 621–622). Finally, assume that the observation function gives the value of the realisation at time zero, i.e., $\Phi_0(m(\tau)) = m(0)$.

Because one only observes the 0-th coordinate, the possible outcomes of a stochastic process and the possible observed values of its deterministic representation coincide. Moreover, because μ is defined by the stochastic process, the probability distribution over the realisations is the same as the one over the sequences of observed values of its deterministic representation. Therefore, *a stochastic process is observationally equivalent to its deterministic representation observed with Φ_0. Consequently, for every process there is at least one observationally equivalent deterministic system.*[7] For instance, consider a semi-Markov process $\{Z_t; t \in \mathbb{R}\}$ (Example 2) and its deterministic representation (M, Σ_M, μ, T_t). Here M is the set of all possible realisations of the semi-Markov process, μ gives the probability distribution over the realisations, T_t shifts a realisation t time steps to the left, and Φ_0 returns the value of the realisation at time 0.

Philosophically speaking, the deterministic representation is a cheat. For it resembles the following argument: Fred was fatally wounded at noon; therefore he died later. The predicate 'being fatally wounded a noon' implies that Fred will die later. Hence the argument is trivially valid. Similarly, the states of the deterministic representation are constructed such that they *encode* the future and past evolution of the process. So there is the question whether deterministic systems which do not involve a cheat can be observationally equivalent to a given stochastic process. It will become clear later that for some stochastic processes the answer is positive. It is unknown whether the answer is positive for every stochastic process. For a formalisation of the notion of observational equivalence underlying Section 3, see Werndl 2009a (the definition carries over to continuous-time).

4   Observational Equivalence: Results II

The previous section only showed that in certain cases there is observational equivalence to abstract mathematical descriptions. Furthermore, the deterministic representation – one of these abstract mathematical descriptions – is not used by scientists. This raises the question of whether there can be observational equivalence not only to *abstract mathematical*

---

[6]The σ-algebra generated by E is the smallest σ-algebra containing E; that is, the σ-algebra $\bigcap_{\sigma\text{-algebras } \Sigma, \, E \subseteq \Sigma} \Sigma$.

[7]The deterministic representation of any stationary process is measure-preserving (the proof in Werndl 2009a, Section 3.3, carries over to continuous time).

descriptions but whether a *stronger* result can hold: namely, that there is observational equivalence between stochastic processes and deterministic systems *of the types used in scientific theorising* (in short, *systems and processes in science*).

If the answer is negative, one could divide the probability distributions found in science into two groups: the ones deriving from observations of deterministic systems in science, and the ones deriving from stochastic processes in science and their observations.[8] Suppose one wants to find out whether a phenomenon is best described as deterministic or stochastic. Then one might argue the following: if the observed probability distributions are of the type of deterministic systems in science, the evidence speaks for a deterministic description. Conversely, if the probability distributions are of the type of stochastic processes in science, this provides evidence for a stochastic description. The idea here is that the probability distributions characteristically found for deterministic (or stochastic) descriptions in science provide inductive evidence for a deterministic (or stochastic) description. There is the question whether such an argument is convincing. Clearly, it can only work if there is indeed a division between the respective probability distributions. So let us ask: *is there a deterministic system in science and is there an observation function such that the deterministic system, when observed, yields a stochastic process in science?*

4.1 *Kolmogorov's Conjecture*

Kolmogorov found it hard to imagine a positive answer to this question. More specifically, it has been nearly forgotten (and has neither been subject to a systematic historical investigation nor been discussed philosophically) that Kolmogorov conjectured the following. *Suppose an observed deterministic system yields a stochastic process in science; then the observed deterministic system produces positive information. Contrary to this, any arbitrary observation of a deterministic system in science never produces positive information.* Employing ideas in information theory, the Kolmogorov-Sinai entropy (KS-entropy) was introduced to capture the highest amount of information, or equivalently the highest amount of uncertainty, that an observation of a deterministic system could produce; where a positive KS-entropy indicates that there are observations which produce positive information. Kolmogorov expected that the KS-entropy would accomplish the separation of deterministic systems in science from the deterministic systems producing stochastic processes in science: the former have zero entropy, the latter have positive entropy.

So it was a big surprise when, from the 1960s onwards, it was found that also many deterministic systems in science have positive KS-entropy and thus have observations which produce positive information. Namely, all the deterministic systems listed in Subsection 3.1 as systems of physical relevance to which Theorem 1 applies have positive KS-entropy; in particular, billiard systems with convex obstacles (Example 1), many hard-ball systems, geodesic flows of constant negative curvature and Lorenz-type systems. For many systems in science, such as for KAM-type systems, it is conjectured that they have positive KS-entropy. *Hence Kolmogorov's attempt to separate deterministic systems in science from deterministic systems which produce stochastic processes in science failed* (Radunskaya 1992,

---

[8]By 'probability distributions' I mean here nontrivial probability distributions (not all probabilities 0 or 1). Otherwise such a division could never exist. For consider a process composed of two semi-Markov processes (Example 2), and an observation function which tells one whether the outcome is among those of the first or the second process. Then the observed probability distribution is trivial.

chapter 1; Sinai 1989, 835–837).

4.2 Deterministic Systems in Science Which Are Observationally Equivalent to Stochastic Processes in Science

So is there a deterministic system in science and an observation such that the observed deterministic system yields a stochastic process in science? To answer this, I need to introduce the definitions of a continuous Bernoulli system and of isomorphism.

Intuitively speaking, Bernoulli systems are strongly chaotic.

Definition 4 *The measure-preserving deterministic system* (M, Σ_M, μ, T_t) *is a* continuous Bernoulli system *iff for all* $t \in \mathbb{R} \setminus \{0\}$ *the discrete measure-preserving deterministic system* (M, Σ_M, μ, T_t) *is a discrete Bernoulli system.*

For the discussion in the main text, the details of this definition are not important.[9] Here it just matters that several deterministic systems in science are continuous Bernoulli system; namely, all the systems listed in the previous subsection as systems with positive KS-entropy, in particular, billiard systems with convex obstacles (Example 1), many hard-ball systems, geodesic flows of constant negative curvature and Lorenz-type systems. Again, for several systems, e.g., for the motion on unstable regions of KAM-type systems, it is conjectured that they are continuous Bernoulli systems.

Isomorphic deterministic systems are probabilistically equivalent, i.e., their states can be put into one-to-one correspondence (via a function ϕ) such that the corresponding solutions have the same probability distributions.

Definition 5 *The 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* isomorphic *iff there are* $\hat{M}_i \subseteq M_i$ *with* $\mu_i(M_i \setminus \hat{M}_i) = 0$ *and* $T_t^i \hat{M}_i \subseteq \hat{M}_i$ *for all* t $(i = 1, 2)$, *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)* $\phi(T_t^1(m)) = T_t^2(\phi(m))$ *for all* $m \in \hat{M}_1, t \in \mathbb{R}$.

