# From von Mises' Impossibility of a Gambling System to Probabilistic Martingales

**MSc Thesis** *(Afstudeerscriptie)*

written by

**Francesca Zaffora Blando**
(born January 27th, 1988 in Chivasso, Italy)

under the supervision of **Prof. Dr. Paul Vitányi** and **Prof. Dr. Michiel van Lambalgen**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

**MSc in Logic**

at the *Universiteit van Amsterdam.*

| **Date of the public defense:** | **Members of the Thesis Committee:** |
|---|---|
| September 16th, 2015 | Prof. Dr. Johan van Benthem |
| | Prof. Dr. Michiel van Lambalgen (supervisor) |
| | Dr. Piet Rodenburg |
| | Dr. Leen Torenvliet |
| | Prof. Dr. Paul Vitányi (supervisor) |
| | Prof. Dr. Ronald de Wolf (chair) |

Institute for Logic, Language and Computation

*Per Siś Più Otto!*

# Abstract

Algorithmic randomness draws on computability theory to offer rigorous formulations of the notion of randomness for mathematical objects. In addition to having evolved into a highly technical branch of mathematical logic, algorithmic randomness prompts numerous methodological questions. This thesis aims at addressing some of these questions, together with some of the technical challenges that they spawn. In the first part, we discuss the work on randomness and the foundations of probability of the Austrian mathematician Richard von Mises [1919], whose theory of *collectives* constitutes the first attempt at providing a formal definition of randomness. Our main objective there is to ascertain the reasons that led to the demise of von Mises' approach in favour of *algorithmic* randomness. Then, we turn to the myriad definitions of randomness that have been proposed within the algorithmic paradigm, and we focus on the issue of whether any of these definitions can be said to be more legitimate than the others. In particular, we consider some of the objections that have been levelled against Martin-Löf randomness [1966] (arguably, the most popular notion of algorithmic randomness in the literature), concentrating on the famous critique of Martin-Löf randomness due to Schnorr [1971a] and on a more recent critique due Osherson and Weinstein [2008], which relies on a learning-theoretic argument. We point out the inconclusiveness of these criticisms, and we recommend a pluralistic approach to algorithmic randomness. While appraising Osherson and Weinstein's critique, we also allow ourselves a brief learning-theoretic digression and further study the notion of Kurtz randomness in learning-theoretic terms. In light of the increasing amount of attention being paid to Schnorr's critique of Martin-Löf randomness in the literature, in the second part of this thesis we consider some of the technical implications of taking said critique seriously. In their paper on probabilistic algorithmic randomness [2013], Buss and Minnes countenance Schnorr's critique by offering a characterisation of Martin-Löf randomness in terms of *computable* probabilistic martingales (betting strategies). Buss and Minnes also ask whether there are any natural conditions on the class of probabilistic martingales that can be used to characterise other common algorithmic randomness notions. We answer their question in the affirmative both in the monotonic and the non-monotonic setting, by providing probabilistic characterisations of Martin-Löf randomness, Schnorr randomness, Kurtz randomness and Kolmogorov-Loveland randomness.

# Contents

# Acknowledgments

First of all, I would like to thank my supervisors, Paul Vitányi and Michiel van Lambalgen, for their guidance and support during my excursion into the exciting world of algorithmic randomness. Paul, I am especially grateful for your availability, even under very stringent time constraints, and for your many insightful comments and suggestions.

Secondly, my gratitude goes to my committee members—Ronald de Wolf, Johan van Benthem, Piet Rodenburg and Leen Torenvliet—for taking the time to read this thesis and for their thought-provoking questions and comments.

During my time at the ILLC, Ulle Endriss, Sonja Smets and Alexandru Baltag have been a constant source of good advice and encouragement in all matters, both academic and personal: thank you.

My warm thanks also go to Tanja Kassenaar, for all her help and her great patience.

Andrea and Michele, thank you for the countless dinners, for helping us wash the dishes in the bathtub, for all the foosball matches and game nights, and for our nocturnal set theory marathons.

Piotr and Magda, thank you for reminding me and Krzyś that there is more to life than 'little worms': there are also STOCHASTIC processes!

(Mamma $\wedge$ Papà), grazie per avermi consentito di arrivare fin qui, per avermi sempre appoggiata in tutti questi anni, incondizionatamente—anche se, dopo che avete letto Logicomix, la morale della storia sembra essere che i logici finiscono per diventare tutti un po' matti.

Sisio, I lost count of all the stairs!

# Chapter 1

# Introduction

> Why should we fear, when chance rules everything, And foresight of the future there is none; 'Tis best to live at random, as one can.
>
> —Sophocles, *Oedipus Rex*

> Chance, too, which seems to rush along with slack reins, is bridled and governed by law.
>
> —Boëthius, *The Consolation of Philosophy*

## 1.1 Motivation

In everyday parlance, the adjective 'random' is typically regarded as being synonymous with 'chancy' or 'unpredictable', an intuition which likely stems from people's experience with games of chance such as coin tossing or roulette. Phenomena standardly categorised as random are encountered all the time in both science and daily life: from the evolution of a bacterial population to the path of a particle in Brownian motion, to the behaviour of the stock market, to the method used to generate the encrypted password needed to access one's Studielink account.

The presence of different conceptions of randomness prompts several fascinating questions: are there any core properties that all such conceptions share? Is there a common randomness notion underlying all of these separate accounts? One may even wonder whether it is at all possible to render such a nebulous concept mathematically precise: after all, is it not part of the elusive essence of randomness not to be amenable to any formal treatment?

Puzzles pertaining to chance and randomness are traditionally taken to fall under the jurisdiction of probability theory. Yet, probability theory does not deal with

the notion of randomness per se: in no textbook on the probability calculus will one find a formal definition of 'random object'. Of course, one will encounter, for instance, the notion of 'random variable', but random variables are simply measurable functions and they need not be unpredictable or random in any way. So, given the ubiquity of randomness (or, at least, of *references* to randomness) in science, a better understanding of this concept appears to be of the essence—but also to be fraught with challenges, both mathematical and methodological.

Like most questions at the basis of scientific inquiry, these questions are philosophical in nature, and there is no incontrovertible way to answer them. Modern attempts at formulating a rigorous definition of randomness do not try to provide an all-encompassing theory of the 'randomness' we perceive as occurring in natural phenomena; instead, they focus on the less ambitious, but more manageable task of providing a reasonable characterisation of randomness for mathematical objects.

The mathematical model of randomness that will take centre stage in this thesis is the theory of *algorithmic randomness*, which combines statistical intuitions with tools from classical computability theory in order to give substance to the common impression that randomness amounts to 'lawlessness', 'disorder', 'irregularity', 'patternlessness' and 'unpredictability'.

The algorithmic theory of randomness has been built upon the posit that the concept of randomness, in spite of hinging on probability theory, cannot be exhausted by it. The language of probability theory—the argument goes—is not expressive enough, for, in the words of Li and Vitányi,

> it can only express expectations of properties of outcomes of random processes, that is, the expectations of properties of the total set of sequences [of experiments] under some distribution [1993/1997/2008, p. 48].

Yet, when talking about randomness, a crucial issue is that of determining whether a *single* sequence of experimental outcomes counts as being random or not. Probability theory, however, is by itself unequipped to formalise the notion of randomness for individual sequences of outcomes.

To see why this is taken to be the case, consider the following familiar example[1]. Suppose that you observe the two binary strings below, both of which record the results of fifty consecutive and independent tosses of a putative fair coin (where 1 stands for 'heads' and 0 for 'tails'):

$$10010001001100111011001010100000110100001100011100,$$
$$01010101010101010101010101010101010101010101010101.$$

While the first sequence of outcomes is at least random-looking, it is safe to presume that very few people would be willing to call the second one random, because of its

[1] See, for instance, [Li and Vitányi, 1993/1997/2008, § 1.8.1].

perceivable regularity. In fact, observing the second sequence could easily prompt one to question the fairness of the coin used in the experiment. This sequence seems to be too easy to predict to have been generated by a random process: after seeing its initial digits, one would likely feel rather confident in their ability to correctly predict the value of the subsequent bits—a feeling that the first sequence does not so readily afford. Yet, a simple probabilistic analysis does not allow one to discriminate between these two sequences, because, under the fair coin assumption, both of them are assigned the same probability: namely, $2^{-50}$.

Along a similar vein, suppose that you notice, among some scrabble pieces scrambled on the table, some letters arranged to spell the word

SUPERCALIFRAGILISTICEXPIALIDOCIOUS[2].

How likely is it that this arrangement is the result of a random process? It would seem more plausible that someone deliberately placed the scrabble pieces in this meaningful way. However, under the uniform distribution, all letter arrangements of length 34 are equally likely. So, how can we account for the intuition that the word SUPERCALIFRAGILISTICEXPIALIDOCIOUS was not randomly generated?

Problems of this sort were already troubling the early probabilists d'Alembert, Condorcet and Laplace. Nowadays, as noted above, a conclusion often drawn is that these considerations show that probability theory is simply not fine-grained enough to characterise the notion of randomness for individual mathematical objects. Moreover, algorithmic randomness is generally regarded as having provided a satisfactory solution to this problem.

Over the past fifty years, algorithmic randomness has evolved into a mature and very rich area of research in mathematical logic, revealing deep connections with Turing degree theory, information theory, complexity theory and computable mathematics in general. However, algorithmic randomness is not the only fish in the sea when it comes to the issue of appropriately formalising the concept of randomness. Notably, in the first half of the twentieth century, the Austrian mathematician Richard von Mises vigorously argued that the notion of randomness, rather than being somewhat subordinate to that of probability, is necessary to account for the empirical meaning of the latter: that is, to explain how the mathematical theory of probability comes to be applicable to real-life phenomena [1919].

Although von Mises' approach has been by and large abandoned, it prompts plentiful interesting questions regarding the connections between randomness and the foundations of probability theory—and, as noted by van Lambalgen [1987a], even the foundations of mathematics as a whole. For example, what exactly are the shortcomings of von Mises' paradigm, and what are the implications of embracing the algorithmic theory of randomness for the debate over the various interpretations

[2]This is a slightly modified version of an example first illustrated by Pierre-Simon Laplace in [1819/1952] (see, for instance, [Li and Vitányi, 1993/1997/2008, Chapter 4]).

of probability? Which notions of algorithmic randomness, if any, render more justice to von Mises' intuitions?

In addition to these issues related to von Mises' foundational project, algorithmic randomness spurs a whole host of methodological questions by itself. For instance, is there one 'correct' or 'true' definition of algorithmic randomness? Are there any uncontroversial criteria that any satisfactory notion of randomness should satisfy? Is it legitimate to appeal to epistemic considerations when trying to persuade someone of the correctness of some algorithmic randomness concept?

Attempts to provide (at least partial) answers to the above questions will keep us busy throughout this thesis.

## 1.2 Thesis outline

Broadly speaking, this thesis is concerned with further exploring some philosophical and technical issues surrounding the problem of providing a satisfactory definition of algorithmic randomness—with a special focus on epistemic characterisations of randomness.

We begin in Chapter 2 by discussing in detail von Mises' theory of *collectives*. We consider the most common objections raised against von Mises' definition of randomness (now known under the label of *stochasticity*) in terms of the impossibility of a gambling system, and we end up agreeing with van Lambalgen [1987a] that these criticisms are cogent only if one already disagrees with von Mises' frequentist interpretation of probability. However, we also argue that there are some convincing objections against von Mises' frequentist approach, especially in view of his own stated intention to reconnect probability theory with its empirical roots. We also highlight a certain tension between the more epistemic aspects of von Mises' theory and its objectivistic core.

In Chapter 3, we review the most common definitions of randomness for finite binary strings and infinite binary sequences from the algorithmic randomness literature. First, we describe the *incompressibility paradigm*, according to which a string is random if, roughly, it is hard to describe by a Turing machine, if it does not display any pattern or regularity that a Turing machine can exploit. Then, we discuss the *measure-theoretic typicality paradigm*, which is based on the intuition that random sequences should satisfy all measure-one properties which can be 'effectively tested'. Lastly, we illustrate the *unpredictability paradigm*, which is closely connected to von Mises' identification of randomness with the impossibility of a successful gambling system. In this setting, a sequence is said to be random if there is no effective betting strategy which allows a gambler to win an infinite amount of money when playing against that sequence. We then review the connections between these different paradigms, and we provide a characterisation in terms of betting strategies of a

stochasticity notion introduced by Vermeeren [2013] called *weak Church stochasticity*. We also offer another proof of the fact that *Schnorr randomness* implies weak Church stochasticity using a compactness-based argument. To nicely complete the picture of the interrelations between randomness and stochasticity notions (see Figure 3.2), we conclude by introducing the concept of *weak Kolmogorov-Loveland stochasticity*.

In Chapter 4, we consider the *monism vs. pluralism* debate in algorithmic randomness: that is, the dispute over whether any single notion of algorithmic randomness may be said to best capture our pre-theoretic intuitions about randomness (monism)—much in the same way as Turing-machine computability is thought to capture the intuitive notion of effective calculability—or whether there are instead several notions of algorithmic randomness which fit the bill (pluralism). We consider three randomness theses that have been proposed in the literature—the *Martin-Löf Thesis* [Delahaye, 1993], *Schnorr's Thesis* [Schnorr, 1971b] and the *Weak 2-Randomness Thesis* [Osherson and Weinstein, 2008]—each of which claims that the corresponding notion of algorithmic randomness coincides with 'true' randomness. We argue that all three theses, taken in isolation, are ultimately wanting and defend a pluralist perspective on algorithmic randomness. However, we note that these theses—and, in particular, the two critiques of *Martin-Löf randomness* that respectively lie at the heart of Schnorr's Thesis and the Weak 2-Randomness Thesis—raise some interesting questions about the role of epistemic considerations in algorithmic randomness. In particular, in spite of pushing for two very different randomness concepts, Schnorr's and Osherson and Weinstein's objections against Martin-Löf randomness are, in the end, surprisingly similar to each other in spirit. After considering Osherson and Weinstein's critique, we also have a brief detour on randomness and computational learning theory, and we present a learning-theoretic characterisation of *Kurtz randomness* in addition to Osherson and Weinstein's own characterisation.

In Chapter 5, in light of the increasing appreciation given to Schnorr's critique of Martin-Löf's definition, we focus on a mathematical framework, introduced by Buss and Minnes in [2013], which takes said critique seriously and allows to provide a characterisation of Martin-Löf randomness in computable terms by appealing to the notion of a probabilistic betting strategy. We extend Buss and Minnes' probabilistic paradigm in a natural way both in the *monotonic* and the *non-monotonic* setting, introducing the following randomness notions: (weak) KEx-*randomness*, (weak) KP1-*randomness*, (weak) WEx-*randomness*, (weak) WP1-*randomness*, *non-monotonic* P1-*randomness* and *non-monotonic* Ex-*randomness*. We then show that several of the above concepts coincide with standard randomness notions (namely, with Martin-Löf randomness, Schnorr randomness, Kurtz randomness or Kolmogorov-Loveland randomness). This indicates that this probabilistic framework can be used to provide a uniform characterisation of the most common algorithmic randomness notions. It is also our hope that it may eventually lead to the identification of novel interesting

randomness concepts, and that it might even shed some light on the long-standing question of whether Kolmogorov-Loveland randomness coincides with Martin-Löf randomness.

In Chapter 6, we conclude this thesis by identifying some open questions and potential future research paths.

## 1.3 Contributions

To sum up, the main technical contributions of this thesis are the following:

(i) We provide characterisations of weak Church stochasticity and weak Kolmogorov-Loveland stochasticity (Proposition 3.3.10 and Proposition 3.3.17, respectively) based on the notion of a 'simple martingale which always eventually bets'.

(ii) We give another proof of Theorem 3.3.11 (whose original proof can be found in [Vermeeren, 2013]), which establishes that Schnorr randomness implies weak Church stochasticity.

(iii) We extend the learning-theoretic framework proposed by Osherson and Weinstein in [2008] by defining the notion of *sequence identification with no mind changes* (Definition 4.2.5), and we prove that Kurtz randomness can be given a further learning-theoretic characterisation via this identification criterion (Proposition 4.2.6).

(iv) We prove that Schnorr randomness is implied by weak KP1-randomness (Proposition 5.3.6) and, *a fortiori*, by KP1-randomness; then, we show that Schnorr randomness is equivalent to weak KEx-randomness (Theorem 5.3.7), and that Martin-Löf randomness is equivalent to KEx-randomness (Theorem 5.3.8). We also prove that Kurtz randomness is equivalent to both weak WEx-randomness and weak WP1-randomness (Theorem 5.3.10 and Theorem 5.3.11, respectively).

(v) We establish that Kolmogorov-Loveland randomness is equivalent to non-monotonic P1-randomness (Theorem 5.4.6), and that Martin-Löf randomness is equivalent to non-monotonic Ex-randomness (Corollary 5.4.8).

## 1.4 Notation and background notions

We close this introductory chapter by fixing the notation that will be used throughout the thesis and by defining a few preliminary notions.

As already mentioned, in what follows we will only be dealing with randomness notions for finite and infinite binary sequences, as is customary within the field of algorithmic randomness.