Now assume that (M, Σ_M, μ, T_t) is isomorphic (via $\phi : \hat{M} \to \hat{M}_2$) to the deterministic representation $(M_2, \Sigma_{M_2}, \mu_2, T_t^2)$ of the stochastic process $\{Z_t; t \in \mathbb{R}\}$. This means that there is a one-to-one correspondence between the states of the system and the realisations of the process. Recall that observing $(M_2, \Sigma_{M_2}, \mu_2, T_t^2)$ with Φ_0 yields $\{Z_t; t \in \mathbb{R}\}$ (where $\Phi_0(m_2(\tau)) = m_2(0)$). Consequently, (M, Σ_M, μ, T_t) observed with the observation function $\Phi := \Phi_0(\phi(m))$ is observationally equivalent to $\{Z_t; t \in \mathbb{R}\}$[10]

---

[9]A discrete Bernoulli system is defined as follows. Recall Definition 12 of a discrete measure-preserving deterministic system. A doubly-infinite sequence of independent rolls of an N-sided die where the probability of obtaining s_k is p_k, $1 \leq k \leq N$, $\sum_{k=1}^{N} p_k = 1$, is a Bernoulli process. Let M be the set of all bi-infinite sequences $m = (\dots m(-1), m(0), m(1) \dots)$ with $m(i) \in \bar{M} = \{s_1, \dots, s_N\}$. Let Σ_M be the σ-algebra generated by the cylinder-sets $C_{i_1 \dots i_n}^{\{s_{l_1}\} \dots \{s_{l_n}\}}$ as defined in (1) with $i_j \in \mathbb{Z}, s_{l_j} \in \bar{M}$. These sets have probability $\bar{\mu}(C_{i_1 \dots i_n}^{\{s_{l_1}\} \dots \{s_{l_n}\}}) = p_{s_{l_1}} \dots p_{s_{l_n}}$. Let μ the unique measure determined by $\bar{\mu}$ on Σ_M. Let $T : M \to M$, $T((\dots m(i) \dots)) = (\dots m(i+1) \dots)$. (M, Σ_M, μ, T) is called a Bernoulli shift. Finally, a discrete measure-preserving deterministic system is a *Bernoulli system* iff it is isomorphic to some Bernoulli shift (isomorphism is exactly defined as for continuous time – see Definition 5).

[10]It does not matter how $\Phi_0(\phi(m))$ is defined for $m \in M \setminus \hat{M}$.

An important application of this principle is as follows. A deep result shows that up to a scaling of time t any continuous Bernoulli systems are isomorphic. That is, given continuous Bernoulli systems (M, Σ_M, μ, T_t) and $(M_2, \Sigma_{M_2}, \mu_2, T_t^2)$ there is a $c > 0$ such that (M, Σ_M, μ, T_t) and $(M_2, \Sigma_{M_2}, \mu_2, T_{ct}^2)$ are isomorphic (Ornstein 1974). Now recall semi-Markov processes (Example 2). Ornstein (1970) proved that the deterministic representation $(M_2, \Sigma_{M_2}, \mu_2, T_t^2)$ of any semi-Markov process is a continuous Bernoulli system. Clearly, for any $c \in \mathbb{R}^+$, $(M_2, \Sigma_{M_2}, \mu_2, T_{ct}^2)$ is still the deterministic representation of a semi-Markov process. Therefore, given any continuous Bernoulli system in science (M, Σ_M, μ, T_t), there is a semi-Markov process whose deterministic representation $(M_2, \Sigma_{M_2}, \mu_2, T_{ct}^2)$ is isomorphic (via ϕ) to (M, Σ_M, μ, T_t), implying that (M, Σ_M, μ, T_t) observed with $\Phi := \Phi_0(\phi(m))$ produces the semi-Markov process. For instance, billiards with convex obstacles (Example 1) are continuous Bernoulli systems. Hence there is an observation function Φ which, applied to the billiard system, yields a semi-Markov processes.

The converse is also true: given any semi-Markov process, there is a Bernoulli system in science which is observationally equivalent to the semi-Markov process. For given any Bernoulli system in science (M, Σ_M, μ, T_t), $(M, \Sigma_M, \mu, T_{ct})$ is also a system in science (e.g., for billiard systems this means that the constant speed of the ball is changed from s to cs). Thus given any semi-Markov process there is a deterministic system in science $(M, \Sigma_M, \mu, T_{ct})$ which is isomorphic to the deterministic representation $(M_2, \Sigma_{M_2}, \mu_2, T_t^2)$ of the semi-Markov process, implying that $(M, \Sigma_M, \mu, T_{ct})$ observed with $\Phi := \Phi_0(\phi(m))$ yields the semi-Markov process. For instance, given any semi-Markov process (Example 2), one can change the constant speed of a billiard system with convex obstacles such that an observation of the billiard system yields the semi-Markov process.

To sum up, deterministic systems in science (namely, certain Bernoulli systems) are observationally equivalent to semi-Markov processes (which are widespread in science).[11] For discrete time Werndl (2009a, Section 4) showed that some deterministic systems in science are observationally equivalent to Bernoulli processes or discrete-time Markov processes (both are widespread in science). Yet this left open whether continuous-time deterministic systems in science can yield stochastic processes in science. Indeed, Werndl's results on Bernoulli or Markov processes do not carry over to continuous time: Bernoulli processes are only defined for discrete time and there are no deterministic systems in science which yield continuous-time Markov processes. The latter holds because a deterministic system can only be observationally equivalent to a continuous-time Markov process if the KS-entropy of the Markov process is not higher than the KS-entropy of the system; but deterministic systems in science have finite KS-entropy, and continuous-time Markov processes have infinite KS-entropy (Feldman and Smorodinsky 1971; Ornstein and Weiss 1991, 19).

So I conclude that the answer to the question advanced at the beginning of this section is positive: *there are deterministic systems in science and observation functions such that the deterministic systems, when observed, yield stochastic processes in science.*[12] So there is observational equivalence not just to abstract mathematical descriptions (as shown in Section 3)[13] but a *stronger* form of observational equivalence between descriptions in science.

---

[11]n-step semi-Markov processes are continuous Bernoulli systems too (Ornstein 1974). Therefore, an analogous argument shows that Bernoulli systems in science are observationally equivalent to n-step semi-Markov processes.

[12]My arguments allow any meaning of 'deterministic systems deriving from scientific theories' that excludes the deterministic representation but is wide enough to include some continuous Bernoulli systems.

[13]Also, compared to the results in Section 3, the results here concern a narrower class of descriptions,

Thus any argument that a phenomenon is best described as deterministic or stochastic (such as the one sketched at the beginning of this section) cannot work when it relies on the premise that there is a division between the probability distributions of deterministic and stochastic descriptions in science.

5   Observational Equivalence: Results III

In the previous section we have seen that there are *some* observations of deterministic systems in science which yield stochastic processes in science. Now let us ask whether an even *stronger* result holds. Namely, *can deterministic systems in science and stochastic processes in science be observationally equivalent at every observation level* (and not just for some observation)? Note that, compared to the previous section, the type of systems considered remain the same, namely deterministic and stochastic descriptions in science.

   Intuitively, one might think that deterministic systems in science can yield stochastic processes in science *only* if *specific coarse* observation functions are applied. If this is true, fine-enough observations of deterministic systems in science would yield different probability distributions than stochastic processes in science. Then, if one wants to find out whether a phenomenon is best described as deterministic or stochastic, one might argue the following: if fine observations yield probability distributions characteristically found for deterministic systems in science, the evidence favours a deterministic description. The idea here is that the existence of characteristic probability distributions for deterministic systems found in science gives us inductive evidence that a phenomenon is best described as deterministic. Such an argument is only tenable if fine-enough observations of deterministic systems indeed yield special probability distributions, i.e., if the question whether deterministic systems in science can be observationally equivalent at every observation level to stochastic processes in science has a negative answer. So let us focus on this question.

5.1   *The Meaning of Observational Equivalence at Every Observation Level*

What does it mean that 'stochastic processes of a certain type are observationally equivalent to a deterministic system *at every observation level*'? I will first introduce the standard notion and then two new notions. I focus on measure-preserving deterministic systems and, correspondingly, stationary stochastic processes because all examples will be of this kind.

The usual meaning based on ε-congruence

To introduce the standard notion, I start by explaining what it means for a deterministic system and a stochastic process to give the same predictions at an observation level $\varepsilon > 0$, $\varepsilon \in \mathbb{R}$. For sufficiently small $\varepsilon_1 > 0$ one cannot distinguish states of the deterministic system which are less than the distance ε_1 apart. Also, for sufficiently small $\varepsilon_2 > 0$ one will not be able to distinguish differences in probabilities of less than ε_2. Assume that $\varepsilon < \min\{\varepsilon_1, \varepsilon_2\}$. Then a deterministic system and a stochastic process give the same

---

namely only continuous Bernoulli systems, semi-Markov processes and n-step semi-Markov processes.

predictions at level ε iff the following shadowing result holds: the solutions can be put into one-to-one correspondence with the realisations such that the state of the system and the outcome of the process are at all time points within distance ε except for a set of probability smaller than ε.

ε-congruence captures this idea mathematically (one assumes that a metric d_M measures distances between states and that the possible outcomes of the process are in M; also, recall the deterministic representation as discussed in Subsection 3.2 and Definition 5 of isomorphism):

Definition 6 *Let (M, Σ_M, μ, T_t) be a measure-preserving deterministic system, where (M, d_M) is a metric space. Let $(M_2, \Sigma_{M_2}, \mu_2, T_t^2)$ be the deterministic representation of the stochastic process $\{Z_t; t \in \mathbb{R}\}$ and let $\Phi_0 : M_2 \to M$, $\Phi_0(m_2(\tau)) = m_2(0)$. (M, Σ_M, μ, T_t) is ε-congruent to $\{Z_t; t \in \mathbb{R}\}$ iff (M, Σ_M, μ, T_t) is isomorphic via $\phi : M \to M_2$ to $(M_2, \Sigma_{M_2}, \mu_2, T_t^2)$ and $d_M(m, \Phi_0(\phi(m))) < \varepsilon$ for all $m \in M$ except for a set of measure $< \varepsilon$.*

Now we generalise over ε and arrive at a plausible meaning of the phrase that 'stochastic processes of a certain type are observationally equivalent at every observation level to a deterministic system'. Namely: *for every $\varepsilon > 0$ there is a process of this type which is ε-congruent to the system* (cf. Ornstein and Weiss 1991, 22–23; Suppes 1999).

The idea for new notions of observational equivalence at every observation level starts from the following thought. ε-congruence does *not* assume that the system is observed with an observation function: the *actual* states of the deterministic system, *not* the observed ones, 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 state of the system and the outcome of the process are less than ε apart. (I should mention that if (M, Σ_M, μ, T_t) and $\{Z_t; t \in \mathbb{R}\}$ are ε-congruent, then $\{\Psi(T_t); t \in \mathbb{R}\}$, where $\Psi(m) := \Phi_0(\phi(m))$, is the process $\{Z_t; t \in \mathbb{R}\}$. Technically, Ψ is an observation function but for ε-congruence it is *not* interpreted in this way. Instead, the meaning of Ψ is as follows: when it is applied to the system, the resulting process shadows the deterministic system.)

In many contexts, such as in Newtonian and statistical mechanics, observations are modeled by observation functions, and one would like to know what happens when specific observation functions are applied. Thus it would be desirable to have a notion of observational equivalence at every observation level based on observation functions. Results about other notions might be regarded as wanting because they leave unclear what happens when observation functions are applied. As explained, ε-congruence is not based on observation functions. For this reason, I now introduce two other notions of observational equivalence at every observation level. Whether a notion is preferable that (i) tells one what happens when one cannot distinguish between states which are less than ε apart (such as the notion based on ε-congruence) or (ii) tells one what stochastic processes are obtained if specific observation functions are applied (such as the notions introduced in the next two paragraphs) will depend on the modeling process and the phenomenon under consideration.

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

First, I have to explain what it means for a stochastic process and a deterministic system as observed with an observation function Φ to give the same predictions relative to accu-

racy $\varepsilon > 0$, $\varepsilon \in \mathbb{R}^+$ (ε indicates that one cannot distinguish differences in probabilistic predictions of less than ε). Plausibly, this means that the possible observed values of the system and the possible outcomes of the process coincide, and that the probabilistic predictions of the observed deterministic system and of the stochastic process differ by less than ε. Strong (Φ, ε)-simulation captures this idea mathematically. I assume that Φ is finite-valued, as it is in practice.

Definition 7 $\{Z_t; t \in \mathbb{R}\}$ strongly (Φ, ε)-simulates *a measure-preserving deterministic system* (M, Σ_M, μ, T_t), *where* $\Phi : M \to \bar{M}$ *is a surjective finite-valued observation function, iff there is a surjective measurable function* $\Psi : M \to \bar{M}$ *such that (i)* $Z_t = \Psi(T_t)$ *for all* $t \in \mathbb{R}$ *and (ii)* $\mu(\{m \in M \,|\, \Psi(m) \neq \Phi(m)\}) < \varepsilon$.

By generalising over Φ and ε one obtains a plausible meaning of the notion that processes of a certain type are observationally equivalent at every observation level to a system. Namely: for every finite-valued Φ and every ε there is a stochastic process of this type which strongly (Φ, ε)-simulates the deterministic system. This notion is attractive because it tells us what probability distributions are obtained when any finite-valued observation function is applied. Yet, to my knowledge, it has not been discussed before.

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

If one asks whether the probability distributions obtained by applying Φ to the deterministic system could derive from a stochastic process of a certain kind, the notion of strong (Φ, ε)-simulation is stronger than what is needed. Is suffices that the observed probability distributions could result from an *observation* of a stochastic process of a certain kind (this will be of some relevance later – see the end of Subsection 5.2). More specifically, if suffices to require the following. Given a deterministic system observed with Φ, there is an observation Γ of a stochastic process such that the following holds: the possible observed outcomes of the process are the possible observed values of the deterministic system, and the probabilistic predictions of the process observed with Γ and the probabilistic predictions of the system observed with Φ differ by less than ε. Technically, weak (Φ, ε)-simulation captures this idea.

Definition 8 $\{Z_t; t \in \mathbb{R}\}$ weakly (Φ, ε)-simulates *a measure-preserving deterministic system* (M, Σ_M, μ, T_t), *where* $\Phi : M \to \bar{M}$ *is a surjective finite-valued observation function, iff 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* $t \in \mathbb{R}$ *and (ii)* $\mu(\{m \in M \,|\, \Psi(m) \neq \Phi(m)\}) < \varepsilon$.