We denote the set of finite binary sequences by $\{0,1\}^*$. We refer to the elements of this set as *strings* and use lowercase Greek letters towards the end of the alphabet (e.g., $\sigma$ or $\tau$) to denote them—except for the empty string, which is denoted by $\varepsilon$. For each $n \in \mathbb{N}$, we let $\{0,1\}^n$ be the subset of $\{0,1\}^*$ consisting of all strings of length $n$. Similarly, $\{0,1\}^{\leq n}$ is the set of all strings of length at most $n$. The length of a string $\sigma$, in turn, is denoted by $|\sigma|$. We denote by $\sigma(n)$ the $n$-th bit of $\sigma$ (where the enumeration is taken to start at 1), while $\sigma \upharpoonright n$ is the initial segment, or prefix, of $\sigma$ consisting of its first $n$ bits (if $n > |\sigma|$, then we set $\sigma \upharpoonright n = \sigma$). For any string $\sigma \neq \varepsilon$, we let $\sigma^-$ denote the initial segment of $\sigma$ consisting of all of its digits except for the last one. By $\sigma\tau$, we mean the concatenation of the two strings $\sigma$ and $\tau$. If $\sigma$ is a prefix of $\tau$, we write $\sigma \sqsubseteq \tau$; we write $\sigma \sqsubset \tau$ if it is a proper prefix. A set $\mathcal{S} \subseteq \{0,1\}^*$ is said to be *prefix-free* if and only if, for any strings $\sigma, \tau \in \mathcal{S}$, neither $\sigma$ nor $\tau$ is a prefix of the other string.

The set of infinite binary sequences is denoted by $\{0,1\}^\omega$. We refer to the elements of this set as *sequences*, and we use uppercase Roman letters towards the end of the alphabet (e.g., $X$ or $Y$) to denote them. The terms $|X|$, $X(n)$, $X \upharpoonright n$, $\sigma X$ and the relation $\sigma \sqsubset X$ are defined analogously to the previous paragraph.

Note that infinite binary sequences can be interpreted both as sets of natural numbers and as real numbers in $[0,1]$: any $X \in \{0,1\}^\omega$ naturally corresponds to the set $S_X = \{n \in \mathbb{N} : X(n) = 1\}$ (i.e., $X$ is the characteristic function of $S_X$) and to the real $\alpha_X = \sum_{n \in \mathbb{N}} X(n) \cdot 2^{-n} \in [0,1]$.

It is important to note that—with the exception of von Mises' definition of randomness (which will be discussed in Chapter 2)—every notion of (algorithmic) randomness that we will consider in this thesis is to be understood relative to an a priori fixed probability measure. Without such a fixed measure, these notions of randomness would not make sense: each measure determines a specific set of random sequences, and a sequence which is random relative to a given measure may not count as random relative to a different one.

Here, we restrict attention to the *uniform* or *Lebesgue measure*[3] over *Cantor space*[4], the topological space consisting of $\{0,1\}^\omega$ together with the *product topology*[5]. The product topology is generated by the collection of *open cylinders*: that is, all sets of the form $[\![\sigma]\!] = \{X \in \{0,1\}^\omega : \sigma \sqsubset X\}$, for some $\sigma \in \{0,1\}^*$. The intersection of a finite number of open cylinders is called a *cylinder set*. Cylinder sets are clopen: i.e., they are both open and closed in the topology. Since they are elements of the topology, cylinder sets are open by definition. Moreover, since the complement of

[3]The reader interested in notions of randomness relative to measures other than the Lebesgue measure may consult, e.g., [Martin-Löf, 1966] or [Levin, 1973, 1976, 1984].

[4]Although the notions of algorithmic randomness discussed in this thesis will all be defined over the Cantor space of infinite binary sequences, it is possible to characterise randomness for spaces other than this. We will briefly return to these possible extensions in Section 4.3.

[5]The set $\{0,1\}^\omega$ may be viewed as the product of countably many copies of $\{0,1\}$, where each copy is equipped with the *discrete topology*.

a cylinder set is a union of cylinders, cylinder sets are also closed. Hence, they are clopen. By the definition of a topology base, every open set in the topology can be written as a union of open cylinders. In particular, every open set is equal to $\bigcup\{[\![\sigma]\!] \subseteq \{0,1\}^\omega : \sigma \in \mathcal{S}\}$, for some prefix-free set of strings $\mathcal{S}$. For ease of notation, we abbreviate $\bigcup\{[\![\sigma]\!] \subseteq \{0,1\}^\omega : \sigma \in \mathcal{S}\}$ as $[\![\mathcal{S}]\!]$.

Intuitively, for every $\mathcal{X} \subseteq \{0,1\}^\omega$, the Lebesgue measure $\lambda(\mathcal{X})$ of $\mathcal{X}$ is the probability that $X \in \mathcal{X}$, when $X$ is the result of infinitely many independent tosses of a fair coin. Formally, for every cylinder $[\![\sigma]\!]$, we set $\lambda([\![\sigma]\!]) = 2^{-|\sigma|}$. A sequence of cylinders $\{[\![\sigma]\!] \subseteq \{0,1\}^\omega : \sigma \in \mathcal{S}\}$ is said to *cover* a set $\mathcal{X} \subseteq \{0,1\}^\omega$ if $\mathcal{X} \subseteq [\![\mathcal{S}]\!]$. The *outer Lebesgue measure* of $\mathcal{X}$ is defined as

$$\lambda^*(\mathcal{X}) = \inf\Big\{ \sum_{n\in\mathbb{N}} 2^{-|\sigma_n|} : \{[\![\sigma_n]\!] \subseteq \{0,1\}^\omega : n \in \mathbb{N}\} \text{ covers } \mathcal{X}\Big\}.$$

The *inner Lebesgue measure* of $\mathcal{X}$ is $\lambda_*(\mathcal{X}) = 1 - \lambda^*(\overline{\mathcal{X}})$, where $\overline{\mathcal{X}}$ is the complement of $\mathcal{X}$ in $\{0,1\}^\omega$. If $\mathcal{X}$ is a measurable set, then $\lambda^*(\mathcal{X}) = \lambda_*(\mathcal{X})$. In this case, we simply write $\lambda(\mathcal{X})$ and we refer to it as the Lebesgue measure of $\mathcal{X}$. A set is said to be a *null set* if it has Lebesgue measure 0. In turn, a set $\mathcal{X} \subseteq \{0,1\}^\omega$ has Lebesgue measure 0 if and only if there is a collection $(\mathcal{U}_n)_{n\in\mathbb{N}}$ of open subsets of $\{0,1\}^\omega$ such that $\lim_{n\to\infty} \lambda(\mathcal{U}_n) = 0$ and $\mathcal{X} \subseteq \bigcap_{n\in\mathbb{N}} \mathcal{U}_n$.

# Chapter 2

# Von Mises' Axiomatisation of Random Sequences

At the second International Congress of Mathematicians held in Paris in 1900, David Hilbert presented a list of twenty-three unsolved problems that he believed mathematicians should focus on in the years to come. Among them, he proposed

> To treat in the same manner [as geometry], by means of axioms, those physical sciences, in which mathematics plays an important part; in the first rank are the Theory of Probabilities and mechanics [Hilbert, 1902].

The idea that probability theory is a physical science just like mechanics—rather common among nineteenth-century mathematicians—was adopted by Richard von Mises, an applied mathematician with a penchant for constructivist mathematics, and a declared logical positivist. Around the second decade of the twentieth century, in an attempt to provide solid and empirically-motivated foundations for probability theory and statistics, von Mises worked out a view of probability known as *strict frequentism*[6] based on the notion of random sequences of trials [1919]. In particular, von Mises was the first one to provide a (more or less) formal definition of the concept of randomness—a definition which, from the very beginning, sparked a lot of discussion and, ultimately, gave rise to the field of algorithmic randomness (which went on to become the orthodox approach to defining randomness in mathematics).

In this chapter, we will review von Mises' seminal work on probability and randomness, and we will appraise some of the objections that have been raised, or can be raised, against his approach.

[6] This name was coined by van Lambalgen in [1987b].

## 2.1 Von Mises' strict frequentism

The frequency interpretation of probability has its origins in the work of Ellis and Venn around the mid nineteenth century, and it may be viewed as a 'British empiricist' reaction contra Laplace's 'continental rationalism' [Gillies, 2000]. Von Mises' main contribution was to enrich Ellis' and Venn's pre-existing frequentist theory by integrating it with his account of randomness—which was aimed at restricting the domain of applicability of probability theory to 'truly' random phenomena.

This first section will be devoted to presenting von Mises' theory in some detail. First, we will illustrate the empirical laws that von Mises singles out as forming the appropriate basis for probability theory. Then, we will discuss von Mises' own axiomatisation of frequentist probability.

### 2.1.1 The empirical laws of probability

While mechanics is concerned with the behaviour of physical bodies when subjected to forces or displacements, the subject matter of probability theory is, according to von Mises, the study of *random* observable iterative events and mass phenomena: i.e., of "problems in which either the same event repeats itself again and again or a great number of uniform elements are involved at the same time" [von Mises, 1928/1961, p. 11]. In particular, von Mises mentions three types of phenomena to which probabilities may be assigned:

(i) all games of chance (e.g., coin tossing or die rolling);

(ii) some branches of physics such as the kinetic theory of gases, Brownian motion, radioactivity and Planck's theory of blackbody radiation;

(iii) biological statistics.

A first thing to note is that von Mises' approach is *objectivistic*, in that probabilities are taken to be ascribable to 'truly' random phenomena only: they are perceived as being a property of the random experiment (e.g. tossing a coin) under consideration[7]. Moreover, in von Mises' view, single-case probability ascriptions do not make sense, as attested by the following quote:

> Our probability theory has nothing to do with questions such as: "Is there a probability of Germany being at some time in the future involved in a war with Liberia?" [1928/1961, p. 9].

[7]This is in clear contrast with the *subjectivist* viewpoint endorsed by, e.g., Ramsey [1931], de Finetti [1937; 1972] and Savage [1954], according to which probabilities are measures of a subject's uncertainty. From this perspective, probabilities can be assigned to anything that the subject is uncertain about, so the question of whether there is an underlying random process becomes irrelevant.

By restricting the scope of probability theory in this manner, von Mises hopes to be able to reduce the concept of probability to an observable and measurable quantity. Following Gillies [2000], we will call the processes singled out by von Mises as the object of probability theory *empirical collectives*.

Being a radical logical positivist, von Mises believes that the *hypothetico-deductive method*[8] is applicable not only to the empirical sciences, but to mathematics, as well. More precisely, von Mises advocates the view that mathematical concepts are obtained by means of abstracting some aspects of our every-day experiences. He contends that mathematicians extrapolate the empirical laws obeyed by the phenomena upon which mathematical concepts rely on the basis of observation. From these empirical laws, mathematicians can then derive the axioms of mathematical theories. In turn, mathematical theories allow mathematicians to deduce consequences, which provide predictions or explanations for further observable phenomena (see Figure 2.1).

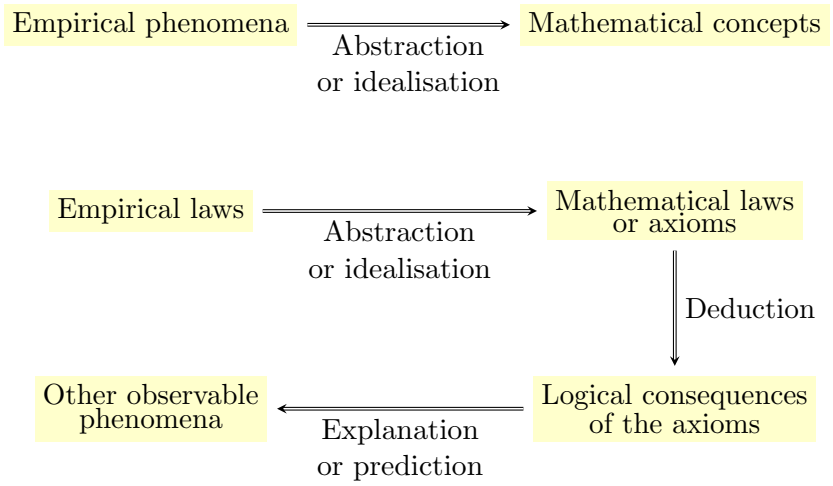



Figure 2.1: Von Mises' view of the formation of mathematical concepts/axioms and their relation to phenomena which can be verified/falsified through observation [Gillies, 2000, p. 91].

According to von Mises, this methodology is applicable to probability theory, as well:

> We take it as understood that probability theory, like theoretical mechanics or geometry, is a scientific theory of a certain domain of observed phenomena. If we try to describe the known modes of scientific research we may say: all exact science starts with observations, which, at the outset, are formulated in ordinary language; these inexact formulations are made

[8] For a discussion of the hypothetico-deductive method see, for instance, [Crupi, 2014].

> more precise and are finally replaced by axiomatic assumptions, which, at the same time, define the basic concepts. Tautological (= mathematical) transformations are then used in order to derive from these assumptions conclusions, which, after retranslation into common language, may be tested by observations, according to operational prescriptions [1964, p. 43].

In particular, von Mises claims that there are two fundamental laws that empirical collectives satisfy, and which can therefore serve as the basis of an empirically-grounded probability calculus. The first one is the so-called *law of stability of statistical frequencies*:

> **Law of stability of statistical frequencies**[9]: in an empirical collective, the relative frequency of an event becomes more and more stable as the number of observations increases.

The validity of the law of stability of statistical frequencies, argues von Mises, is corroborated by abounding observations in all games of chance. Such a conviction appears to have been prevailing within the wider mathematical community, as well:

> The fact that in a number of instances the relative frequency of random events in a large number of trials is almost constant compels us to presume the existence of certain laws, independent of the experimenter, that govern the course of these phenomena and that manifest themselves in this near constancy of the relative frequency [Gnedenko, 1968, p. 55].

> This observed regularity, that the frequency of appearance of any random event oscillates about some fixed number when the number of experiments is large, is the basis of the notion of probability [Fisz, 1963, p. 5].

> In spite of the irregular behavior of individual results, the average results of long sequences of random experiments show a striking regularity [Cramér, 1957, p. 141].

Von Mises' second law, on the other hand, states that empirical collectives, although characterised by a certain global regularity which derives from the law of stability of statistical frequencies, are, locally, extremely irregular or 'random'. Consider, as an illustration, the case of coin tossing with a fair coin: as amply confirmed by experience, a gambler placing bets on, say, 'heads' cannot be better off in the long run by wagering according to some strategy, instead of wagering in accordance with the outcomes of a second fair coin's flips. As colourfully put by von Mises,

[9] This terminology was introduced by Keynes [1921/1963].

> This impossibility of affecting the chances of a game by a system of selection, this uselessness of all systems of gambling, is the characteristic and decisive property common to all sequences of observations or mass phenomena which form the proper subject of the probability calculus. [...] Everybody who has been to Monte Carlo, or who has read descriptions of a gambling bank, knows how many 'absolutely safe' gambling systems, sometimes of an enormously complicated character, have been invented and tried out by gamblers; and new systems are still being suggested every day. The authors of such systems have all, sooner or later, had the sad experience of finding out that no system is able to improve their chances of winning in the long run, i.e., to affect the relative frequencies with which different colours or numbers appear in a sequence selected from the total sequence of the game. This experience forms the total basis of our definition of probability [1941, pp. 24-25].

Von Mises' second law is thus known as the *law of excluded gambling strategy*, or as the *principle of the impossibility of a gambling system*:

> **Law of excluded gambling strategy**: if we select, according to some rule, a subcollection of an empirical collective—while having at our disposal, at each step, only the knowledge of the results of the previous trials—then the relative frequency of an event in this subcollection is approximately the same as its relative frequency in the original collective.

Having the law of stability of statistical frequencies and the law of excluded gambling strategy at hand, the next step consists in formulating a mathematical theory of empirical collectives.

### 2.1.2 The axioms of probability

According to von Mises, to arrive at a satisfactory probability calculus one must abstract from the incidental properties of empirical collectives. For instance, when dealing with coin tosses, the fact that the coin is being flipped on a Monday afternoon while the birds are chirping is, arguably, inessential to the formulation of a mathematically rigorous theory of probability. Another property which renders empirical collectives mathematically less tractable is their intrinsic finiteness (which is due to the fact that we can only ever carry out a finite number of experiments). For instance, formally characterising the notion of having 'approximately the same' relative frequency of heads in the context of two finite series of coin tosses is a very thorny issue.

To deal with problems of this sort, von Mises introduces what Gilles calls *mathematical collectives* [2000]: given a set of possible outcomes or sample space, a mathematical collective is simply an infinite sequence of elements from the sample space satisfying the mathematical counterparts of the law of stability of statistical

frequencies and the law of excluded gambling strategy. Such mathematical counterparts are known as the *axiom of convergence* (or the *limit axiom*) and the *axiom of randomness*, respectively.