Clearly, if a stochastic process strongly (Φ, ε)-simulates a system, it also weakly (Φ, ε)-simulates it (choose $\Gamma(s) = s$). The converse is generally not true. By generalising over Φ and ε, one obtains a plausible meaning of the notion that stochastic processes of a certain type are observationally equivalent at every observation level to a deterministic system. Namely: for every finite-valued Φ and every ε there is a process of this type which weakly (Φ, ε)-simulates the system. To my knowledge, this notion has never been discussed.

According to all three notions, at every observation level, the data could derive from the deterministic system or a stochastic process of a certain type. So let us see what results obtain for the three notions.

5.2 Stochastic Processes in Science Which Are Observationally Equivalent at Every Observation Level to Deterministic Systems in Science

The next three theorems show that the following holds for our three notions: continuous Bernoulli systems (Definition 4), including several systems in science, are observationally equivalent at every observation level to semi-Markov processes (notion one and three) or n-step semi-Markov processes (notion two), which are widespread in science.

Theorem 2 *Let (M, Σ_M, μ, T_t) be a continuous Bernoulli system where the metric space (M, d_M) is separable*[14] *and Σ_M contains all open sets of (M, d_M). Then for every $\varepsilon > 0$, (M, Σ_M, μ, T_t) is ε-congruent to a semi-Markov process.*

For a proof see Ornstein and Weiss (1991, 93–94). The assumptions of this theorem are fulfilled by all continuous Bernoulli systems in science. For discrete time Werndl (2009a, Section 4) showed that several deterministic systems in science are ε-congruent for all $\varepsilon > 0$ to discrete-time Markov processes (which are widespread in science). However, this left open whether continuous-time deterministic systems in science can be ε-congruent for all $\varepsilon > 0$ to continuous-time stochastic processes in science. Indeed, Werndl's (2009a) results on Markov processes do not carry over to continuous time because, as explained at the end of Subsection 4.2, deterministic systems in science cannot yield continuous-time Markov processes.

Theorem 3 *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 n-step semi-Markov process strongly (Φ, ε)-simulates (M, Σ_M, μ, T_t).*

For a proof see Ornstein and Weiss (1991, 94–95).

Theorem 4 *Let (M, Σ_M, μ, T_t) be a continuous Bernoulli system. Then for every finite-valued observation function Φ and every $\varepsilon > 0$ a semi-Markov process $\{Z_t, t \in \mathbb{R}\}$ weakly (Φ, ε)-simulates (M, Σ_M, μ, T_t).*

For a proof, see Subsection 8.3.

For instance, consider a billiard system with convex obstacles (Example 1). According to Theorem 2, the solutions of such a billiard system are shadowed by the realisations of a semi-Markov process for any accuracy ε. According to Theorem 3, such a billiard system, observed with any finite-valued observation function, is observationally equivalent to an n-step semi-Markov process (disregarding differences in probabilistic predictions of less than ε). According to Theorem 4, such a billiard system, observed with any finite-valued observation function, gives the same predictions as some observed semi-Markov process (disregarding differences in probabilistic predictions of less than ε).

Consequently, I conclude that the answer to the question advanced at the beginning of this section is positive: *there are deterministic systems in science which are observationally equivalent to stochastic processes in science at every observation level.* Compared to the

---

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

results in Section 4 these results are *stronger* because there is observational equivalence at every observation level (and not just for some observation).[15] Note that the increase in strength is very different for the move from the first to the second results (observational equivalence not just to abstract mathematical descriptions but between descriptions in science) and for the move from the second to the third results (observational equivalence between descriptions in science at every, and not just at some, observation level). The third results imply that arguments that the evidence favours a deterministic description fail when they rely on the premise that fine-enough observations of deterministic systems in science yield probability distributions different from the ones of stochastic processes in science. And for this reason the argument advanced at the beginning of this subsection fails. This, of course, is not to claim that there can be no sound arguments for preferring a deterministic or stochastic description, but just that arguments based on alleged general differences between probability distributions of deterministic and stochastic descriptions in science will not work.[16]

In this context note the following. Assume one wants to find out, by applying observation functions, whether the evidence favours a deterministic or stochastic description. According to Theorem 3, certain deterministic systems are observationally equivalent at every observation level to n-step semi-Markov processes. Scientists use n-step semi-Markov processes in many contexts (from physics and biology to the social sciences) where they do not, and we think it is rational that they do not, automatically conclude that this indicates a deterministic description. Thus it is plausible to argue that n-step semi-Markov processes do not indicate deterministic descriptions. Still, n-step semi-Markov processes show correlations in the following sense: the next outcome depends on the past n outcomes, not only on the previous outcome. Suppose that, because of this, n-step semi-Markov processes are taken to indicate a deterministic description. Then one still cannot conclude that fine-enough observations of Bernoulli systems indicate deterministic descriptions. This is so because all one needs is that the observed probability distributions could derive from stochastic process showing no correlations. And, according to Theorem 4, all observations could derive from semi-Markov processes, where the next outcome depends only on the past outcome and thus there are no such correlations.[17]

All in all, the results on observational equivalence are stronger than what one might have expected: observational equivalence holds even between deterministic and stochastic descriptions *found in science* at *every observation level.*

6   Previous Philosophical Discussion

There are hardly any philosophical reflections in the mathematics literature on observational equivalence. Because I discussed the continuous-time results, I can comment on the main exception:

---

[15]The results here concern the same kind of descriptions as in Section 4, namely continuous Bernoulli systems, semi-Markov processes and n-step semi-Markov processes.

[16]The question of which description is preferable is an interesting one but requires another paper. Werndl (2011) is a paper devoted to this question.

[17]It seems worth noting that for the semi-Markov processes of Theorem 4 any arbitrary outcome will be followed by at least two different outcomes. And for many Bernoulli systems any of these semi-Markov processes is such that none of the transition probabilities is close to 0 or 1 (see Subsection 8.3).

Our theorem [Theorem 2] 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 processes governed by roulette wheels and processes governed by Newton's laws. {The popular literature emphasises 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.[18] (Ornstein and Weiss 1991, 39–40)

It is hard to tell what this comment expresses because it is vague and unclear.[19] For instance, why do Ornstein and Weiss highlight coin flipping even though Theorem 2 does not tell us anything about coin flipping (Bernoulli processes) but only about semi-Markov processes? Disregarding that, possibly, Ornstein and Weiss think that semi-Markov processes are random and hence claim that both deterministic and stochastic descriptions can be random. This is widely acknowledged in philosophy (e.g., Eagle 2005). Or maybe Ornstein and Weiss (1991) want to say that deterministic systems in science, when observed with specific observation functions, can be observationally equivalent to stochastic processes in science or, if semi-Markov processes are random, even random processes.[20] This is true and important. However, as discussed in Section 4, this was generally known before Theorem 2 was proven and has been established by theorems which are weaker than Theorem 2. One might have expected Ornstein and Weiss (1991) to say that Theorem 2 shows what I argued that it does, namely that deterministic systems in science are observationally equivalent at every observation level to stochastic processes in science (cf. Subsection 5.2). But they do not seem to say this: because, if they did, it would be unclear why the deterministic representation is mentioned; also, they do not talk about all possible observation levels.

In any case, even if Theorem 2 establishes observational equivalence, 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 there is no conceptual distinction between deterministic and indeterministic descriptions. Regardless of any results on observational equivalence, this distinction remains.

In the philosophy literature the significance of Theorem 2 is taken to be that (at every observation level) there is a choice between a deterministic description in science and a stochastic process: "What is fundamental is that [...] we are in a position to choose either between a deterministic or stochastic model" (Suppes and de Barros 1996, 196). "The fact that a Bernoulli flow can be partitioned in such a way as to yield a (semi-) Markov process illustrates what has been acknowledged all along: Some deterministic systems, when partitioned, generate stochastic processes" (Winnie 1998, 317).

---

[18]The text in braces is in a footnote.

[19]Ornstein and Weiss are mathematicians and not philosophers. So one should not blame them for misguided philosophical claims. Still, one needs to know whether their philosophical claims are tenable; thus I criticise them.

[20]If (M, Σ_M, μ, T_t) and a semi-Markov process $\{Z_t;\ t \in \mathbb{R}\}$ are ε-congruent, there is a finite-valued Φ such that $\{\Phi(T_t);\ t \in \mathbb{R}\}$ is $\{Z_t;\ t \in \mathbb{R}\}$ (cf. Subsection 5.1).

Theorem 2 indeed implies that there is a choice (at every observation level) between deterministic descriptions in science and stochastic descriptions. Yet note that this is already shown by the first level (cf. Subsection 3.1). For Theorem 1 already shows that for every finite-valued observation function there is a choice between a nontrivial stochastic or a deterministic description. This implies that, according to our first notion, the deterministic systems to which Theorem 1 applies are observationally equivalent at every observation level to nontrivial stochastic processes. Formally:

Proposition 1 *Let (M, Σ_M, μ, T_t) be a measure-preserving deterministic system where (M, d_M) is separable and where Σ_M contains all open sets of (M, d_M). Assume that (M, Σ_M, μ, T_t) satisfies the assumption of Theorem 1 and has finite KS-entropy.*[21] *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 are $o_i, o_j \in M_O$ such that $0 < P\{Z_{t+k} = o_j \mid Z_t = o_i\} < 1$.*

This proposition is easy to establish (see Subsection 8.4). And, clearly, according to the second and third notion, every deterministic system (M, Σ_M, μ, T_t) to which Theorem 1 applies is observationally equivalent at every observation level to nontrivial stochastic processes. This is so because the second and third notion quantifies over all finite-valued observations Φ, and $\{\Phi(T_t); t \in \mathbb{R}\}$ is nontrivial by Theorem 1. Thus $\{\Phi(T_t); t \in \mathbb{R}\}$ strongly and weakly (Φ, ε)-simulates (M, Σ_M, μ, T_t) for every $\varepsilon > 0$.

The significance of a theorem is constituted by the new knowledge added and by the conclusions shown by the theorem that have not already been shown by weaker theorems. As just explained, that there is a choice between deterministic and stochastic descriptions follows from much *weaker* theorems than Theorem 2 and was known to the community long before Theorem 2 was proven. Consequently, *the choice (at every observation level) between a deterministic description in science and a stochastic description cannot be the significance of Theorem 2.* Instead, its significance is that deterministic systems in science are observationally equivalent at every observation level even to stochastic processes of the kinds found in science.[22]

Moreover, Suppes and de Barros (1996, 198–200) and Suppes (1999, 189, 192) seem to think that what it means for a deterministic system to be ε-congruent to a certain type of stochastic process for every $\varepsilon > 0$ (the first notion of observational equivalence) is that the system observed with any finite-valued observation function yields a process of a certain type: "we can form a finite partition [...] of the phase space of possible trajectories" (Suppes 1999, 189). That is, they seem to think that the first notion of observational equivalence expresses something like my second notion. But, as discussed in Subsection 5.1, the first and the second notion are quite different (e.g., only the latter tells us what happens if any arbitrary finite-valued observation function is applied).

---

[21]Deterministic systems in science have finite KS-entropy (Ornstein and Weiss 1991, 19).

[22]Suppes (1999, 182) and Winnie (1998, 317) similarly claim that the philosophical significance of the result that some *discrete-time* systems are ε-congruent for all $\varepsilon > 0$ to *discrete-time* Markov processes is that there is a choice between deterministic and stochastic descriptions. Werndl (2009a, Section 4.2.2) already criticised that these claims are too weak but did not explain why (so I did this here).

7 Conclusion

This paper presented and philosophically assessed three types of results on the observational equivalence of deterministic and indeterministic descriptions. These results were about continuous-time measure-theoretic deterministic systems and stochastic processes, which are ubiquitous in science. The main contribution of this paper was to show that there are results on observational equivalence for continuous-time descriptions of increasing strength (namely, about descriptions in science and for every observation level) and to assess their philosophical significance and consequences.

The first results were about observational equivalence to abstract mathematical descriptions. I showed how, when starting with deterministic systems, one finds observationally equivalent stochastic processes, and conversely. The second results were stronger in the sense that they were not about observational equivalence to abstract mathematical descriptions but showed observational equivalence between descriptions of the types found in science. They establish, e.g., that billiard systems with convex obstacles are observationally equivalent to semi-Markov processes. Here I also discussed Kolmogorov's failed attempt to separate deterministic systems in science from deterministic systems which yield stochastic processes in science. Compared to the second results, the third results were stronger because they concerned observational equivalence between descriptions in science at every (and not just at some) observation level. I introduced two new meanings of 'observational equivalence at every observation level'. And I showed that deterministic systems in science can be observationally equivalent even at every observation level to stochastic processes in science. For example, billiard systems with convex obstacles are observationally equivalent at every observation level to semi-Markov or n-step semi-Markov processes.

A consequence of these results is as follows. Suppose one wants to find out whether a phenomenon is best described as deterministic or stochastic. Then one might think of arguing that there is evidence for a deterministic or stochastic description by appealing to the difference between the probability distributions of stochastic processes in science and the probability distributions of (possibly fine-enough) observations of deterministic systems in science. The results of this paper showed that such arguments fail because there is no general difference between the respective probability distributions.

What, then, can we say about the question of whether the deterministic or indeterministic description is preferable? Let me briefly outline my view on this issue. For a thorough treatment, the reader is referred to Werndl (2011). Generally, the answer to this question will vary: in some contexts the deterministic description, in others the indeterministic description might be preferable, and in still other contexts there might be underdetermination. But I argue that there is a clear answer for the specific case of the choice between a deterministic description D which follows from Newtonian mechanics and an indeterministic description I which does not follow from any theory. This includes the main example considered by the extant literature, namely the Newtonian deterministic description of the billiard ball and the corresponding stochastic description which does not follow from any theory. Here the same probabilistic predictions follow from D and I. Still, there are many descriptions in Newtonian mechanics which are not derivable from D or I and which are supported by evidence. Because of the unifying power of Newtonian mechanics, these descriptions provide indirect evidence for D but not for the indeterministic description I. Hence, for these specific cases, indirect evidence provides a reason to prefer

the deterministic description.

8 Appendix

8.1 *Semi-Markov and n-step Semi-Markov Processes*

First, we need the following definition.