To explain what von Mises' axioms postulate, let us focus once again on a paradigmatic game of chance: coin tossing[10]. In this context, our sample space is the set $\{0,1\}$, corresponding to the two possible outcomes of a coin-tossing experiment: heads, or 0, and tails, or 1. We wish to define the probability of each subset of $\{0,1\}$ (in modern parlance, we take our algebra of events to coincide with the power-set algebra). Now, each sequence in $\{0,1\}^\omega$ may be taken to represent the consecutive outcomes of an infinite coin-tossing experiment[11]. In this setting, given some $X \in \{0,1\}^\omega$, let $\#\mathsf{zeroes}(X \upharpoonright n)$ denote the number of 0's in the first $n$ bits of $X$. Then, following Li and Vitányi [1993/1997/2008], the axiom of convergence can be stated as follows:

> **Axiom of convergence:** if $X \in \{0,1\}^\omega$ is a collective, then, the limiting relative frequency
>
> $$\lim_{n\to\infty} \frac{\#\mathsf{zeroes}(X \upharpoonright n)}{n} = p$$
>
> exists, for some $0 < p < 1$.

The axiom of randomness is more difficult to formulate. To this end, von Mises devises the notion of *insensitivity to admissible place selections*.

Let us begin by making more precise the notion of a *place selection rule*. A selection rule is a (possibly partial) function[12] $s : \{0,1\}^* \to \{\mathsf{select}, \mathsf{scan}\}$ which may be thought of as a gambling system: given the prefix of some infinite sequence as input, it selects the next bit of that sequence to be bet on (without specifying how much should be bet of whether the bet should be placed on 0 or on 1).

For all $\sigma \in \{0,1\}^*$ and selection rule $s$, $s[\sigma]$ will denote the substring of $\sigma$ consisting of all bits progressively selected by $s$ from $\sigma$. For example, if $\sigma = 10010$ and $s$ selects the first, the third and the fifth bit of $\sigma$, while simply scanning the two remaining bits, then $s[\sigma] = 100$. It then follows that, given $X \in \{0,1\}^\omega$, the sequence $(s[X \upharpoonright n])_{n\in\mathbb{N}}$ is non-decreasing with respect to $\sqsubseteq$: that is, for any two initial segments $\sigma, \tau$ of $X$, if $\sigma \sqsubseteq \tau$, then $s[\sigma] \sqsubseteq s[\tau]$, as well. If $s$ selects only finitely many

[10] For a characterisation of von Mises' axioms in their full generality, see, for instance, [van Lambalgen, 1987a].

[11] It should be noted that this notation may be slightly misleading: according to van Lambalgen [1987a], von Mises, being a constructivist, rejects the notion of a *completed infinity*; hence, in his work, sequences should be thought of as being on a par with Brouwer's *free choice sequences* [1918] rather than as elements of Cantor space.

[12] In his original paper on collectives [1919], von Mises does not specify that selection rules should be *functions*, possibly due to his constructivist views. However, he later on settles for the definition given above.

bits from $X$, then we let $s[X]$ denote the finite string consisting of such selected bits. Otherwise, $s[X]$ will denote the infinite sequence of bits selected from $X$ by $s$. If $s$ is a partial selection rule that is undefined for some bit of $X$, then $s[X]$ is taken to be undefined, too.

According to von Mises, a selection rule is *admissible* if the decision of whether to select the $n$-th digit from the original sequence $X$ *depends* on the number $n$ and on the outcomes $X(1), X(2), ..., X(n-1)$ of the $(n-1)$ preceding trials, but not on the outcome of the $n$-th or any subsequent trial[13]. This notion allows us to state the following axiom:

> **Axiom of randomness:** let $X \in \{0,1\}^\omega$ be a sequence satisfying the axiom of convergence, as witnessed by the limiting relative frequency $p$. If $X$ is a collective, then, for any admissible place selection rule $s : \{0,1\}^* \to \{\mathsf{select}, \mathsf{scan}\}$ such that $s[X]$ is infinite, we have that
>
> $$\lim_{n\to\infty}\left(\frac{\#\mathsf{zeroes}(s[X] \upharpoonright n)}{n}\right) = p.$$

So, according to von Mises, a sequence $X$ is random if and only if no admissible strategy for selecting bits from $X$ can induce an infinite subsequence that allows odds for gambling different from those allowed by a subsequence selected arbitrarily (e.g., by making one's decision on the basis of coin tosses).

Von Mises uses the term 'chance' to refer to the limiting relative frequency of some event in a sequence that satisfies his axiom of convergence. The notion of 'probability', on the other hand, is earmarked for the limiting relative frequency of events within mathematical collectives—i.e., those sequences which also satisfy the axiom of randomness. In particular, the set of all limiting relative frequencies within a collective is called its *probability distribution*.

It is then clear that von Mises' goal is not to *bridge* the concept of frequency and that of probability: he wants to explicitly *define* probabilities in terms of frequencies within collectives. As von Mises pithily puts it: "First the collective, then the probability" [1928/1961, p. 18]. Collectives, in turn, are taken as primitives—although one may also say that the axiom of convergence and the axiom of randomness provide an *implicit* definition of collectives, just like Kolmogorov's axioms may be said to offer an implicit definition of probability.

In von Mises' view, "[a]ll problems of the theory of probability consist in deriving, according to certain rules, new collectives from given ones and calculating the distributions of these new collectives" [Li and Vitányi, 1993/1997/2008, p. 51]. Observations are in fact formalised in terms of probabilities, and these give rise to predictions, which are themselves probabilities. Place selection rules induce a

[13] This definition of admissibility is rather vague. Unfortunately, von Mises never attempted to make it more precise—a fact which, as we shall see in Section 2.2, has prompted much criticism.

type of transformation on collectives. In addition, von Mises specifies three different operations on collectives—*mixing, partition*, and *combination*[14]. These operations are collective-preserving, a fact that can be derived from the axiom of convergence and the axiom of randomness. Moreover, with these operations at hand, one can derive (i) Kolmogorov's axioms[15], (ii) the multiplication rule and (iii) the formula for conditional probability. So, in von Mises' theory, the axiom of randomness guarantees that certain laws of probability theory are satisfied by the limiting relative frequencies in collectives [van Lambalgen, 1987a].

## 2.2 Objections to von Mises' definition of randomness

There are two kinds of criticisms that may be levelled against von Mises' theory: on the one hand, one may be dissatisfied with von Mises' definition of randomness; on the other hand, one may have more general worries regarding von Mises' commitment to the frequency interpretation of probability. In what follows, we will consider both sorts of objections.

On the randomness front, von Mises' approach seems to be a bit of a sitting duck: by and large, his account of randomness is regarded as being both flawed and superfluous to the foundations of probability theory. However, many of its alleged shortcomings have been extensively defended by van Lambalgen [1987a], who claims that the downfall of von Mises' definition of randomness may be traced back to the triumph of intuitions concerning the nature of probability—and, more in general, concerning the foundations of mathematics—antithetical to those of von Mises. As maintained by van Lambalgen, the most popular objections to von Mises' characterisation of randomness are in fact concealed attacks at his views on the interpretation of probability. In particular, van Lambalgen contends that von Mises' critics seem to favour some form or another of the propensity interpretation of probability (additionally, such critics work within the framework of classical mathematics, while von Mises leaned towards constructivism).

The goal of this section is to assess the tenability of von Mises' definition of randomness. Then, in Section 2.3, we will address some more general criticisms against his frequentist interpretation of probability.

### 2.2.1 Do collectives exist?

An aspect of von Mises' work which has been the object of much criticism is the notion of *admissibility* of place selection rules. Von Mises deliberately leaves this concept imprecise. Of course, it is possible to concoct several examples of admissible selection rules: for instance, 'select the $n$-th digit of a sequence if and only if $n$ is a

[14] See [van Lambalgen, 1987a] for a rigorous characterisation of these operations and [Childers, 2013, § 1.2.4] for a humourous one.

[15] Up to finite additivity.

prime number' or 'select the $n$-th digit of a sequence if and only if $n$ is even' or 'select a digit if and only if it is preceded by the block 001', and so forth. However, a handful of examples does not suffice to arrive at a general characterisation of admissibility.

Von Mises' sole explicit pronouncement concerning admissibility consists in stating that, while playing against a sequence, one is allowed to employ only selection rules that decide whether to select a given bit independently of the value of that particular bit. Unfortunately, this idea, although rather intuitive in the context of actual betting scenarios, does not translate in any obvious way into a precise mathematical definition.

This ambiguity is precisely what prompts Erich Kamke, a German mathematician, to accuse von Mises' notion of collective of being vacuous [1933]. Kamke's argument proceeds as follows: suppose that $X \in \{0,1\}^\omega$ is a sequence satisfying von Mises' axiom of convergence, so that both 0 and 1 occur infinitely often in it. Let $p_1$ denote the limiting relative frequency of 1 in $X$. Now, consider the collection of all increasing infinite sequences of natural numbers. Such a collection exists independently of $X$. Moreover, within this collection, one will find the sequence $n_1, n_2, n_3, ...$ corresponding to all positions in $X$ whose value is 1. Then, define the following place selection rule: retain the $n$-th bit of $X$ if and only if $n$ is one of the numbers that appear in the sequence $n_1, n_2, n_3, ...$ This function selects from $X$ a subsequence that consists only of 1's. So, in this subsequence, the limiting relative frequency of 1 is different from $p_1$, which violates von Mises' axiom of randomness. Since $X$ was chosen arbitrarily, this argument purports to show that there exist no collectives.

A first point worth mentioning is that it is not entirely clear whether Kamke's selection rule indeed chooses which bits of $X$ to retain independently of their value (in fact, it seems to fall short of this requirement, albeit in a rather roundabout way). As remarked by van Lambalgen, it also appears that Kamke's argument fails to take into account von Mises' constructivist views about mathematics. Place selection rules which are not explicitly constructible cannot be deemed admissible by von Mises:

> Kamke speaks as a set theorist: the set of all infinite binary sequences exists 'out there', together with all its elements, some of which are [collectives]. Hence the set $\{n \in \mathbb{N} : X(n) = 1\}$ is available for admissible place selection in much the same sense as is the set of primes. Von Mises, on the other hand, considers [collectives] to be new objects which, like choice sequences, are not pre-existent; hence $\{n \in \mathbb{N} : X(n) = 1\}$ is not available. For him, $n \mapsto X(n)$ is not a legitimate mathematical function; functions are objects which have been constructed [van Lambalgen, 1987a, pp. 28-29].

Of course, this observation alone does not suffice to exonerate von Mises' theory from the accusation of being vacuous: calling for a departure from classical mathematics could hardly count as a watertight counter-argument. So, von Mises is still

left in the tricky position of having to provide a proof of the existence of sequences satisfying his axioms. In 1937, Abraham Wald comes to the rescue by proving that, for any countable collection $\mathcal{S}$ of 'constructive'[16] place selection rules and for any $p \in (0,1)$, there is at least one[17] sequence $X \in \{0,1\}^\omega$ which satisfies von Mises' axioms with respect to $\mathcal{S}$ and with limiting relative frequency $p$ [1937].

Wald contends that restricting one's attention to countably many selection rules is a rather weak requirement. Being sympathetic towards von Mises' constructivism, he reasons that one should only allow functions which can be 'given by a mathematical law'. Mathematical laws are formulated within a system of formal logic, and logical systems can only involve countably many symbols; in turn, these symbols can be combined to construct at most countably many formulas. Hence, it does not make sense to consider more than countably many place selection functions.

Wald's result induces von Mises to abandon the project of singling out a specific class of admissible place selections in favour of a more pragmatic approach: for each problem in the probability calculus that one may have to solve, one should choose the countable collection of place selections which ensures that the calculations necessary to solve that problem can indeed be carried out. Of course, this solution is rather dubious, for it is not clear that pinpointing the countable family of place selections that are appropriate for one's experiment is always such a self-evident task. Yet, von Mises seems to think that this predicament simply does not pertain to probability theory: assessing the correctness of a given choice of selection functions falls within the business of a theory of inductive inference; probability theory, however, is not concerned with inductive reasoning, but with mass phenomena and repetitive events. We will return to this issue in Section 2.3, where we will appraise von Mises' strict frequentism.

### 2.2.2 Stochasticity and Ville's Theorem

Possibly the most notable attempt at demarcating the class of admissible selection rules is due to Alonzo Church [1940], one of the fathers of computability theory.

Church criticises Wald's appeal to formal systems due to the arbitrariness of any choice of language:

> [Wald's interpretation of gambling systems] is unavoidably relative to the

[16] Wald's concept of constructivity is a rather informal one: it is based on an unspecified notion of *procedure* for successively determining the values of the digits of a sequence in a finite number of steps.

[17] In fact, as noted by Shafer in the introduction to his translation of a passage from Jean Ville's *Étude critique de la notion de collectif*,

> The set of sequences that do qualify [as collectives] have $p$-measure one. [...] This means that this set has probability one with respect to the probability distribution for $X(1), X(2), ...$ obtained by assuming that the $X(n)$ are independent random variables, each equal to 1 with probability $p$ and 0 with probability $1-p$ [1939/2005, p. 3].

> choice of the particular system $\mathcal{L}$ and thus has an element of arbitrariness which is artificial. If used within the system $\mathcal{L}$, it requires the presence in $\mathcal{L}$ of the semantical relation of *denotation* (known to be problematical on account of the Richard paradox[18]). If it is used outside of $\mathcal{L}$, it becomes necessary to say more exactly what is meant by 'definable in $\mathcal{L}$', and the questions of consistency and completeness of $\mathcal{L}$ are likely to be raised in a peculiarly uncomfortable way [1940, p. 135].

According to Church, an admissible selection rule should instead be defined in terms of an effective algorithm, because a gambling system is nothing but a rule telling us, at each step and in a finite amount of time, whether to bet or not.

> To a player who would beat the wheel at roulette, a system is unusable which corresponds to a mathematical function known to exist but not given by explicit definition; and even the explicit definition is of no use unless it provides a means of calculating the particular values of the function. [...] Thus a [gambling system] should be represented mathematically, not as a function, or even as a definition of a function, but as an effective algorithm for the calculation of the values of a function [1940, p. 133.].

So, the set of admissible selection rules is identified by Church with that of *computable* selection rules. Since this latter collection is countable, Wald's theorem ensures the existence of collectives that are invariant under selection rules in this class.

The notion emerging from this restriction is known as *Church stochasticity*.

**Definition 2.2.1** (Church stochasticity)**.** Let $X \in \{0,1\}^\omega$. Then, $X$ is said to be *Church stochastic* if and only if

$$\lim_{n\to\infty}\left(\frac{\#\mathsf{zeroes}(s[X]\restriction n)}{n}\right) = \frac{1}{2} \tag{✠}$$

[18] Richard's paradox, due to the French mathematician Jules Richard [1905], goes as follows. We know that certain linguistic expressions (in, say, the English language) define real numbers unambiguously, while other natural language expressions do not. For instance, the expression 'The real number whose integer part is 17 and whose $n$-th decimal place is 0 for even $n$'s and 1 for odd $n$'s' uniquely defines the real number $17.1010101... = \frac{1693}{99}$. The expression 'Lewis Carroll is Charles Dodgson', on the other hand, does not define any real number. Now, consider the infinite list of English expressions of finite length which define real numbers unambiguously. Arrange this list by length and then order it lexicographically, so that we obtain an infinite list $r_1, r_2, r_3, ...$ of real numbers that are enumerated in a canonical way. We can then define a new real number $r$ as follows: let the integer part of $r$ be 0, and its $n$-th decimal place be 1 if the $n$-th decimal place of $r_n$ is *not* 1, and 2 otherwise. This definition is clearly a natural language expression which unambiguously defines $r$. Hence, $r$ must be one of the $r_n$'s on the list described above. However, $r$ was constructed in such a way as to differ from each of the $r_n$'s, which leads to a contradiction. (This reasoning is clearly reminiscent of Cantor's famous diagonal argument.)

for all total computable selection rules $s$ such that $s[X]$ is infinite[19].

By requiring selection rules to be partial computable rather than computable functions, one can obtain the stronger notion of von Mises-Wald-Church stochasticity; moreover, a concept weaker than Church stochasticity may also be defined.

**Definition 2.2.2** (von Mises-Wald-Church stochasticity)**.** Let $X \in \{0,1\}^\omega$. Then,

(a) $X$ is said to be *von Mises-Wald-Church stochastic* if and only if $s[X]$ satisfies Equation (✠) for all partial computable selection rules $s$ such that $s[X]$ is infinite.

(b) $X$ is said to be *weakly Church stochastic* if and only if $s[X]$ satisfies Equation (✠) for all total computable selection rules $s$ such that $s[Y]$ is infinite for *all* $Y \in \{0,1\}^\omega$.

While von Mises-Wald-Church stochasticity and Church stochasticity are well-established concepts and have been extensively studied, weak Church stochasticity was recently introduced by Stijn Vermeeren in his doctoral dissertation [2013]. The significance of this notion will become apparent in § 3.3.3, where we will see that weak Church stochasticity, as opposed to von Mises-Wald-Church stochasticity and Church stochasticity, is implied by Schnorr randomness (a notion of algorithmic randomness defined by Schnorr in [1971a]).

The reason why we use the term 'Church stochasticity' rather than 'Church randomness' is a theorem due to the French mathematician Jean Ville known as *Ville's Theorem*, which is generally taken to show that von Mises' theory of collectives is too weak to give rise to proper randomness notions. In what follows, we will illustrate Ville's result (Theorem 2.2.3 below) and discuss to what extent it thwarts von Mises' enterprise.