Definition 9 *A discrete stochastic process $\{Z_t; t \in \mathbb{Z}\}$ is a family of random variables Z_t, $t \in \mathbb{Z}$, from a probability space $(\Omega, \Sigma_\Omega, \nu)$ to a measurable space $(\bar{M}, \Sigma_{\bar{M}})$.*

Semi-Markov and n-step semi-Markov processes are defined via Markov and n-step Markov processes (for a definition of these well-known processes, see Doob (1953), where it is also explained what it means for these processes to be irreducible and aperiodic).

A semi-Markov process[23] 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, where at time 0 the outcome 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 process. Because at time 0 the semi-Markov process is in S_0 and the process is in S_0 for the time $u(S_0)$, $T_{-1} = u(S_0) - T_0$.

Technically, $\{(S_k, T_k), k \in \mathbb{Z}\}$ 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 surjective; 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 stationary irreducible and aperiodic Markov process with outcome space S; $p_{s_i} = P\{S_0 = s_i\} > 0$, 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, $1 \leq i \leq N$, $P(T_0 \in A \,|\, S_0 = s_i) = \int_A 1/u(s_i) d\lambda$ for all $A \in L((0, u(s_i)])$, where $L((0, u(s_i)])$ is the Lebesgue σ-algebra and λ is the Lebesgue measure on $(0, u(s_i)]$.

Definition 10 *A* semi-Markov process *is a process $\{Z_t; t \in \mathbb{R}\}$ with outcome space S constructed via a process $\{(S_k, T_k), k \in \mathbb{Z}\}$ as follows:*

$Z_t = S_0$ 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 $-T_{-1} - \ldots - T_{-k-1} \leq t < -T_{-1} - \ldots - T_{-k}; k \geq 1$ and thus $t < -T_{-1}$,

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

$$P(Z_0 = s_i) = \frac{p_{s_i} u(s_i)}{p_{s_1} u(s_1) + \ldots + p_{s_N} u(s_N)}. \tag{2}$$

Semi-Markov processes are stationary (Ornstein 1970 and 1974, 56–61). In this paper I assume that the elements of U are irrationally related (u_i and u_j are *irrationally related*

---

[23]The term 'semi-Markov process' is not defined unambiguously. I follow Ornstein and Weiss (1991).

iff $\frac{u_i}{u_j}$ is irrational; and the elements of $U = \{u_1, \dots, u_{\bar{N}}\}$ are *irrationally related* iff for all $i, j, i \neq j, u_i$ and u_j are irrationally related).

n-*step* semi-Markov processes[24] generalise semi-Markov processes.

Definition 11 n-step semi-Markov processes *are defined like semi-Markov processes except that condition (iii) is replaced by: (iii') $\{S_k;\ k \in \mathbb{Z}\}$ is a stationary irreducible and aperiodic n-step Markov process with outcome space S and $p_{s_i} = P\{S_0 = s_i\} > 0$, for all $i, 1 \leq i \leq N$.*

Again, n-step semi-Markov processes are stationary (Park 1982), and I assume that the elements of U are irrationally related.

8.2   *Proof of Theorem 1*

Iff for a 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 (esmz.) $T_n(C) = C$, then the following holds: for every nontrivial finite-valued observation function $\Phi : M \to M_O$, every $k \in \mathbb{R}^+$ and $\{Z_t;\ t \in \mathbb{R}\} = \{\Phi(T_t);\ t \in \mathbb{R}\}$ there are $o_i, o_j \in M_O$ with $0 < P\{Z_{t+k} = o_j \mid Z_t = o_i\} < 1$.

Proof: We need the following definitions.

Definition 12 *A discrete measure-preserving deterministic system is a quadruple (M, Σ_M, μ, T) where (M, Σ_M, μ) is a probability space and $T : M \to M$ is a bijective measurable function such that T^{-1} is measurable and $\mu(T(A)) = \mu(A)$ for all $A \in \Sigma_M$.*

Definition 13 *A discrete measure-preserving deterministic system (M, Σ_M, μ, T) is* ergodic *iff there is no $A \in \Sigma_M$, $0 < \mu(A) < 1$, such that, esmz., $T(A) = A$.*

(M, Σ_M, μ, T) is ergodic iff for all $A, B \in \Sigma_M$ (Cornfeld et al. 1982, 14–15):

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} (\mu(T^n(A) \cap B) - \mu(A)\mu(B)) = 0. \tag{3}$$

Definition 14 $\alpha = \{\alpha_1, \dots, \alpha_n\}, n \in \mathbb{N}$, *is a* partition *of (M, Σ_M, μ) iff $\alpha_i \in \Sigma_M$, $\mu(\alpha_i) > 0$, 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 $M = \bigcup_{i=1}^{n} \alpha_i$.*

A partition is nontrivial iff $n \geq 2$. Given two partitions $\alpha = \{\alpha_1, \dots, \alpha_n\}$ and $\beta = \{\beta_1, \dots, \beta_l\}$ of (M, Σ_M, μ), $\alpha \vee \beta$ is the partition $\{\alpha_i \cap \beta_j \mid i = 1, \dots, n; j = 1, \dots, l\}$. Given a deterministic system (M, Σ_M, μ, T_t), if α is a partition, $T_t\alpha = \{T_t(\alpha_1), \dots, T_t(\alpha_n)\}, t \in \mathbb{R}$, is also a partition.

Note that any finite-valued observation function Φ (cf. Subsection 2.1) can be written as: $\Phi(m) = \sum_{i=1}^{n} o_i \chi_{\alpha_i}(m)$, $M_O = \{o_i \mid 1 \leq i \leq n\}$, for some partition α of (M, Σ_M, μ), where χ_A is the characteristic function of A (cf. Cornfeld et al. 1982, 179).[25] It suffices to prove the following:

---

[24]Again, the term 'n-step semi-Markov process' is not used unambiguously; I follow Ornstein and Weiss (1991).

[25]That is, $\chi_A(m) = 1$ for $m \in A$ and 0 otherwise.

($*$) Iff 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, $1 \leq j \leq r$, $\mu(T_k(\alpha_i) \backslash \alpha_j) > 0$.

For 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 $M_O = \cup_{l=1}^r o_l$. Consequently, the right hand side of ($*$) expresses that for any nontrivial finite-valued $\Phi : M \to M_O$ and all $k \in \mathbb{R}^+$ there is an $o_i \in M_O$ such that for all $o_j \in M_O$, $P\{Z_{t+k} = o_j \mid Z_t = o_i\} < 1$, or equivalently, that there are $o_i, o_j \in M_O$ with $0 < P\{Z_{t+k} = o_j \mid Z_t = o_i\} < 1$.

$\Leftarrow$: 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 $\alpha = \{C, M \backslash C\}$ it holds that $\mu(T_n(C) \backslash C) = 0$ and $\mu(T_n(M \backslash C) \backslash (M \backslash C)) = 0$.

$\Rightarrow$: 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 deterministic system (M, Σ_M, μ, T_k) is ergodic (cf. Definition 13).

Case 1: For every i there is a j such that, esmz., $T_k(\alpha_i) = \alpha_j$. Because the discrete system (M, Σ_M, μ, T_k) is ergodic (equation (3)), 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: There exists an i and a j with, esmz., $T_k(\alpha_i) \subset \alpha_j$ and with $\mu(\alpha_i) < \mu(\alpha_j)$. Because the discrete system (M, Σ_M, μ, T_k) is ergodic (equation (3)), there is a $h \in \mathbb{N}$ such that, esmz., $T_{hk}(\alpha_j) \subseteq \alpha_i$. Hence $\mu(\alpha_j) \leq \mu(\alpha_i)$, contradicting $\mu(\alpha_i) < \mu(\alpha_j) \leq \mu(\alpha_i)$.