Before we state Ville's Theorem, it should be noted that Ville's approach to probability theory is rather different from that of von Mises: in line with the school of French probabilists of which he is a member, Ville believes that the probability calculus makes contact with the world only by making predictions with probability near or equal to one. An example of such predictions is the law of large numbers, which von Mises' collectives indeed satisfy. However, do collectives also satisfy all other probability-one predictions made by probability theory? Ville's result provides a negative answer to this question: no matter what countable collection of selection rules one picks[20], there exists a sequence which fails to satisfy the law of the iterated logarithm—which, in the context of the uniform distribution, states that the relative frequency of 0 in an infinite sequence should, with probability one, oscillate above and below $\frac{1}{2}$ while converging to $\frac{1}{2}$.

[19] Clearly, the selection rule $s$ given by $s(\sigma) = \mathsf{select}$ for all $\sigma \in \{0,1\}^*$ is total computable and such that $s[X] = X$. This is why, in Definition 2.2.1, one need not add a separate clause requiring the existence of the limiting relative frequency of 0 and 1 in $X$. Also note that, in Definition 2.2.1, the value $\frac{1}{2}$ could be replaced by any other $p \in (0,1)$.

[20] So, the collection of computable selection rules, as well.

**Theorem 2.2.3** (Ville [1939])**.** *Let $\mathcal{S}$ be a countable collection of (possibly partial) selection rules. Then, there exists a sequence $X \in \{0,1\}^\omega$ such that*

(a) $\lim_{n\to\infty}\left(\frac{\#\mathsf{zeroes}(X \upharpoonright n)}{n}\right) = \frac{1}{2}$;

(b) $\lim_{n\to\infty}\left(\frac{\#\mathsf{zeroes}(s[X] \upharpoonright n)}{n}\right) = \frac{1}{2}$ *for every $s \in \mathcal{S}$ such that $s[X]$ is infinite;*

(c) $\frac{\#\mathsf{zeroes}(X \upharpoonright n)}{n} \geq \frac{1}{2}$ *for all $n \in \mathbb{N}$.*

Ville's Theorem shows that there is a collective in which the relative frequency of 0 converges to $\frac{1}{2}$ from above. Hence, frequency convergence from above constitutes a kind of systematicity (i.e., a measure-zero property) that von Mises' definition of randomness is not capable of avoiding. So, Ville's Theorem is usually employed to argue that von Mises' theory is too weak: collectives are vulnerable to Ville's counterexample because there are regularities which cannot be uncovered by merely considering the convergence of the relative frequencies of zeroes and ones in a sequence.

In view of Theorem 2.2.3, one may further speculate that, in order to prevent potential analogous results involving some other measure-one statistical property, a satisfactory notion of randomness should display all asymptotic regularities proved by measure-theoretic methods—i.e., it should satisfy all measure-one statistical properties, and not only the law of large numbers and the law of the iterated logarithm[21].

Ville's own diagnosis is that von Mises' characterisation of gambling systems is not general enough. To solve this problem, he suggests an alternative based on the notion of a betting strategy—that is, a finitely describable rule which specifies whether and how much to bet at any turn in the game, given the outcomes of the previous turns. More precisely, Ville introduces a type of functions which he calls *martingales*[22].

[21] As we shall see, intuitions of this sort lie at the heart of the so-called measure-theoretic typicality approach to algorithmic randomness (Section 3.2).

[22] The kind of martingales employed in algorithmic randomness are but a particular case of the broader definition of martingales from probability theory. In probability theory, a martingale is defined, in full generality, as a sequence $X_0, X_1, ...$ of real-valued random variables (possibly taking negative values) such that, for all $n \in \mathbb{N}$, $\mathbb{E}[X_{n+1} | X_0, X_1, ..., X_n] = X_n$—i.e., the expectation of each random variable conditional on the observed values of the previous ones equals the observed value of the immediately preceding random variable. To avoid confusion, we shall call such a sequence a *martingale process* (see, for instance, [Hitchcock and Lutz, 2006] or [Downey and Hirschfeldt, 2010, § 6.3.4] for a comparison of martingales and martingale processes). It should also be noted that, in the context of games of chance, the word 'martingale' has been used for centuries to designate the betting strategy that doubles one's bet after every loss [Bienvenu et al., 2009; Mansuy, 2009]. Here, however, in accordance with Ville's definition, by 'martingale' we shall mean a more general notion of betting strategy.

Martingales formalise the capital processes emerging from betting strategies which satisfy the following condition: at each stage of the game, the gambler may bet only a fraction $0 \leq \rho \leq 1$ of her current capital on the next bit of some sequence being a 0 (or a 1). This ensures that her capital will remain non-negative no matter how the trial comes out.

Formally, a martingale function is defined as follows.

**Definition 2.2.4** (Martingale)**.** Let $\mathbb{R}^{\geq 0}$ denote the set of non-negative real numbers.

(1) A *martingale* is a function $d : \{0,1\}^* \to \mathbb{R}^{\geq 0}$ which satisfies, for all $\sigma \in \{0,1\}^*$, the averaging condition $2 \cdot d(\sigma) = d(\sigma 0) + d(\sigma 1)$.

(2) A martingale $d$ is said to be *normed* if the gambler's starting capital $d(\varepsilon)$ is 1.

(3) A *supermartingale* is a function $d : \{0,1\}^* \to \mathbb{R}^{\geq 0}$ such that, for all $\sigma \in \{0,1\}^*$, $2 \cdot d(\sigma) \geq d(\sigma 0) + d(\sigma 1)$.

The value $d(\sigma)$ of a martingale $d$ stands for the capital attained by a gambler after betting on the bits of $\sigma \in \{0,1\}^*$ according to the betting strategy underlying $d$. The requirement that $2 \cdot d(\sigma) = d(\sigma 0) + d(\sigma 1)$ is generally referred to as the *fairness condition*, for it ensures that the amount of money gained from an outcome of 0 is the same that would be lost from an outcome of 1. In the case of supermartingales, the fairness condition is relaxed by allowing the gambler to waste some money along the way.

It is then possible to formally specify what it means for a martingale—and, thus, a betting strategy—to be successful.

**Definition 2.2.5** (Martingale success)**.** Let $d$ be a martingale and $X \in \{0,1\}^\omega$. Then, $d$ is said to *succeed* on $X$ if and only if $\limsup_{n\to\infty} d(X \upharpoonright n) = \infty$. Similarly, $d$ is said to succeed on $\mathcal{X} \subseteq \{0,1\}^\omega$ if and only if it succeeds on each $X \in \mathcal{X}$.

So, a martingale is successful if it can make a gambler unboundedly rich. Then, according to Ville, randomness should still be equated with the impossibility of a successful gambling system, but where the gambling system in question is a martingale function rather than a selection rule. This is because, with the above definitions at hand, Ville is able to show that, given a sequence $X \in \{0,1\}^\omega$ which satisfies conditions (a), (b) and (c) of Theorem 2.2.3, one can construct a martingale that succeeds on $X$.

Let us consider a simple construction to illustrate Ville's point. First, let $\delta_n = \#\mathsf{zeroes}(X \upharpoonright n) - \#\mathsf{ones}(X \upharpoonright n)$. By condition (c) of Theorem 2.2.3, we have that $\delta_n \geq 0$ for all $n \in \mathbb{N}$. Thus, $\liminf_{n\to\infty} \delta_n \geq 0$, which means that either $\liminf_{n\to\infty} \delta_n = \infty$ or $\liminf_{n\to\infty} \delta_n = \ell$, for some non-negative integer $\ell$. If $\liminf_{n\to\infty} \delta_n = \infty$, then a gambler playing against $X$ can become infinitely rich by always betting a fixed amount, not exceeding her starting capital, that the next bit will be 0. Since $\delta_n \geq 0$ for all $n \in \mathbb{N}$,

the gambler will never lose all of her capital. Moreover, the fact that $\liminf_{n\to\infty} \delta_n = \infty$ ensures that, eventually, she will gain unbounded capital. If $\liminf_{n\to\infty} \delta_n = \ell$, on the other hand, we have that $\delta_n < \ell$ for finitely many $n$ only. So, there is some $N$ such that, for all $n \geq N$, $\delta_n \geq \ell$ and, for infinitely many $n \geq N$, $\delta_n = \ell$. In this case, a gambler playing against $X$ can become infinitely rich by waiting until stage $N$ has passed and $\delta_n = \ell$: since $X(n+1)$ has to be a 0, the gambler can bet on it risk-free. It should be noted that the martingale described for the case in which $\liminf_{n\to\infty} \delta_n = \infty$ works on any sequence $Y$ with $\liminf_{n\to\infty} \delta'_n = \infty$ (where $\delta'_n = \#\mathsf{zeroes}(Y \upharpoonright n) - \#\mathsf{ones}(Y \upharpoonright n)$). On the other hand, when the liminf is finite, we have a countable collection of martingales $(d_{N,\ell})_{N,\ell\in\mathbb{N}}$—one for each pair $N, \ell$. Even so, this countable collection of martingales can be combined into one single martingale $\sum_{N,\ell} \alpha_{N,\ell} \cdot d_{N,\ell}$, where the $\alpha_{N,\ell}$'s are positive reals adding up to one, and $\sum_{N,\ell} \alpha_{N,\ell} \cdot d_{N,\ell}$ succeeds on a sequence whenever one of the $d_{N,\ell}$'s does.

By an argument analogous to the one above, Ville concludes that von Mises' definition of randomness is too permissive: it counts as random sequences which possess regularities that can be successfully exploited by a bettor. Hence, randomness à la von Mises does not really coincide with the impossibility of a gambling system.

Another famous result by Ville shows that the notion of martingale success is tightly connected with the property of being a measure zero set.

**Theorem 2.2.6** (Ville [1939])**.** *For any class $\mathcal{N} \subseteq \{0,1\}^\omega$, $\mathcal{N}$ has Lebesgue measure zero if and only if there exists a martingale which succeeds on $\mathcal{N}$.*

Theorem 2.2.6 vindicates the intuition that satisfying all measure-one statistical properties (such as the strong law of large numbers or the law of the iterated logarithm) does correspond to the impossibility of devising a successful gambling strategy.

Ville's claim that von Mises' gambling systems should be replaced by martingales may be countered as follows. First of all, as argued by van Lambalgen, one should realise that von Mises' objective never was to model infinite sequences for which no successful gambling strategy exists per se. Although von Mises might have mistakenly regarded his requirement of invariance under admissible place selection functions as being sufficient to single out all and only those sequences whose sole regularity is their limiting relative frequency, his ultimate objective was merely to formalise strict frequentism, while also being able to recover Kolmogorov's axiomatisation of probability. In other words, von Mises was after a minimal notion of 'randomness' which would suffice to formally validate the frequency interpretation of probability.

> Von Mises' rules are set up so that they preserve the frequency interpretation; this no longer holds for the limiting operations of measure theory. [T]he theorem [of the iterated logarithm] does not have a frequency interpretation (in the space of infinite binary sequences). To be

> more precise: von Mises distinguishes between measure theoretic and probabilistic derivations. [The law of the iterated logarithm], as a statement about infinite sequences of trials, is not (probabilistically) derivable using operations such as place selections, although it is of course measure theoretically derivable using properties of the infinite product measure (essentially the Borel-Cantelli lemmas) [van Lambalgen, 1996, p. 16].

So, in view of von Mises' goal, one should ask: in spite of Ville's Theorem, are von Mises' axioms indeed adequate as a theory of frequentist probability? In his doctoral dissertation [1987a], van Lambalgen convincingly argues that, modulo being able to characterise the notion of admissibility for place selections in a satisfactory manner, von Mises' theory is indeed successful[23]. We will not repeat those arguments here, for they are not central to the more general point we wish to stress: namely, that Ville's argument, which is often depicted as having dealt a deathblow to von Mises' theory, only succeeds in showing that von Mises' notion of randomness is not an adequate formalisation of the concept of unpredictability for infinite binary sequences, which is not a concept von Mises was after in the first place.

Of course, this is not to say that Ville's result is of no significance. For one, as we shall see in Section 3.3, Ville's martingale functions and the concept of unpredictability (understood as the failure of a martingale to accrue unbounded capital) play a fundamental role in the field of algorithmic randomness. Yet, in passing from von Mises' characterisation of randomness to the algorithmic approach discussed in Chapter 3, the connections between randomness and probability theory become much murkier. The goal ceases to be finding an empirically sound basis for probability theory, and the focus shifts towards an approach to randomness that takes probability (in its measure-theoretic formulation) for granted.

The upshot of this discussion appears to be that, provided that one agrees that von Mises' notion of randomness (or stochasticity) indeed yields a satisfactory bridge between frequentism and Kolmogorov's probability calculus, the debate over the tenability and usefulness of von Mises' theory of collectives should instead focus on putting under a microscope von Mises' frequentist interpretation of probability.

[23] Van Lambalgen also claims that a frequentist interpretation of probability in fact forces collective-like properties:

> [A]nybody who believes in the frequency interpretation and in the validity of the usual rules for probability is bound to believe in [collectives]. That is, not necessarily in the idealised, infinite [collectives] as they occur in von Mises' axioms, but rather as finite approximations to these. In other words, [*collectives*] *are a necessary consequence of the frequency interpretation* [1987a, p. 31].

By means of an example, van Lambalgen argues that if probability is taken to mean limiting relative frequency, and if one endorses Kolmogorov's axiom together with the product rule and the rule for conditional probability, then an infinite sequence of experiments will be invariant under place selection rules. For a detailed discussion of this argument, see [van Lambalgen, 1987a, § 2.4.1 and § 2.4.2].

Since von Mises tries to reduce the concept of probability to that of limiting relative frequency, in what follows we will address some objections against his reductionist approach. In presenting these arguments, we do not aspire to conclusively refute von Mises' account: after all, the debate over the foundations of probability is still raging among both philosophers and statisticians. Our more modest goal is to remark that subjectivism seems to often lurk behind the objectivist edifice of von Mises' theory.

## 2.3 Objections to von Mises' frequentist approach

A common objection against strict frequentism (voiced, among others, by Kolmogorov) concerns von Mises' appeal to limits. The problem is that von Mises' limiting relative frequency approach appears to be unsuitable for real-life applications: in every-day life, we only ever deal with finitely many trials, so we can never be sure what the limiting relative frequency of an event is or whether such limiting relative frequency exists at all. As noted by von Mises himself, we never have

> sufficient reasons to believe that the relative frequency of the observed attribute [event] would tend to a fixed limit if the observations were indefinitely continued [1928/1961, p. 15].

Hence, generalising from the law of stability of statistical frequencies to the axiom of convergence may be an unwarranted step altogether.

One may even start to wonder if von Mises' theory of collectives has any empirical significance: in particular, whether it has any bearing on the question of how to ascertain the probability of a given event from finite experimental data, or on the question of how to forecast the outcomes of an experiment given a certain probability assignment.

A standard reply to this objection is that it reflects a failure to understand the purpose of von Mises' (and any other) limiting relative frequency theory. After all, von Mises remarks in several occasions that his foundational project does not have anything to do with inductive or statistical reasoning. His work focuses on formulating a *definition* of probability which can serve as the basis for the applicability of the probability calculus to real-life random phenomena, independently of any epistemic concerns about how to assign probabilities to events from finite observations. In other words, von Mises' project is a purely definitional enterprise, aimed at providing a "coherent idealization which knits together all [the] principles [used by statisticians]" [Hacking, 1965, p. 7]. So, limiting relative frequencies should be understood as a useful idealisation.

In light of this response, let us compare for a moment von Mises' limiting relative frequency interpretation and the subjective interpretation, which views probability theory as the language of inductive reasoning. The difference between these two views, it seems, is not unlike the philosophical disagreement between the

*correspondence theory of truth* and *epistemic theories of truth*[24]. Correspondence theorists attempt to define truth as a relational property involving a special connection (correspondence, conformity, congruence, agreement, etc.) to certain aspects of reality (facts, states of affairs, conditions, situations, etc.). Truth so understood has no bearing on the epistemological question of how we can *know* whether something is true: understanding the notion of truth is taken to be a purely analytical enterprise. Epistemic theories of truth, on the other hand, tend to associate the concept of truth with that of justification or rational acceptability or perhaps instrumental success. The type of definition sought after in this setting is one which would allow inferring from the available evidence that one has good support for the claim that a given proposition is true—and, so, good support for accepting and asserting that proposition. Epistemic truth, then, is roughly equivalent to warranted assertibility.

From the perspective of someone with an epistemic penchant (or, perhaps, from the perspective of a probability theorist or statistician), this comparison may be seen as highlighting the fact that both the correspondence theory of truth and the limiting relative frequency interpretation of probability are guilty of committing the same mistake: they extrapolate from experience idealisations which are excessively abstract and even, perhaps, metaphysically dubious. In the context of truth, one may question the soundness of chasing after a definition of absolute truth, considering that all we can ever conclusively establish is to what extent our theories about the world are predictively or instrumentally successful. In the context of probability theory, on the other hand, one may puzzle over the value of focusing on an abstract definition of probability which does not get us any closer to answering the question of how to ascertain the probability of an event. Especially in view of von Mises' concerns about reducing probability to an observable and measurable quantity, postulating the existence of limiting relative frequencies when all we have at our disposal are finite relative frequencies, which do not warrant any particular choice of limit value, appears to be a rather unavailing (if not wrong) idealisation. A similar point is made by Hacking:

> It is sometimes said that the Euclidean plane or spherical geometry used in surveying involves an idealization. Perhaps this means that surveyors take a measurement, make simplifying assumptions, use Euclid for computing their consequences, and finally assume that these consequences are also, approximately, properties of the world from which the original measurements were taken.
>
> It is true that some of the laws of von Mises' collective apply to frequency in the long run, and that these laws are used in computing new frequencies from old. But it is the laws, and not the infinite collective, which are of use here. Never, in the journals, will one find a statistician using the

[24] For an introduction to the philosophical debate over theories of truth, see, for instance, [Glanzberg, 2014]. For an overview of the correspondence theory of truth, see, for example, [David, 2015].

> peculiar characteristics of a collective in making a statistical inference, whereas surveyors really do use some of the attributes peculiar to the Euclidean plane or spherical geometry. So whatever its interest in its own right, the theory of collectives seems redundant as an idealization in the study of frequency in the long run [1965, pp. 6-7].

These observations raise the question of whether it might be possible to recover Kolmogorov's axiomatisation from finite frequentism, in lieu of the limiting relative frequency interpretation of probability advocated by von Mises. As a matter of fact, considerations of this kind constitute the prime motive behind Kolmogorov's own attempt at providing a finitary version of von Mises' collectives, where probability assignments reflect finite relative frequencies, as they are recorded after *sufficiently long* sequences of trials. In particular, in [1963], Kolmogorov offers an alternative definition of randomness[25] which is clearly inspired by Church's introduction of computable place selection rules: a binary string counts as being random if the relative frequencies of 0 and 1 are *roughly* invariant under *admissible selection algorithms*[26]. In turn, admissible selection algorithms are essentially non-monotonic place selection functions: instead of proceeding linearly along a string, they can extract a substring in any computably specified order, as long as no position is ever considered more than once[27].

In spite of being ontologically more parsimonious, finite frequentism, as a reductionist project, has many problems of its own. For instance, how can we determine whether a sequence of trials is long enough to exhibit the required frequency stability? Moreover, how close to one another should the observed relative frequencies be in order for us to be able to infer the correct probability assignment?

A possible escape route that addresses the objection from empirical significance and some of the issues raised above would consist in adopting the view that probabilities are nothing more than a 'rational' agent's degree of belief in the occurrence of a given event, and the relative frequencies observed in actual trials constitute *evidence* that can and should be employed (together with all other data that the agent has at her disposal) to calibrate her credences. Embracing such a viewpoint, however, would amount to radically departing from von Mises' reductionist and objectivist approach in favour of a form of *constrained subjectivism*, to use Hájek's terminology [2007], where rational subjective probability assignments are required to track as much as possible the observed relative frequencies. This position is encapsulated

[25] This definition of randomness predates Kolmogorov's work on algorithmic complexity [1965], which we will discuss in § 3.1. The algorithmic approach will allow Kolmogorov to connect the notion of randomness more directly to applications, without any detour through relative frequencies.

[26] See [Li and Vitányi, 1993/1997/2008, Chapter 1].

[27] Kolmogorov's definition of randomness based on non-monotonic selection functions is the precursor of Kolmogorov-Loveland randomness and Kolmogorov-Loveland stochasticity, which we will discuss in § 3.3.2 and § 3.3.3, respectively.

by the following principle, due to Hacking[28] [1965], where $Pr$ denotes a probability function meant to represent an agent's subjective credences:

> **Principle of Direct Probability** Let $\mathsf{relfreq}(E)$ denote the observed relative frequency of some event $E$. Then, $Pr(E \,|\, \mathsf{relfreq}(E) = x) \approx x$, for all events $E$ such that $Pr(\mathsf{relfreq}(E) = x) > 0$.

Such a change of perspective would allow one to remain within a finitistic setting, while also being able to dodge some of the criticisms that finite frequentism is liable to[29]. Yet, taking on a subjectivist viewpoint would also render von Mises' attempt at confining the range of applicability of probability theory to truly random phenomena superfluous, for subjective probabilities can be assigned to any event, random or non-random, as long as such probabilities measure a rational agent's uncertainty. Hence, a subjectivist account of randomness would by its very own nature be disconnected from the question of how probabilities are to be defined. As we shall see, a subjectivist interpretation of probability appears to be more compatible with the algorithmic approach to randomness that we will discuss in Chapter 3.

On a separate note, it appears that such an epistemic and subjectivistic approach may be more consonant with Church's identification of the class of admissible selection functions with the collection of *computable* selection functions: in this setting, computability is taken to mark the boundaries of our cognitive and epistemic limitations, and a sequence is deemed random if no computable selection rule can find any *evidence* for the presence of some underlying pattern which could be exploited to select a biased subsequence.

Another issue facing von Mises' account is what Hájek calls the *reference sequence problem* [2009]. According to von Mises, probabilities are always dependent on a specific collective. This means, in particular, that probabilities hinge on the order in which sampling or experiments are performed. To elucidate this dependency, consider once again the case of coin flipping. When computing the limiting relative frequencies in a collective corresponding to a coin-tossing experiment, one usually takes the temporal ordering to be the most salient one: one simply records the outcomes of the experiments, as those experiments are successively carried out. However, this choice is ultimately arbitrary. For example, imagine performing a coin-tossing experiment in an elevator that constantly moves up and down: one could legitimately choose the up-down spatial dimension to determine the ordering in the collective. But then, how can one ensure that the resulting limiting relative frequencies will be the same, no matter what ordering one picks? From an objectivist viewpoint, this

[28] Hacking's Principle of Direct Probability is analogous to Lewis' *Principal Principle* [1980], according to which subjective probability assignments should track objective chances.

[29] Of course, this is not to say that the subjective interpretation of probability does not face objections of its own. A conclusive appraisal of this debate, however, is beyond the scope of this thesis.

seems to be a serious problem, unless one were able to show that invariance under admissible place selection rules also ensures that permuting a collective never changes the limiting relative frequencies. This issue is even more poignant in the context of random sampling in statistics, where there does not seem to be any obviously 'natural' ordering one can impose. If, for instance, we wanted to test the hypothesis that a given party will win the next elections, how could we make sure that the ordering in the collective corresponding to the sample from the population being tested will not affect the resulting limiting relative frequencies? Of course, von Mises could simply bite the bullet and reiterate that (i) probabilities do depend on collectives (and, consequently, on the given collective's ordering) and that (ii) choosing the relevant collective is a purely pragmatic issue. This is a viable reply, but it seems to render von Mises' approach more agent-relative and subjectivistic than he might have been happy to concede.

These are only a couple of objections that one may raise against von Mises' strict frequentism. There are many more arguments that one could consider against the limiting relative frequency interpretation of probability—and against its finite counterpart, as well. Hájek, for example, famously discusses fifteen arguments against each type of frequentism [1997; 2009] (although not every argument from [2009] is applicable to von Mises' theory).

The moral of our story so far seems to be that some of the criticisms that von Mises' frequentist approach faces may be resolved by abandoning his reductionist approach in favour of a subjectivistic view of probability that takes the observed relative frequencies seriously. However, when passing from an objectivist position to a full-on subjectivist one, the need for a notion of randomness à la von Mises fades away. Hence, the tenability and usefulness of von Mises' definition of randomness ultimately rests on the correctness of strict frequentism, which is not short of problems.

# Chapter 3

# The Many Faces of Algorithmic Randomness

This chapter will be devoted to providing an overview of algorithmic randomness. Although inspired by von Mises' views, the algorithmic approach to defining randomness has considerably distanced itself from the theory of collectives. First of all, the account of probability upon which it is based is very different from that of von Mises and more in line with that of Ville and the French probabilists (§ 2.2.2): as we will see, in algorithmic randomness, the focus tends to be on those special statistical properties which occur 'with measure one'. Moreover, one of the main features of this approach, inherited from Church, is a deep connection with computability theory (hence the name *algorithmic* randomness).

In what follows, we will review the three main existing approaches to algorithmic randomness:

(i) the *incompressibility paradigm*, built on the notion of *Kolmogorov complexity* and pioneered by Solomonoff [1960; 1964], Kolmogorov [1965] and Chaitin [1969];

(ii) the *measure-theoretic typicality paradigm*, launched by Martin-Löf [1966] (and Demuth [1975]) and based on the notion of *effective statistical test*;

(iii) the *unpredictability paradigm*, inaugurated by Ville's work on martingale functions [1939], and then developed by Schnorr [1971a; 1971b] and Levin [1970; 1973] in a computability-theoretic setting.

Unless otherwise specified, our presentation will follow Li and Vitányi's monograph [1993/1997/2008].

In § 3.3.3, we will also provide a martingale-based characterisation of a stochasticity notion called *weak Church stochasticity*, which was introduced by Vermeeren in [2013]. Then, we will present an alternative proof of the fact that weak Church

stochasticity is implied by Schnorr randomness, as opposed to the orthodox stochasticity notions defined in § 2.2.2. To further 'symmetrise' the multitude of implications that obtain between randomness and stochasticity notions, we will also introduce the non-monotonic variant of weak Church stochasticity (*weak Kolmogorov-Loveland stochasticity*), which can be easily seen to be implied by Kolmogorov-Loveland stochasticity and to entail weak Church stochasticity.

## 3.1 Randomness as incompressibility

In the 1960's, Kolmogorov [1965] proposes a definition of randomness for finite strings inspired by Shannon's work on information theory[30]. The basic idea behind this definition is that randomness amounts to irregularity; in other words, randomness is identified with a certain lack of recognisable patterns[31].

In spite of being very intuitive, this identification also prompts the following question: who should be entrusted with deciding what counts as random, given that pattern-detecting skills vary wildly across people? For instance, although (as already remarked in Chapter 1) hardly anybody would judge the string

$$0101010101010101010101010101010101010101$$

to be random, the string

$$1112112111112213122111311222111132132 11$$

does not exhibit any immediately recognisable patterns. Yet, upon further reflection, one may notice that this second string (known as the *Conway sequence* [1986]) is describable by means of a relatively simple rule, too: begin by partitioning it as

$$1,\ 11,\ 21,\ 1211,\ 111221,\ 312211,\ 13112221,\ 1113213211;$$

each block except for the first is generated on the basis of the previous one by reading off the digits of the previous block and counting the number of digits in groups of the same digit. More concretely: the first block, 1, generates the block 'one 1' or 11; the second block, in turn, generates the block 'two 1s' or 21; the third block generates the block 'one 2, then one 1' or 1211; and so forth.

To characterise the notion of 'lack of recognisable patterns' as objectively as possible, Kolmogorov appeals to computability theory and to the Church-Turing Thesis [1936], which postulates that the intuitive notion of 'effective procedure' coincides with that of Turing machine[32]. In particular, in Kolmogorov's framework,

[30] See, for instance, [Shannon, 1948].

[31] Around the same time, Solomonoff [1960; 1964] and Chaitin [1969] independently arrived at the same notion of randomness.

[32] It should be noted that the Church-Turing Thesis cannot be proven. This is because it hinges on intuitions that have no mathematical counterpart, it cannot be formulated as a theorem. Yet, so far, all attempts to disprove it have failed (see [Vitányi, 2009]).

randomness is equated with patternlessness in the eyes of a Turing machine. More specifically, Kolmogorov's great insight is that, if a string is regular, then it should be possible to provide a Turing machine with a simple set of instructions to obtain that string as an output. Conversely, if a string is irregular or random, then all of its descriptions (equivalently, all programmes that generate it) should be very complex.

To make these intuitions precise, what Kolmogorov needs is a way to measure the complexity of strings. From among all possible descriptions of a string, Kolmogorov suggests that the length of a shortest description be taken as a measure of the string's complexity. Then, he defines a string to be random if it is highly incompressible: that is, if all of its descriptions are roughly as long as the string itself[33]. On the other hand, he defines a string to be non-random if it is compressible: i.e., if it has one description shorter than itself.

In what follows, we present Kolmogorov's definition of randomness for finite strings, which is based on the notion of Kolmogorov complexity sketched above, and which has given rise to a field known as *algorithmic information theory*. In Section 3.2, we shall explain how this framework can be extended to provide a characterisation of randomness for infinite sequences, as well.

**Definition 3.1.1** (Plain Kolmogorov complexity)**.** Let $f : \{0,1\}^* \to \{0,1\}^*$ be a partial computable function and $M$ a Turing machine that computes $f$. The *plain Kolmogorov complexity* of a string $\sigma$ with respect to $M$ is given by

$$C_M(\sigma) = \min\{|\tau| \in \mathbb{N} : \tau \in \{0,1\}^* \wedge M(\tau) \downarrow = \sigma\}.$$

If there is no $\tau \in \mathrm{dom}(f)$ such that $M(\tau) \downarrow = \sigma$, then $C_M(\sigma) = \infty$.

Put in (plain) English: the plain Kolmogorov complexity of a string $\sigma$ with respect to a Turing machine $M$ is the minimal length of a programme which makes $M$ compute output $\sigma$ without any additional input.

A string $\sigma$ is then said to be *random* with respect to a Turing machine $M$ if and only if $C_M(\sigma) \geq |\sigma|$.

Clearly, this complexity measure crucially rests on the particular choice of Turing machine. Is it possible to get rid of this dependency? Recall that a *universal* Turing machine is one which can simulate the behaviour of any other Turing machine. Moreover, given a class $\mathcal{C}$ of Turing machines, we have that a Turing machine $M$ is *additively optimal* for $\mathcal{C}$ if

(i) $M \in \mathcal{C}$, and

(ii) for every $N \in \mathcal{C}$, there is a constant $c_{M,N}$ such that, for all $\sigma \in \{0,1\}^*$,

$$C_M(\sigma) \leq C_N(\sigma) + c_{M,N}.$$

[33] Programmes, or descriptions, may be viewed as decompression algorithms.

The following theorem, known as Kolmogorov's *Invariance Theorem*, establishes that there is a proper subset of the collection of universal Turing machines—namely, the one consisting of all additively optimal universal Turing machines—which allows for a general definition of Kolmogorov complexity[34].

**Theorem 3.1.2** (Kolmogorov [1965])**.** *There exists an additively optimal universal Turing machine.*

Fix some universal additively optimal Turing machine $U$: then, $C_U$ is a universal complexity measure. In particular, Theorem 3.1.2 entails that altering one's choice of universal additively optimal Turing machine can only make the complexity measure vary by an additive constant. Thus, we can define the *plain Kolmogorov complexity* of a string $\sigma$ as

$$C(\sigma) = C_U(\sigma).$$

It should be pointed out, however, that Theorem 3.1.2, although striking, does not render the notion of plain Kolmogorov complexity fully independent from one's choice of compression scheme. It is in fact possible, given two strings $\sigma, \tau$ and two additively optimal universal Turing machines $U, V$, to have that $C_U(\sigma) = 0$, $C_V(\tau) = 0$, and yet $\sigma$ is random with respect to $V$ and $\tau$ is random with respect to $U$: then, $C_V(\sigma) \geq |\sigma|$ and $C_U(\tau) \geq |\tau|$.

One may then wonder: is there any (objective) quantity measured via Kolmogorov complexity whose value is not underdetermined by the specific choice of compression scheme? It turns out that Kolmogorov complexity does not hold invariant any property that applies to only finitely many objects; however, it provides a machine-independent characterisation of asymptotic complexity: i.e., all Kolmogorov complexity measures agree on the asymptotic complexity of infinite sequences.

An immediate corollary of Kolmogorov's celebrated Invariance Theorem is that there exists a constant $c$ such that, for all $\sigma \in \{0,1\}^*$, $C(\sigma) \leq |\sigma| + c$. To prove this, one can simply take the Turing machine which computes the identity function.

It is also possible to provide a bound for the complexity of pairs of strings. At first, one may think that, for all $\sigma, \tau \in \{0,1\}^*$, the following should hold for some constant $c$:

$$C(\sigma\tau) \leq C(\sigma) + C(\tau) + c.$$

This is however not the case. The problem is that, by concatenating the descriptions of $\sigma$ and $\tau$, one loses the information of where a description ends and the next one begins. For a machine to be able to correctly parse $\sigma$ and $\tau$, one has to feed it the extra information $|\sigma|$ and $|\tau|$. Then, the following result holds: for all $\sigma, \tau \in \{0,1\}^*$,

$$C(\sigma\tau) \leq C(\sigma) + C(\tau) + 2 \cdot \log\big(\min(C(\sigma), C(\tau))\big).$$

[34] For a proof of the Invariance Theorem, see [Li and Vitányi, 1993/1997/2008, Lemma 2.1.1, p. 104 and Theorem 2.1.1, p. 105].

Crucially, the logarithmic term in the above inequality cannot be done away with[35].

Now, for $k \in \mathbb{N}$, we have that $\sigma \in \{0,1\}^*$ is said to be *$k$-random$_C$* (or *$k$-incompressible$_C$*) if and only if $C(\sigma) \geq |\sigma| - k$. Note that this definition does not allow us to partition the set of finite strings into the random and the non-random ones: $C$ is not a measure of absolute randomness, it is a measure of the *degree of randomness* of a string.

The following important result hinges on intuitions similar to those underlying *Berry's paradox*[36].

**Proposition 3.1.3** (Kolmogorov [1965])**.** *The plain Kolmogorov complexity function $\sigma \mapsto C(\sigma)$ is uncomputable.*

The good news, however, is that $C$ can be computably approximated.