8.3 *Proof of Theorem 4*

Theorem 4 *Let (M, Σ_M, μ, T_t) be a continuous Bernoulli system. Then for every finite-valued Φ and every $\varepsilon > 0$ a semi-Markov process $\{Z_t, t \in \mathbb{R}\}$ weakly (Φ, ε)-simulates (M, Σ_M, μ, T_t).*

Proof: Let (M, Σ_M, μ, T_t) be a continuous Bernoulli system, let $\Phi : M \to S$, $S = \{s_1, \ldots, s_N\}$, $N \in \mathbb{N}$, be a surjective observation function, and let $\varepsilon > 0$. Theorem 3 implies that there is a surjective observation function $\Theta : M \to S$, $\Theta(m) = \sum_{i=1}^N s_i \chi_{\alpha_i}(m)$, for a partition α (cf. Definition 14), such that $\{Y_t = \Theta(T_t); t \in \mathbb{R}\}$ is an n-step semi-Markov process with outcomes s_i and corresponding times $u(s_i)$ which strongly (Φ, ε)-simulates (M, Σ_M, μ, T_t).

I need the following definition:

Definition 15 $(M_2, \Sigma_{M_2}, \mu_2, T_t^2)$ *is a* factor *of $(M_1, \Sigma_{M_1}, \mu_1, T_t^1)$ (where both deterministic systems measure-preserving) iff there are $\hat{M}_i \subseteq M_i$ with $\mu_i(M_i \backslash \hat{M}_i) = 0$, $T_t^i \hat{M}_i \subseteq \hat{M}_i$ for all t ($i = 1, 2$), 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) $\phi(T_t^1(m)) = T_t^2(\phi(m))$ for all $m \in \hat{M}_1, t \in \mathbb{R}$.*

Note that the deterministic representation $(X, \Sigma_X, \mu_X, W_t, \Lambda_0)$ of $\{Y_t; t \in \mathbb{R}\}$ is a factor of (M, Σ_M, μ, T_t) via $\phi(m) = r_m$ (r_m is the realisation of m) (cf. Ornstein and Weiss, 1991, 18).

Now I construct a measure-preserving 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 n-step Markov process corresponding to $\{Y_t; t \in \mathbb{R}\}$. 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}, $1 \le i \le N$, 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 σ-algebra of $[0, u(s_i))$. Let μ_{K_i} be the product measure

$$(\mu_\Omega^{\Sigma_\Omega \cap \beta_i} \times \lambda([0, u(s_i)))) / \sum_{j=1}^N u(s_j) \mu_\Omega(\beta_j), \tag{4}$$

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 $\Sigma_\Omega \cap \beta_i$. Let Σ_K be the σ-algebra generated by $\cup_{i=1}^N \Sigma_{K_i}$. Define a pre-measure $\bar{\mu}_K$ on $H = (\cup_{i=1}^N (\Sigma_\Omega \cap \beta_i \times L([0, s_i)))) \cup K$ 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 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 measure-preserving system (a 'flow built under the function f'). $(X, \Sigma_X, \mu_X, W_t)$ is proven to be isomorphic (via a function ψ) to $(K, \Sigma_K, \mu_K, R_t)$ (Park 1982).

Consider $\gamma = \{\gamma_1, \ldots, \gamma_l\} = \beta \vee V\beta \vee \ldots \vee V^{n-1}\beta$ and $\Pi(\omega) = \sum_{j=1}^l o_j \chi_{\gamma_j}(\omega)$, $o_i \ne o_j$ for $i \ne j$, $1 \le i, j \le l$. I now show that $\{B_t = \Pi(V^t(\omega)); t \in \mathbb{Z}\}$ is an irreducible and aperiodic Markov process. By construction, for all t and all i, $1 \le i \le l$, there are $q_{i,0}, \ldots, q_{i,n-1} \in S$ such that

$$P\{B_t = o_i\} = P\{S_t = q_{i,0}, S_{t+1} = q_{i,1}, \ldots, S_{t+n-1} = q_{i,n-1}\}. \tag{5}$$

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

$$P\{B_{t+1} = o_i \mid B_t = o_{j_1}, \ldots, B_{t-k+1} = o_{j_k}\} = \tag{6}$$

$$P\{S_{t+1} = q_{i,0}, \ldots, S_{t+n} = q_{i,n-1} \mid S_t = q_{j_1,0}, \ldots, S_{t+n-1} = q_{i,n-2}, S_{t-1} = q_{j_2,0}, \ldots, S_{t-k+1} = q_{j_k,0}\}$$
$$= P\{S_{t+1} = q_{i,0}, \ldots, S_{t+n} = q_{i,n-1} \mid S_t = q_{j_1,0}, \ldots, S_{t+n-1} = q_{i,n-2}\}$$
$$= P\{B_{t+1} = o_i \mid B_t = o_{j_1}\},$$

if $P\{B_{t+1} = o_i, B_t = o_{j_1}, \ldots, B_{t-k+1} = o_{j_k}\} > 0$. Hence $\{B_t; t \in \mathbb{Z}\}$ is a Markov process. Now a discrete measure-preserving deterministic system (M, Σ_M, μ, T) is mixing iff for all $A, B \in \Sigma_M$

$$\lim_{t \to \infty} \mu(T^t(A) \cap B) = \mu(A)\mu(B). \tag{7}$$

The deterministic representation of every irreducible and aperiodic n-step Markov process, and hence $(\Omega, \Sigma_\Omega, \mu_\Omega, V, \Xi_0)$, is mixing (Ornstein 1974, 45–47). This implies that $\{B_t; t \in \mathbb{Z}\}$ is irreducible and aperiodic.

Let $\Delta(k) = \sum_{i=1}^l o_i \chi_{\gamma_i \times [0, u(o_i))}(k)$, where $u(o_i)$, $1 \le i \le l$, is defined as follows: $u(o_i) = u(s_r)$ where $\gamma_i \subseteq \beta_r$. It follows immediately that $\{X_t = \Delta(R_t); t \in \mathbb{R}\}$ is a semi-Markov

26

process. Consider the surjective function $\Psi : M \to \{o_1, \ldots, o_l\}$, $\Psi(m) = \Delta(\psi(\phi(m)))$ for $m \in \hat{M}$ and o_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, $\{Z_t = \Psi(T_t); t \in \mathbb{R}\}$ is a semi-Markov process with outcomes o_i and times $u(o_i)$, $1 \leq i \leq l$.

Now consider the surjective observation function $\Gamma : \{o_1, \ldots, o_l\} \to S$, where $\Gamma(o_i) = s_r$ for $\gamma_i \subseteq \beta_r$, $1 \leq i \leq l$. By construction, esmz., $\Gamma(\Psi(T_t(m))) = \Theta(T_t(m)) = Y_t(m)$ for all $t \in \mathbb{R}$. Hence, because $\{Y_t; t \in \mathbb{R}\}$ strongly (Φ, ε)-simulates (M, Σ_M, μ, T_t), $\mu(\{m \in M \mid \Gamma(\Psi(m)) \neq \Phi(m)\}) < \varepsilon$.