**Proposition 3.1.4** (Kolmogorov [1965])**.** *The function $C$ is* upper semi-computable*: i.e., there is a total computable function $f : \{0,1\}^* \times \mathbb{N} \to \mathbb{N}$ such that, for all $\sigma \in \{0,1\}^*$, the sequence $f(\sigma,0), f(\sigma,1), f(\sigma,2), \ldots$ is monotonically decreasing and converges to $C(\sigma)$.*

As already noted, the plain Kolmogorov complexity of a pair of strings—i.e., when two strings are concatenated—is bounded by the sum of the complexities of the two individual strings plus a logarithmic factor. One may then ask: is it possible to come up with another complexity measure that allows to dispense with this logarithmic factor (i.e., which is *subadditive*)? As it turns out, this is indeed feasible via the notion of *prefix-free Kolmogorov complexity*.

Recall from Section 1.4 that a set $\mathcal{S} \subseteq \{0,1\}^*$ is prefix-free if and only if, if $\sigma, \tau \in \mathcal{S}$, then neither $\sigma \sqsubset \tau$ nor $\tau \sqsubset \sigma$. A function is then said to be prefix-free if it has a prefix-free domain: for any two strings $\sigma, \tau \in \{0,1\}^*$, if $\sigma \sqsubset \tau$, then the function is defined for at most one of these two strings.

**Definition 3.1.5** (Prefix-free Kolmogorov complexity)**.** Let $g : \{0,1\}^* \to \{0,1\}^*$ be a prefix-free partial computable function and $N$ a prefix-free Turing machine which computes $g$. The *prefix-free Kolmogorov complexity* of a string $\sigma$ with respect to $N$ is given by

$$K_N(\sigma) = \min\{|\tau| \in \mathbb{N} : \tau \in \{0,1\}^* \wedge N(\tau) \downarrow = \sigma\}.$$

[35] See [Li and Vitányi, 1993/1997/2008, Example 2.1.5, p. 109 and Example 2.2.3, p. 118].

[36] Berry's paradox is discussed by Russell and Whitehead in their celebrated *Principia Mathematica* [1910/1912/1913], where the authors credit G. G. Berry, a librarian at the Bodleian Library in Oxford. As explained by Li and Vitányi [1993/1997/2008], the paradox goes as follows. Let $\sigma$ be the lexicographically least binary string which cannot be univocally described in less than twenty words. If such a $\sigma$ exists, then we have just managed to describe it using only eighteen words, which contradicts its definition. If such a $\sigma$ does not exist, then all binary strings can be univocally described in less than twenty words, which is of course false.

Just as in the case of the plain Kolmogorov complexity $C$ from Definition 3.1.1, we then have that $\sigma$ is *random* with respect to a prefix-free Turing machine $N$ if and only if $K_N(\sigma) \geq |\sigma|$.

An analogue of Theorem 3.1.2 holds for prefix-free Kolmogorov complexity, as well: there exists an additively optimal universal prefix-free Turing machine. The *prefix-free Kolmogorov complexity* of a string $\sigma$ can then be defined as

$$K(\sigma) = K_V(\sigma),$$

where $V$ is some fixed additively optimal universal prefix-free Turing machine.

With this definition at hand, one can show that there is a constant $c$ such that, for all $\sigma \in \{0,1\}^*$,

$$K(\sigma) \leq C(\sigma) + 2 \cdot \log C(\sigma) + c.$$

The above definition also gives rise to a new randomness notion: for $k \in \mathbb{N}$, $\sigma \in \{0,1\}^*$ is said to be *$k$-random$_K$* (or *$k$-incompressible$_K$*) if and only if $K(\sigma) \geq |\sigma| - k$.

Just like $\sigma \mapsto C(\sigma)$, the function $\sigma \mapsto K(\sigma)$ is uncomputable. However, $K$ is subadditive. For consider some prefix-free partial computable function $g$ such that $\sigma, \tau \in \mathrm{dom}(g)$ and a Turing machine $N$ which computes $g$. Then, the string $\xi$ corresponding to the concatenated pair $\sigma\tau$ can be decoded without extra information: one only need consider all possible splittings of $\xi$ into two adjacent substrings $\sigma'$ and $\tau'$, and then compute $N(\sigma')$ and $N(\tau')$. Since it is prefix-free, $N$ halts only when $\sigma' = \sigma$ and $\tau' = \tau$.

Now that we have at our disposal two notions of randomness for finite strings ($k$-randomness$_C$ and $k$-randomness$_K$), can we employ similar ideas to provide a definition of randomness for infinite sequences? Kolmogorov's initial suggestion was to say that a sequence $X \in \{0,1\}^\omega$ is random if and only if there is a constant $c$ such that $C(X \upharpoonright n) \geq n - c$ for all $n \in \mathbb{N}$. Yet, as shown by Martin-Löf, no sequence satisfies this condition.

**Theorem 3.1.6** (Martin-Löf [1966])**.** *Let $X \in \{0,1\}^\omega$. There exist infinitely many $n \in \mathbb{N}$ such that $C(X \upharpoonright n) \leq n - \log n$.*

Theorem 3.1.6 is known as Martin-Löf's *oscillation theorem*, and it shows that, for any infinite sequence, it is always possible to find an initial segment of low complexity, which renders Kolmogorov's original proposal vacuous.

However, as we will see in § 3.2.1, a satisfactory definition of randomness for infinite sequences which is not affected by Theorem 3.1.6 can be given in terms of prefix-free Kolmogorov complexity[37].

This concludes our review of the incompressibility paradigm. Next, we discuss Martin-Löf's approach to randomness, which is based on the notion of effective statistical tests.

[37] Recently, Miller and Yu managed to prove that randomness for infinite sequences is actually characterisable in terms of plain Kolmogorov complexity, as well [2008].

## 3.2 Randomness as measure-theoretic typicality

The *measure-theoretic typicality* paradigm was introduced by Martin-Löf [1966] in an attempt to extend Kolmogorov's definition of randomness to infinite binary sequences (at the time, Martin-Löf was Kolmogorov's student).

In this setting, a sequence is random if it does not possess any rare properties. More precisely, the basic intuitions behind this approach are that (i) randomness amounts to lawlessness or patternlessness, and (ii) order is an exceptional feature: almost all infinite sequences are random and only a few of them exhibit some orderly pattern that makes them easily characterisable. Hence, the set of non-random sequences should have measure zero, while the set of random sequences should have measure one.

The problem is: how can we choose a measure one subset of Cantor space corresponding to the set of all random sequences in a non-arbitrary manner? At first, one may think that the most obvious way to characterise random sequences is by taking the intersection of all measure one subsets of $\{0,1\}^\omega$. This intuitively appealing idea, however, does not work. Under any reasonable probability measure on Cantor space, the singleton $\{X\}$ of each sequence $X \in \{0,1\}^\omega$ is in fact a measure zero set; hence, given each infinite sequence, the complement of its singleton will have measure one. This means that the intersection of all measure one sets is the empty set.

One must therefore restrict one's attention to measure one subsets of Cantor space which satisfy certain desirable properties. For instance, in line with Ville's reasoning (§ 2.2.2), it seems reasonable to require of random sequences that they obey the strong law of large numbers and the law of the iterated logarithm. Yet, within this setting, it is unclear how many and exactly which measure one stochastic properties one should expect random sequences to satisfy.

In accordance with Kolmogorov's line of research, Martin-Löf's own proposal relies on computability theory: a sequence is said to be random if it satisfies all measure one stochastic properties which can be *effectively* specified. Equivalently, a sequence is deemed random if and only if it cannot be effectively determined to violate a measure one stochastic property. As we will see, this restriction ensures that the set of random sequences has (effective) measure one.

Effectively determining whether a measure one stochastic property has been violated can be thought of as performing a statistical test for randomness—where, given some sequence $X \in \{0,1\}^\omega$, the conjecture (or *null hypothesis*) is that $X$ is a typical outcome. Then, $X$ is categorised as a random sequence if and only if it passes all such performable statistical tests (or randomness tests).

Over the years, Martin-Löf's original definition has inspired a whole variety of algorithmic randomness notions based on the concept of a randomness test. Here, we will consider three distinct types of such tests. The first is Martin-Löf's own concept, which gives rise to Martin-Löf randomness; the second is due to Schnorr and it gives

rise to Schnorr randomness [1971a; 1971b]; the third and last one is due to Kurtz, and it gives rise to Kurtz randomness [1981]. As proven by Schnorr, Martin-Löf randomness is strictly stronger than Schnorr randomness which, in turn, is strictly stronger than Kurtz randomness.

### 3.2.1 Martin-Löf randomness

Part of the rationale behind Martin-Löf's definition of randomness comes from statistical hypothesis testing, which prescribes that a hypothesis be discarded if, upon supposing that the hypothesis is true, one observes a statistically significant outcome according to some pre-specified significance level. As we will see, Martin-Löf focuses on significance levels of the form $2^{-n}$. A sequence $X \in \{0,1\}^\omega$ is rejected at level $2^{-n}$ if and only if there is $m \in \mathbb{N}$ such that we would reject $X \upharpoonright m$ at level $2^{-n}$. On the other hand, a sequence $X$ is rejected if, for every significance level, there is an initial segment of $X$ that we would discard at that level. A Martin-Löf random sequence, then, is one which is *not* rejected at every significance level.

Before we can make these ideas mathematically precise, we need a couple of preliminary definitions.

**Definition 3.2.1** (C.e. open set)**.** An open set $\mathcal{U} \subseteq \{0,1\}^\omega$ is said to be an *effectively open set* or a *computably enumerable open set* (c.e. open set) if and only if there exists a computably enumerable prefix-free sequence $(\sigma_n)_{n\in\mathbb{N}}$ of strings such that $\mathcal{U} = \bigcup \{[\![\sigma_n]\!] \subseteq \{0,1\}^\omega : n \in \mathbb{N}\}$.

**Definition 3.2.2** (Computable sequence of c.e. open sets)**.** A sequence $(\mathcal{U}_n)_{n\in\mathbb{N}}$ of open subsets of $\{0,1\}^\omega$ is said to be a *computable sequence of c.e. open sets* if and only if there exists a sequence $(\mathcal{S}_n)_{n\in\mathbb{N}}$ of subsets of $\{0,1\}^*$ that are computably enumerable uniformly in $n$—i.e., for each $n \in \mathbb{N}$, $\mathcal{S}_n = (\sigma_{n,i})_{i\in\mathbb{N}}$—and such that, for all $n \in \mathbb{N}$, $\mathcal{U}_n = [\![\mathcal{S}_n]\!] = \bigcup \{[\![\sigma_{n,i}]\!] \subseteq \{0,1\}^\omega : i \in \mathbb{N}\}$.

Martin-Löf randomness is then defined as follows.

**Definition 3.2.3** (Martin-Löf randomness)**.** (a) Let $(\mathcal{U}_n)_{n\in\mathbb{N}}$ be a computable sequence of c.e. open sets satisfying, for all $n \in \mathbb{N}$, $\lambda(\mathcal{U}_n) \leq 2^{-n}$. Such a sequence is called a *Martin-Löf test*.

(b) For every Martin-Löf test $(\mathcal{U}_n)_{n\in\mathbb{N}}$, a set $\mathcal{N} \in \wp(\bigcap_{n\in\mathbb{N}} \mathcal{U}_n)$ is called a *Martin-Löf null set*.

(c) A sequence $X \in \{0,1\}^\omega$ is said to be *Martin-Löf random* if and only if there is no Martin-Löf null set $\mathcal{N}$ such that $X \in \mathcal{N}$.

Note that Definition 3.2.3(a) does not require that a Martin-Löf test be composed of nested sets. Given any Martin-Löf test, it is however possible to construct another, nested test as follows. Take a Martin-Löf test $(\mathcal{V}_n)_{n\in\mathbb{N}}$. Set $\mathcal{U}_n = \bigcup_{m>n} \mathcal{V}_m$. Then,

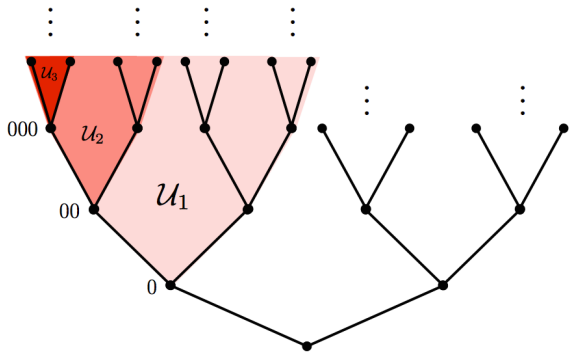



Figure 3.1: Graphical representation of the first three levels of a (nested) Martin-Löf test $(\mathcal{U}_n)_{n\in\mathbb{N}}$, where each level corresponds to a c.e. open set $\mathcal{U}_n$. From [Porter, 2014].

$(\mathcal{U}_n)_{n\in\mathbb{N}}$ is a Martin-Löf test, $\mathcal{U}_0 \supseteq \mathcal{U}_1 \supseteq \mathcal{U}_2 \supseteq \cdots$, and $\bigcap_{n\in\mathbb{N}} \mathcal{V}_n = \bigcap_{n\in\mathbb{N}} \mathcal{U}_n$. Moreover, the value $2^{-n}$ in Definition 3.2.3(a) is arbitrary: it can be shown that replacing it with the value of any other computable function which approaches 0 in the limit still gives rise to Martin-Löf randomness.

One of the most salient features of Martin-Löf randomness is the existence of a *universal* Martin-Löf test: namely, a Martin-Löf test $(\mathcal{V}_n)_{n\in\mathbb{N}}$ such that, for every Martin-Löf test $(\mathcal{U}_n)_{n\in\mathbb{N}}$, $\bigcap_{n\in\mathbb{N}} \mathcal{U}_n \subseteq \bigcap_{n\in\mathbb{N}} \mathcal{V}_n$. This means that there is a *single* effective statistical test which allows for a definition of randomness for individual sequences.

As proven by Schnorr (while he was serving as a referee for a paper by Chaitin [1975]), Martin-Löf randomness has an equivalent characterisation in terms of prefix-free Kolmogorov complexity.

**Theorem 3.2.4** (Schnorr)**.** *Let $X \in \{0,1\}^\omega$. Then, $X$ is Martin-Löf random if and only if there is a constant $c$ such that $K(X \upharpoonright n) \geq n - c$ for all $n \in \mathbb{N}$.*

This striking result—known as *Schnorr's Theorem*—has been taken to show that Martin-Löf randomness is indeed a reasonable notion of algorithmic randomness[38]. We will see in Section 3.3 that Martin-Löf randomness has a natural characterisation within the unpredictability paradigm, as well.

Note that Martin-Löf randomness is also referred to as 1-*randomness* in the algorithmic randomness literature, because Definition 3.2.3 can be generalised so as

[38] As we will see in Section 4.1, Schnorr's Theorem has even been taken as an indication that Martin-Löf randomness truly captures our 'pre-theoretic concept of randomness'.

to obtain a whole hierarchy of $n$-randomness concepts ($n \in \mathbb{N}$), which are all strictly stronger than Martin-Löf's original definition (we will expand on this in Chapter 4).

### 3.2.2 Schnorr randomness

Although Martin-Löf randomness is often described in the literature as the central notion of algorithmic randomness—due to its mathematical elegance and to its robustness (as witnessed by Theorem 3.2.4)—Martin-Löf's definition does not enjoy universal consensus. According to Schnorr, for instance, the concept of a Martin-Löf test is not effective enough. This is because knowing that an infinite sequence belongs to a c.e. open set of small measure is not enough, given Martin-Löf's definition, to effectively predict the bits of that sequence.

We will get a better handle on Schnorr's critique of Martin-Löf randomness in Section 4.1. For the time being, we simply illustrate the alternative notion of Schnorr randomness, introduced by Schnorr in [1971a], in terms of statistical tests.

**Definition 3.2.5** (Schnorr randomness)**.** (a) Let $(\mathcal{U}_n)_{n\in\mathbb{N}}$ be a Martin-Löf test such that the measures $\lambda(\mathcal{U}_n)$ are uniformly computable. Then, $(\mathcal{U}_n)_{n\in\mathbb{N}}$ is a *Schnorr test*.

(b) For every Schnorr test $(\mathcal{U}_n)_{n\in\mathbb{N}}$, a set $\mathcal{N} \in \wp(\bigcap_{n\in\mathbb{N}} \mathcal{U}_n)$ is called a *Schnorr null set*.

(c) A sequence $X \in \{0,1\}^\omega$ is said to be *Schnorr random* if and only if there is no Schnorr null set $\mathcal{N}$ such that $X \in \mathcal{N}$.

A Schnorr test may be equivalently defined as a computable sequence of c.e. open sets $(\mathcal{U}_n)_{n\in\mathbb{N}}$ such that, for all $n \in \mathbb{N}$, $\lambda(\mathcal{U}_n) = 2^{-n}$. Just as in the case of Martin-Löf randomness, the value $2^{-n}$ is arbitrary.

As opposed to Martin-Löf randomness, Schnorr randomness lacks a universal element, a fact which is often regarded as being a major flaw of Schnorr's definition.

### 3.2.3 Kurtz randomness

We conclude this section on the measure-theoretic typicality paradigm by discussing Kurtz' definition of randomness, which he proposed in his doctoral dissertation [1981]. As opposed to Martin-Löf and Schnorr randomness, which both hinge on the idea that a random sequence should avoid all effective measure zero sets, Kurtz defines random sequences as those which belong to all effective measure one sets.

**Definition 3.2.6** (Kurtz randomness)**.** A sequence $X \in \{0,1\}^\omega$ is *Kurtz random* if and only if it belongs to all c.e. open subsets of $\{0,1\}^\omega$ of measure one.

As shown by Wang [1996], Kurtz randomness can also be defined in terms of statistical tests.

**Definition 3.2.7** (Kurtz null test)**.** Let $(\mathcal{U}_n)_{n\in\mathbb{N}}$ be a Martin-Löf test. Then, $(\mathcal{U}_n)_{n\in\mathbb{N}}$ is a *Kurtz null test* if and only if there is a sequence $(\mathcal{S}_n)_{n\in\mathbb{N}}$ of finite, uniformly computable subsets of $\{0,1\}^*$ such that, for each $n \in \mathbb{N}$, $\mathcal{U}_n = [\![\mathcal{S}_n]\!]$.

It follows from Definition 3.2.7 that if $(\mathcal{U}_n)_{n\in\mathbb{N}}$ is a Kurtz null test, then both $(\mathcal{U}_n)_{n\in\mathbb{N}}$ and its complement $(\overline{\mathcal{U}_n})_{n\in\mathbb{N}}$ are computable sequences of c.e. open sets.

**Theorem 3.2.8** (Wang [1996])**.** *Let $X \in \{0,1\}^\omega$. Then, $X$ is Kurtz random if and only if, for any Kurtz null test $(\mathcal{U}_n)_{n\in\mathbb{N}}$, there is no $\mathcal{N} \subseteq \bigcap_{n\in\mathbb{N}} \mathcal{U}_n$ such that $X \in \mathcal{N}$.*

## 3.3 Randomness as unpredictability

According to the unpredictability paradigm, the essence of a random experiment is that it is not possible to make any reasonable predictions on the experiment's future outcomes. Of course, unpredictability was already an important ingredient of von Mises' project; in the context of algorithmic randomness, however, this game-theoretic perspective is taken much more seriously, and a sequence is defined as being random if and only if it is impossible for a gambler to predict the bits of that sequence and gain infinite wealth by successively wagering on them. Unsurprisingly, the betting strategies employed to define randomness in this setting are the martingale functions that we discussed in § 2.2.2 (and which were first introduced by Jean Ville in an attempt at improving on von Mises' definition of randomness). As we will see, this approach—championed by Schnorr—combines Ville's martingales with computability theory, so that randomness is defined in terms of *effective* betting strategies.

We will begin by presenting the notions of *computable randomness* and *partial computable randomness* introduced by Schnorr in [1971a], and that of *Kolmogorov-Loveland randomness* due to Muchnik et al. [1998]. Then, we will see that the randomness concepts defined within the measure-theoretic typicality paradigm have equivalent formulations in terms of martingales with varying computational power. In light of Theorem 2.2.6, these results should come as no big surprise: after all, the unpredictability and the measure-theoretic typicality paradigm hinge on analogous intuitions.

### 3.3.1 Computable and partial computable randomness

Recall from Definition 2.2.5 that a martingale $d : \{0,1\}^* \to \mathbb{R}^{\geq 0}$ (where $\mathbb{R}^{\geq 0}$ denotes the set of non-negative real numbers) is said to succeed on a sequence $X$ if and only if it accrues unbounded capital when run against $X$: that is, $\limsup_{n\to\infty} d(X \upharpoonright n) = \infty$. Computable randomness is then defined as follows, in terms of *computable martingales* (i.e., those martingales that are total computable functions).

**Definition 3.3.1** (Computable randomness)**.** Let $X \in \{0,1\}^\omega$. Then, $X$ is *computably random* if and only if there is no computable martingale that succeeds on it.

With this definition at hand, one can effectivise Theorem 2.2.6: a class $\mathcal{N} \subseteq \{0,1\}^\omega$ is said to have *computable (Lebesgue) measure zero* if and only if there is a computable martingale which succeeds on $\mathcal{N}$.

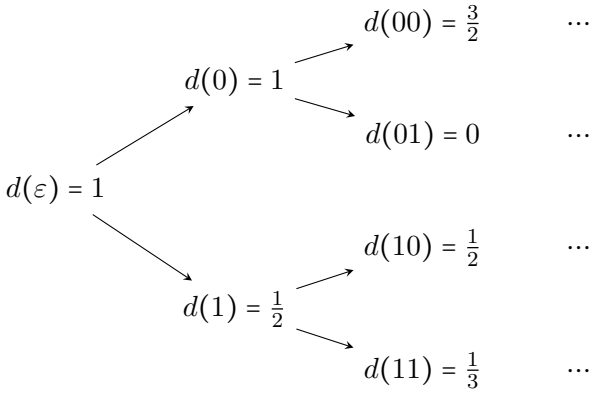
It is possible to strengthen Definition 3.3.1 by considering *partial computable martingales* instead of computable ones—where a partial computable martingale is a partial computable function $d : \{0,1\}^* \to \mathbb{R}^{\geq 0}$ such that

(1) for $\iota \in \{0,1\}$ and $\sigma \in \{0,1\}^*$, if $d(\sigma\iota)$ is defined, then so are $d(\sigma)$ and $d(\sigma(1-\iota))$;

(2) $d$, where defined, satisfies the fairness condition from Definition 2.2.4(1).

Then, one may provide a stronger variant of Definition 3.3.1 by saying that a sequence is *partial computably random* if and only if there is no partial computable martingale which succeeds on it. Clearly, for a partial computable martingale $d$ to succeed on a sequence $X$, $d$ must be defined on each initial segment of $X$.

It should be noted that, for any (computable or partial computable) *supermartingale* $d$ (Definition 2.2.4(3)), there exists a (computable or partial computable) *martingale* $d'$ that succeeds on the same, and possibly more, sequences: $d'$ behaves exactly like $d$, except that it saves all the money that $d$ fritters away. Intuitively, one may view $d'$ as having its capital split between a *checking account* and a *savings account*. All of the martingale's initial capital is stored in the checking account, ready to be used. However, as soon as supermartingale $d$ squanders some money, $d'$ places that same amount in its savings account and does not use it for any further betting: it simply leaves it there to ensure that, at each step, the fairness condition from Definition 2.2.4(1) is met. To see how this works, consider the following example. Let $d$ be a supermartingale whose capital evolves as follows:




Then, we can define a martingale $d'$ in such a way that, for each $\sigma \in \{0,1\}^*$, $d'(\sigma) = C(\sigma) + S(\varepsilon) + \cdots + S(\sigma^-) + S(\sigma)$, where $C : \{0,1\}^* \to \mathbb{R}^{\geq 0}$ is the 'checking account' function, while $S : \{0,1\}^* \to \mathbb{R}^{\geq 0}$ is the 'savings account' function. Begin by setting $C(\varepsilon) = d(\varepsilon) = 1$ and $S(\varepsilon) = 0$, so that $d'(\varepsilon) = C(\varepsilon) + S(\varepsilon) = 1 = d(\varepsilon)$. Then, set $C(0) = d(0) = 1$ and $S(0) = \frac{1}{2}$, so that $d'(0) = C(0) + S(\varepsilon) + S(0) = \frac{3}{2}$; moreover,

set $C(1) = d(1) = \frac{1}{2}$ and $S(1) = 0$, so that $d'(1) = C(1) + S(\varepsilon) + S(1) = \frac{1}{2}$. This ensures that $2 \cdot d'(\varepsilon) = d'(0) + d'(1)$. Then, let $C(00) = d(00) = \frac{3}{2}$, $S(00) = 0$, $C(01) = d(01) = 0$ and $S(01) = \frac{1}{2}$. This gives us that $d'(00) = C(00) + S(\varepsilon) + S(0) + S(00) = 2$, while $d'(01) = C(01) + S(\varepsilon) + S(0) + S(01) = 1$, which guarantees that $2 \cdot d'(0) = d'(00) + d'(01)$. This reasoning can of course be carried on so as to define $d'(\sigma)$ for all $\sigma \in \{0,1\}^*$, and it illustrates why both computable randomness and partial computable randomness could have been equivalently defined in terms of supermartingales (of course, one direction trivially holds because any martingale is also a supermartingale).

As observed by Vermeeren [2013, p. 67], in the definition of martingale success

> We formalised infinite profits by requiring that the limsup of the capital is infinity. One might call this jokingly the *American concept for success*: it does not matter if you lose almost all of your money repeatedly, because in the *land of opportunity* you will always have the possibility to grow rich again. A more *European condition* for success would be to require a more steady growth of capital, without repeated bankruptcies, i.e., the *limit* (and not just the limsup) of the capital should be infinity.

However, it is known that adopting the more 'European condition' for success in terms of limits does not make any difference for the notions of computable randomness and partial computable randomness. This is because, given any (super)martingale $d$, one can effectively define a (super)martingale $d'$ such that (i) $d$ and $d'$ succeed on the very same sequences, and (ii) for all $X \in \{0,1\}^\omega$, $\limsup_{n\to\infty} d(X \upharpoonright n) = \infty$ if and only if $\lim_{n\to\infty} d'(X \upharpoonright n) = \infty$[39].

Without loss of generality, we can also restrict Definition 3.3.1 to (super)martingales with values in the rational numbers. As proved by Schnorr [1971a], for each computable real-valued martingale $d$, there exists a computable rational-valued martingale $f$ that succeeds on exactly the same sequences as $d$. In fact, $f$ is such that $d(\sigma) < f(\sigma) < d(\sigma) + 1$. This result will turn out to be particularly useful in Chapter 5, where we will discuss several notions of algorithmic randomness and stochasticity in terms of rational-valued probabilistic martingales.

### 3.3.2 Kolmogorov-Loveland randomness

Kolmogorov-Loveland randomness is built on the notion of 'non-monotonic betting strategy'. Non-monotonic betting strategies were first defined by Muchnik, Semenov and Uspensky [1998] as a generalisation of Kolmogorov's [1963] and Loveland's [1966] non-monotonic selection functions (which will be formally defined in § 3.3.3). These strategies were introduced in order to rebut Schnorr's critique of Martin-Löf randomness by means of considering computable betting strategies more powerful than those used to define computable randomness, and which could potentially be

[39] The proof of this familiar result is based on the so-called savings trick (see, for instance, [Downey and Hirschfeldt, 2010, Proposition 6.3.8, p. 237]).

employed to find a more natural (martingale-based) characterisation of Martin-Löf randomness.

A non-monotonic betting strategy differs from a monotonic one in that the gambler is allowed to bet in whatever order she might prefer (that is, the gambler can choose which position in the sequence to bet against next), provided that she does not place more than one bet on any one digit of a given sequence. The gambler does not have to bet on all the bits of a sequence; moreover, her choice of whether to bet, and on which position, has to be based on the string of bits that have already been revealed (which is of course not necessarily going to be an initial segment of the infinite sequence that the gambler is playing against) and not on digits yet to be observed.

In the following definition, let $\pi_1$ denote the left projection of a pair (later on, $\pi_2$ will also be used to denote the right projection of a pair).

**Definition 3.3.2** (Non-monotonic betting strategy)**.** A *non-monotonic betting strategy* (non-monotonic strategy, for short) is a function $b : \{0,1\}^* \to \mathbb{N} \times \big(\{\mathsf{scan}\} \cup [-1,1]\big)$ such that, when run against a sequence, no position in that sequence is ever visited more than once. That is, if $\pi_1(b(\sigma)) = n$, then, for all $\tau$ with $\sigma \sqsubset \tau$, $\pi_1(b(\tau)) \neq n$.

Intuitively, given some $\sigma \in \{0,1\}^m$, if $b(\sigma) = (n, \rho)$ (with $\rho \in [-1,1]$), then it means that, after having read the bits $\sigma(1), ..., \sigma(m)$, strategy $b$ has decided to bet a fraction $\rho$ of its current capital on the fact that the $n$-th bit of the infinite sequence being played against is 0—as before, if $\rho$ is negative, then the strategy is betting a fraction $-\rho$ of its current capital on the fact that the $n$-th bit of the sequence is 1. On the other hand, if $b(\sigma) = (n, \mathsf{scan})$, then strategy $b$ has decided to read the $n$-th bit of the sequence against which it is playing without betting anything. Now, some useful notation. Suppose that we are running strategy $b$ against some sequence $X \in \{0,1\}^\omega$. Let $\chi^{(n)} \in \{0,1\}^n$ denote the string consisting of all bits from $X$ that have been observed up to the $n$-th move (so, $\chi^{(0)} = \varepsilon$). For instance, if $X = 10101010...$, $n = 2$, and $b$ is such that $\pi_1(b(\varepsilon)) = 3$ and $\pi_1(b(1)) = 6$, then $\chi^{(2)} = 10$. Let $\mathsf{seen}(n) \subseteq \mathbb{N}$ denote the set of positions that have been visited by $b$ after $n$ moves (so, $\mathsf{seen}(0) = \varnothing$). In the previous example, we then have that $\mathsf{seen}(2) = \{3, 6\}$. Similarly, let $\#\mathsf{bets}(n) \in \mathbb{N}$ be the number of bets that have been placed after $n$ moves (clearly, $\#\mathsf{bets}(n) \leq n$ for all $n$). Finally, let $C(\chi^{(n)}) \in \mathbb{R}^{\geq 0}$ denote the capital accrued by strategy $b$ after having observed $\chi^{(n)}$, where $C(\chi^{(0)}) = 1$. At step $n+1$, we then have that

(i) If $b(\chi^{(n)}) = (k, \mathsf{scan})$ (and $k \notin \mathsf{seen}(n)$), then
   - $\chi^{(n+1)} = \chi^{(n)} X(k)$;
   - $\mathsf{seen}(n+1) = \mathsf{seen}(n) \cup \{k\}$;
   - $\#\mathsf{bets}(n+1) = \#\mathsf{bets}(n)$;
   - $C(\chi^{(n+1)}) = C(\chi^{(n)})$.

(ii) If $b(\chi^{(n)}) = (k, \rho)$, with $\rho \in [-1, 1]$ (and $k \notin \mathsf{seen}(n)$), then

- $\chi^{(n+1)} = \chi^{(n)} X(k)$;
- $\mathsf{seen}(n+1) = \mathsf{seen}(n) \cup \{k\}$;
- $\#\mathsf{bets}(n+1) = \#\mathsf{bets}(n) + 1$;
- $C(\chi^{(n+1)}) = \begin{cases} (1+\rho) \cdot C(\chi^{(n)}) & \text{if } X(k) = 0; \\ (1-\rho) \cdot C(\chi^{(n)}) & \text{if } X(k) = 1. \end{cases}$

Clearly, the capital function $C$ is a martingale. To make the dependence on strategy $b$ and on the sequence being played against more explicit, we write $C_b^X(\chi^{(n)})$ to denote the capital gained by strategy $b$ after $n$ moves when running against $X \in \{0,1\}^\omega$. Then, we say that a non-monotonic betting strategy $b$ *succeeds* on $X$ if and only if $\limsup_{n\to\infty} C_b^X(\chi^{(n)}) = \infty$.

This allows us to define the notion of Kolmogorov-Loveland randomness:

**Definition 3.3.3** (Kolmogorov-Loveland randomness)**.** Let $X \in \{0,1\}^\omega$. Then, $X$ is *Kolmogorov-Loveland random* if and only if there is no computable non-monotonic strategy which succeeds on it.

As shown by Merkle et al. [2006], the concept of Kolmogorov-Loveland randomness is left unchanged if, in Definition 3.3.3, one replaces 'computable' by 'partial computable'.

The following theorem establishes that Kolmogorov-Loveland randomness is implied by Martin-Löf randomness.

**Theorem 3.3.4** (Muchnik et al. [1998])**.** *Let $X \in \{0,1\}^\omega$. If $X$ is Martin-Löf random, then it is Kolmogorov-Loveland random.*

Whether the converse of Theorem 3.3.4 holds as well is still an open question[40].

### 3.3.3 Typicality and stochasticity via martingales

The notion of martingale function can be used to give a uniform characterisation of many of the algorithmic randomness concepts found in the literature.

First of all, note that a martingale $d : \{0,1\}^* \to \mathbb{R}^{\geq 0}$ is said to be c.e. if there exists a computable function $h : \{0,1\}^* \times \mathbb{N} \to \mathbb{Q}$ such that, for all $\sigma \in \{0,1\}^*$, the sequence $(h(\sigma, n))_{n\in\mathbb{N}}$ is non-decreasing and converges to $d(\sigma)$. With this definition at hand, one can obtain a novel, martingale-based characterisation of Martin-Löf randomness.

**Theorem 3.3.5** (Schnorr [1971a])**.** *Let $X \in \{0,1\}^\omega$. Then, $X$ is Martin-Löf random if and only if no c.e. (super)martingale succeeds on it.*

[40] For a compact discussion of some partial results on this issue involving the notions of *permutation* and *injective randomness*, see [Downey and Hirschfeldt, 2010, Section 7.9].

Schnorr and Kurtz randomness are also characterisable in terms of martingales.