8.4 *Proof of Proposition 1*

Proposition 1 *Let (M, Σ_M, μ, T_t) be a measure-preserving deterministic system where (M, d_M) is separable and where Σ_M contains all open sets of (M, d_M). Assume that (M, Σ_M, μ, T_t) satisfies the assumption of Theorem 1 and has finite KS-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 are $o_i, o_j \in M_O$ such that $0 < P\{Z_{t+k} = o_j \mid Z_t = o_i\} < 1$.*

Proof: A partition α (cf. Definition 14) is *generating* for the discrete measure-preserving system (M, Σ_M, μ, T) iff 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 $\mu((A \setminus C) \cup (C \setminus A)) < \varepsilon$ (Petersen 1983, 244). α is *generating* for the measure-preserving system (M, Σ_M, μ, T_t) iff for all $A \in \Sigma_M$ there is a $\tau \in \mathbb{R}^+$ and a set C of unions of elements in $\bigcup_{\text{all } m} \bigcap_{t=-\tau}^{\tau} (T^{-t}(\alpha(T^t(m))))$ such that $\mu((A \setminus C) \cup (C \setminus A)) < \varepsilon$ ($\alpha(m)$ is the set $\alpha_j \in \alpha$ with $m \in \alpha_j$).

By assumption, there is a $t_0 \in \mathbb{R}^+$ such that the discrete system $(M, \Sigma_M, \mu, T_{t_0})$ is ergodic (cf. Definition 13). For discrete ergodic systems Krieger's (1970) theorem implies that there is a partition α which is generating for $(M, \Sigma_M, \mu, T_{t_0})$ and hence generating for (M, Σ_M, μ, T_t). 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, Theorem 1 implies that for the process $\{\Psi(T_t); t \in \mathbb{R}\}$ for all $k \in \mathbb{R}^+$ there are o_i, o_j such that $0 < P\{Z_{t+k} = o_j \mid Z_t = o_i\} < 1$. Because α is generating for (M, Σ_M, μ, T_t), β is generating too. This implies that (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, 274). And, by construction, $d_M(m, \Phi_0(\phi(m))) < \varepsilon$ except for a set in M smaller than ε.

Acknowledgements

References

Berkovitz, Joseph, Roman Frigg and Fred Kronz 2006, "The Ergodic Hierarchy, Randomness and Hamiltonian Chaos", Studies in History and Philosophy of Modern Physics 37: 661–691.

Butterfield, Jeremy 2005, "Determinism and Indeterminism", Routledge Encyclopaedia of Philosophy Online.

Cornfeld, Isaak P., Sergej V. Fomin and Yakuv G. Sinai 1982, Ergodic Theory, Berlin: Springer.

Doob, Joseph L. 1953, Stochastic Processes, New York: John Wiley & Sons.

Eagle, Antony 2005, "Randomness Is Unpredictability", The British Journal for the Philosophy of Science 56: 749–790.

Eckmann, Jean-Paul and David Ruelle 1985, "Ergodic Theory of Chaos and Strange Attractors", Reviews of Modern Physics 57: 617–654.

Feldman, Joel and Meir Smorodinksy 1971, "Bernoulli Flows With Infinite Entropy", The Annals of Mathematical Statistics 42: 381–382.

Halmos, Paul 1944, "In General a Measure-preserving Transformation is Mixing", The Annals of Mathematics 45: 786–792.

Halmos, Paul 1949, "Measurable transformations", Bulletin of the American Mathematical Society 55: 1015–1043.

Hopf, Eberhard 1932, "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.

Janssen, Jacques and Nikolaos Limnios 1999, Semi-Markov Models and Applications, Dotrecht: Kluwer Academic Publishers.

Krieger, Wolfgang 1970, "On Entropy and Generators of Measure-preserving Transformations", Transactions of the American Mathematical Society 149: 453–456.

Lorenz, Edward 1963, "Deterministic Nonperiodic Flow", Journal of the Atmospheric Sciences 20: 130–141.

Luzzatto, Stefano, Ian Melbourne and Frederic Paccaut 2005, "The Lorenz Attractor Is Mixing", Communications in Mathematical Physics 260: 393–401.

Ornstein, Donald 1970, "Imbedding Bernoulli Shifts in Flows", in Albrecht Dold and Beno Eckmann (eds.), Contributions to Ergodic Theory and Probability, Proceedings of the First Midwestern Conference on Ergodic Theory. Springer: Berlin, 178–218.

Ornstein, Donald 1974, Ergodic Theory, Randomness, and Dynamical Systems, New Haven and London: Yale University Press.

Ornstein, Donald and Giovanni Galavotti 1974, "Billiards and Bernoulli Schemes", Communications in Mathematical Physics 38: 83–101.

Ornstein, Donald and Benjamin Weiss 1991, "Statistical Properties of Chaotic Systems", Bulletin of the American Mathematical Society 24: 11–116.

Park, Kong 1982, "A Special Family of Ergodic Flows and Their $\bar{d}$-Limits", Israel Journal of Mathematics 42: 343–353.

Petersen, K. 1983, Ergodic Theory, Cambridge: Cambridge University Press.

Radunskaya, Amy 1992, Statistical Properties of Deterministic Bernoulli Flows, Ph.D. Dissertation, Stanford: University of Stanford.

Simányi, Nándor 2003, "Proof of the Boltzmann-Sinai Ergodic Hypothesis for Typical Hard Disk Systems", Inventiones Mathematicae 154: 123–178.

Sinai, Yakov G. 1989, "Kolmogorov's Work on Ergodic Theory", The Annals of Probability 17: 833–839.

Strogatz, Steven H. 1994, Nonlinear Dynamics and Chaos, with Applications to Physics, Biology, Chemistry, and Engineering, New York: Addison Wesley.

Suppes, Patrick 1999, "The Noninvariance of Deterministic Causal Models", Synthese 121: 181–198.

Suppes, Patrick and Acacio de Barros 1996, "Photons, Billiards and Chaos", in Paul Weingartner and Gerhard Schurz (eds.), Law and Prediction in the Light of Chaos Research. Berlin: Springer, 189–207.

Uffink, Jos 2007, "Compendium to the Foundations of Classical Statistical Physics", in Jeremy Butterfield and John Earman (eds.), Philosophy of Physics (Handbooks of the Philosophy of Science B). Amsterdam: North-Holland, 923–1074.

Werndl, Charlotte 2009a, "Are Deterministic Descriptions and Indeterministic Descriptions Observationally Equivalent?", Studies in History and Philosophy of Modern Physics 40: 232–242.

Werndl, Charlotte 2009b, "What Are the New Implications of Chaos for Unpredictability?", The British Journal for the Philosophy of Science 60: 195–220.

Werndl, Charlotte 2011, "On Choosing Between Deterministic and Indeterministic Models: Underdetermination and Indirect Evidence". Forthcoming in: Synthese.

Winnie, John 1998, "Deterministic Chaos and The Nature of Chance", in John Earman and John Norton (eds.), The Cosmos of Science – Essays of Exploration. Pittsburgh: Pittsburgh University Press, 299–324.