**Theorem 3.3.6** (Schnorr [1971a])**.** *Let $X \in \{0,1\}^\omega$. Then, $X$ is Schnorr random if and only if there are no computable martingale $d$ and computable unbounded non-decreasing function $h : \mathbb{N} \to \mathbb{N}$ such that $d(X \upharpoonright n) \geq h(n)$ for* infinitely many *$n \in \mathbb{N}$.*

**Theorem 3.3.7** (Kurtz [1981])**.** *Let $X \in \{0,1\}^\omega$. Then, $X$ is Kurtz random if and only if there are no computable martingale $d$ and computable unbounded non-decreasing function $h : \mathbb{N} \to \mathbb{N}$ such that $d(X \upharpoonright n) \geq h(n)$ for* all *$n \in \mathbb{N}$.*

The notions of Church stochasticity and von Mises-Wald-Church stochasticity from Definition 2.2.1 and Definition 2.2.2(a), on the other hand, may be elegantly characterised in terms of a special type of martingales (first defined by Ambos-Spies et al. [1996]) which always bet a fixed fraction of their current capital on the next bit being 0.

**Definition 3.3.8** (Simple martingale)**.** Let $d$ be a martingale. Then, $d$ is *simple* if and only if there is a rational $\rho \in \mathbb{Q} \cap (0,1)$ such that, for all $\sigma \in \{0,1\}^*$ and $\iota \in \{0,1\}$,

$$d(\sigma\iota) \in \{d(\sigma), (1+\rho) \cdot d(\sigma), (1-\rho) \cdot d(\sigma)\}.$$

**Theorem 3.3.9** (Ambos-Spies et al. [1996])**.** *Let $X \in \{0,1\}^\omega$. Then,*

(a) *$X$ is Church stochastic if and only if there is no total computable simple martingale which succeeds on it;*

(b) *$X$ is von Mises-Wald-Church stochastic if and only if there is no partial computable simple martingale which succeeds on it.*

Now, we say that a martingale $d$ *always eventually bets* if there is no $Y \in \{0,1\}^\omega$ for which there is some $N \in \mathbb{N}$ such that, for all $n \geq N$, $d(Y \upharpoonright n) = d(Y \upharpoonright n-1)$. With this notion at our disposal, Theorem 3.3.9 may be extended to provide a characterisation of weak Church stochasticity (Definition 2.2.2(b)) in terms of simple martingales which always eventually bet[41].

**Proposition 3.3.10.** *Let $X \in \{0,1\}^\omega$. Then, $X$ is weakly Church stochastic if and only if there is no total computable simple martingale which always eventually bets that succeeds on it.*

*Proof.* ($\Rightarrow$) Suppose that there is a total computable simple martingale $d$ which always eventually bets that succeeds on $X$. Since $d$ is a simple martingale, by

[41] Note that Ambos-Spies et al.'s proof of Theorem 3.3.9 relies on the notion of *prediction function* (see Definition 13 in [1996]), while, in our proof of Proposition 3.3.10, we avoid this detour.

definition, whenever it bets, it wagers a positive fraction $\rho$ of its current capital on the next bit being 0. Since $d$ succeeds on $X$ by assumption, we have that

$$\limsup_{n\to\infty} \frac{|\{y < n : d(X \upharpoonright y) = (1+\rho)\cdot d(X \upharpoonright y-1)\}|}{|\{y < n : d(X \upharpoonright y) = (1-\rho)\cdot d(X \upharpoonright y-1)\}|} > 1.$$

We can then define a total computable selection function $s_d : \{0,1\}^* \to \{\mathsf{select}, \mathsf{scan}\}$ by letting

$$s_d(Y \upharpoonright n) = \begin{cases} \mathsf{select} & \text{if } d(Y \upharpoonright n+1) = (1+\rho)\cdot d(Y \upharpoonright n) \\ & \text{or } d(Y \upharpoonright n+1) = (1-\rho)\cdot d(Y \upharpoonright n) \\ \mathsf{scan} & \text{if } d(Y \upharpoonright n+1) = d(Y \upharpoonright n) \end{cases}$$

for all $Y \in \{0,1\}^\omega$. Then, the limsup of the proportion of the bits of $X$ selected by $s_d$ that are 0 is greater than $\frac{1}{2}$, which means that $s_d[X]$ does not satisfy Equation (✠) from the definition of Church stochasticity. Moreover, since, by assumption, $d$ always eventually bets, $s_d[Y]$ is infinite for all $Y \in \{0,1\}^\omega$. Therefore, $X$ is not weakly Church stochastic.

($\Leftarrow$) Suppose that $X$ is not weakly Church stochastic. Then, there is a total computable selection function $s$ such that $s[Y]$ is infinite for all $Y \in \{0,1\}^\omega$, but such that $s[X]$ does not satisfy Equation (✠) from Definition 2.2.1. Recall that, for any $Y \in \{0,1\}^\omega$, $\#\mathsf{zeroes}(Y \upharpoonright n)$ denotes the number of 0's among the first $n$ bits of $Y$. Then, w.l.o.g., there is some $\varepsilon > 0$ such that there are infinitely many $n \in \mathbb{N}$ with

$$\frac{\#\mathsf{zeroes}(s[X] \upharpoonright n)}{n} > \frac{1}{2} + \varepsilon.$$

Employing a well-known technique, we will now turn the selection rule $s$ into a countable collection of martingales. For all $k \in \mathbb{N}$, define $d_k : \{0,1\}^* \to \mathbb{R}^{\geq 0}$ recursively as follows:

(i) $d_k(\varepsilon) = 1$;

(ii) given $d_k(\sigma)$, let $d_k(\sigma 0) = \begin{cases} d_k(\sigma) & \text{if } s(\sigma) = \mathsf{scan} \\ (1+2^{-k})\cdot d_k(\sigma) & \text{if } s(\sigma) = \mathsf{select} \end{cases}$, and let

$$d_k(\sigma 1) = \begin{cases} d_k(\sigma) & \text{if } s(\sigma) = \mathsf{scan} \\ (1-2^{-k})\cdot d_k(\sigma) & \text{if } s(\sigma) = \mathsf{select}. \end{cases}$$

Clearly, for each $k \in \mathbb{N}$, $d_k$ is a total computable simple martingale. Moreover, since $s[Y]$ is infinite for all $Y \in \{0,1\}^\omega$, each $d_k$ always eventually bets. Now, given that there are infinitely many $n \in \mathbb{N}$ such that $\#\mathsf{zeroes}(s[X] \upharpoonright n) > (\frac{1}{2} + \varepsilon)\cdot n$, for each such $n$, consider the bit $s[X](n)$. This digit will occur in $X$ at some position, say,

$m$; in other words, $m$ is the $n$-th integer such that $s(X \upharpoonright m) = \mathsf{select}$. Then, for any $k \in \mathbb{N}$, we have that

$$\begin{aligned} d_k(X \upharpoonright m+1) &= (1+2^{-k})^{\#\mathsf{zeroes}(s[X]\upharpoonright n)} \cdot (1-2^{-k})^{\#\mathsf{ones}(s[X]\upharpoonright n)} \\ &\geq (1+2^{-k})^{(\frac{1}{2}+\varepsilon)\cdot n} \cdot (1-2^{-k})^{(\frac{1}{2}-\varepsilon)\cdot n}. \end{aligned}$$

Hence, $\log d_k(X \upharpoonright m+1) \geq n \cdot \big((\frac{1}{2}+\varepsilon)\cdot\log(1+2^{-k}) + (\frac{1}{2}-\varepsilon)\cdot\log(1-2^{-k})\big)$. We estimate this expression as follows. Let $h : \mathbb{R} \to \mathbb{R}$ be given by $h(r) = (\frac{1}{2}+\varepsilon)\cdot\log(1+r) + (\frac{1}{2}-\varepsilon)\cdot(1-r)$ for all $r \in \mathbb{R}$. Then, $h(0) = 0$. Moreover, since the derivative of $h$ is $h'(r) = \dfrac{\frac{1}{2}+\varepsilon}{1+r} - \dfrac{\frac{1}{2}-\varepsilon}{1-r}$, $h'(0) = 2\varepsilon > 0$. Hence, if $r$ is sufficiently small, we have that $h(r) > 0$. By choosing $k$ large enough, we thus have that $\log d_k(X \upharpoonright m+1) \geq \delta \cdot n$, for some $\delta > 0$ that is independent of $n$. Since $n$ can be chosen arbitrarily large, we have that, for this $k$, $\limsup_{m\to\infty} d_k(X \upharpoonright m) = \infty$. □

Von Mises-Wald-Church stochasticity is implied by both Martin-Löf randomness and partial computable randomness, while Church stochasticity is implied by computable randomness (and, a fortiori, by Martin-Löf randomness and partial computable randomness). Interestingly, as shown by Wang [1996], Schnorr randomness is too weak to imply von Mises-Wald-Church stochasticity and Church stochasticity. However, as proven by Vermeeren, Schnorr randomness does imply weak Church stochasticity.

**Theorem 3.3.11** (Vermeeren [2013])**.** *Let $X \in \{0,1\}^\omega$. If $X$ is Schnorr random, then $X$ is weakly Church stochastic.*

So, weak Church stochasticity nicely fits in the picture, adding to the symmetry underlying the interconnections between measure-theoretic typicality and stochasticity notions (see Figure 3.2).

Since Vermeeren's proof of Theorem 3.3.11 is rather succinct, we present here a more detailed argument. While Vermeeren's original proof relies on König's Lemma, our proof will exploit the compactness of Cantor space.

*Proof of Theorem 3.3.11.*
Suppose that $X$ is not weakly Church stochastic. Then, there is a total computable selection rule $s$ such that

(i) $s[Y]$ is infinite for all $Y \in \{0,1\}^\omega$, and

(ii) $s[X]$ does not satisfy Equation (✠). Then, w.l.o.g., there is some $\varepsilon > 0$ such that there are infinitely many $n \in \mathbb{N}$ with

$$\frac{\#\mathsf{zeroes}(s[X] \upharpoonright n)}{n} > \frac{1}{2} + \varepsilon.$$

Define the function $g : \mathbb{N} \to \mathbb{N}$ as $g(m) = \min_{\sigma \in \{0,1\}^m} |s[\sigma]|$, for all $m \in \mathbb{N}$. Since $s$ is total computable, so is $g$. Moreover, $g$ is non-decreasing. For suppose that there are $\sigma \in \{0,1\}^k$ and $\tau \in \{0,1\}^{k+1}$ such that (i) $g(k) = |s[\sigma]|$ and $g(k+1) = |s[\tau]|$ for some $k \in \mathbb{N}$ (i.e., $\sigma$ and $\tau$ minimise the number of bits selected by $s$ at step $k$ and step $k+1$, respectively), and such that (ii) $g(k+1) < g(k)$. Then, $|s[\tau]| < |s[\sigma]|$, which implies that $|s[\tau^-]| < |s[\sigma]|$, contradicting the minimality of $\sigma$. Finally, we have that $g$ is unbounded. To see that this is indeed the case, suppose towards a contradiction that $g$ is bounded. We then have that $g$ has a maximum $N \in \mathbb{N}$. Take $m_0$ such that $g(m_0) = N$. Then, for all $m \geq m_0$, $g(m) = N$ because, as shown earlier, $g$ is non-decreasing. For each $n \in \mathbb{N}$, let $\mathcal{M}_n$ denote the set of $s$-minimal strings of length $n$. Then, for each $m \geq m_0$, define the set

$$\mathcal{S}_m = \{Y \in \{0,1\}^\omega : Y \upharpoonright m \in \mathcal{M}_m\}.$$

Then, $\mathcal{S}_m = \bigcup\{[\![\sigma]\!] \subseteq \{0,1\}^\omega : \sigma \in \mathcal{M}_m\}$. Since each $[\![\sigma]\!]$ is clopen and $\mathcal{M}_m$ is finite, we then have that $\mathcal{S}_m$ is itself a clopen subset of Cantor space. For all $m \geq m_0$, we also have that $\mathcal{S}_m \supseteq \mathcal{S}_{m+1}$. To see why this is the case, take an arbitrary $m' \geq m_0$ and suppose that there is $Y \in \mathcal{S}_{m'+1}$ which is not in $\mathcal{S}_{m'}$. Since $m'+1 > m_0$, $|s[Y \upharpoonright (m'+1)]| = N$. But $|s[Y \upharpoonright (m')]| > g(m') = N$—because $Y \notin \mathcal{S}_{m'}$—which is a contradiction. This allows us to conclude that, for all $m > m_0$, $\mathcal{S}_{m_0} \cap \ldots \cap \mathcal{S}_m \neq \varnothing$. Since Cantor space is compact, we then have that $\bigcap_{m \geq m_0} \mathcal{S}_m \neq \varnothing$. Hence, there is $Y' \in \{0,1\}^\omega$ such that $g(m) = N = |s[Y' \upharpoonright m]|$ for all $m \geq m_0$. But this means that, past $m_0$, $s$ does not select any more bits from $Y'$, which contradicts our assumption that $s[Y]$ is infinite for every $Y \in \{0,1\}^\omega$. Hence, $g$ must be unbounded.
Now, let $g' : \mathbb{N} \to \mathbb{N}$ be any computable unbounded non-decreasing function that grows more slowly than $\exp(g)$. We know from our initial assumptions that there are infinitely many $n \in \mathbb{N}$ such that $\#\mathsf{zeroes}(s[X] \upharpoonright n) > (\frac{1}{2} + \varepsilon) \cdot n$. Just like in the proof of Proposition 3.3.10, for any such $n$, consider $s[X](n)$: this digit will occur in $X$ at some position, say, $m$. So, $m$ is the $n$-th integer such that $s(X \upharpoonright m) = \mathsf{select}$. Then, for infinitely many values of $n$, we have that

$$\log\Big(d_k\big(X \upharpoonright m+1\big)\Big) > n \cdot h(k) \geq g(m) \cdot h(k),$$

where $d_k$ and $h$ are defined as in the proof of Theorem 3.3.10. Then,

$$\limsup_{m \to \infty} \frac{d_k(X \upharpoonright m)}{g'(m)} = \infty,$$

which establishes that $X$ is not Schnorr random.

□

In Section 2.3, we mentioned that Kolmogorov's own attempt at fixing von Mises' theory of collectives and at providing a frequentist basis for the applicability of probability theory relied on the notion of 'non-monotonic selection function'. In the context of infinite sequences, a non-monotonic selection function is defined as follows:

**Definition 3.3.12** (Non-monotonic selection function)**.** A *non-monotonic selection function* is a function $s : \{0,1\}^* \to \mathbb{N} \times \{\mathsf{select}, \mathsf{scan}\}$ such that if $\pi_1(s(\sigma)) = n$, then, for all $\tau$ with $\sigma \sqsubset \tau$, $\pi_1(s(\tau)) \neq n$.

Intuitively, given some $\sigma \in \{0,1\}^m$, $s(\sigma) = (n, \mathsf{select})$ means that, after having read the bits $\sigma(1), ..., \sigma(m)$, $s$ has decided to include the $n$-th bit of the infinite sequence being played against in the selected subsequence. If, on the other hand, $s(\sigma) = (n, \mathsf{scan})$, then $s$ has decided to simply view (or scan) the $n$-th bit of the sequence being played against, without including it in the selected subsequence. More formally, suppose that $X \in \{0,1\}^\omega$ is the sequence being played against. Then, let $\chi^{(n)} \in \{0,1\}^n$ denote the string of bits from $X$ seen by $s$ after $n$ moves (i.e., all the bits seen, independently of whether they were selected or simply scanned) and $\beta^{(n)} \in \{0,1\}^*$ denote the string of bits selected by $s$ from $X$ after $n$ moves (so, $\chi^{(0)} = \beta^{(0)} = \varepsilon$). Moreover, let $\mathsf{seen}(n) \subseteq \mathbb{N}$ denote the set of positions in $X$ visited by $s$ up to the $n$-th move (so, $\mathsf{seen}(0) = \varnothing$). Then, the $n+1$-st move is given by

(i) If $s(\chi^{(n)}) = (k, \mathsf{scan})$, then
- $\chi^{(n+1)} = \chi^{(n)} X(k)$;
- $\beta^{(n+1)} = \beta^{(n)}$;
- $\mathsf{seen}(n+1) = \mathsf{seen}(n) \cup \{k\}$.

(ii) If $s(\chi^{(n)}) = (k, \mathsf{select})$, then
- $\chi^{(n+1)} = \chi^{(n)} X(k)$;
- $\beta^{(n+1)} = \beta^{(n)} X(k)$;
- $\mathsf{seen}(n+1) = \mathsf{seen}(n) \cup \{k\}$.

If $s$ non-monotonically selects from $X$ an infinite subsequence $(\beta^{(n)})_{n \in \mathbb{N}}$, we denote such sequence by $s[X]$.

Kolmogorov-Loveland stochasticity is then defined as follows.

**Definition 3.3.13** (Kolmogorov-Loveland stochasticity)**.** A sequence $X \in \{0,1\}^\omega$ is *Kolmogorov-Loveland stochastic* if and only if

$$\lim_{n \to \infty} \left( \frac{\#\mathsf{zeroes}(s[X] \upharpoonright n)}{n} \right) = \frac{1}{2}$$

for all computable *non-monotonic* selection functions such that $s[X]$ is infinite.

Given Definition 3.3.13, the following result is immediate (as any computable monotonic selection rule for which Equation (✠) fails is also a computable non-monotonic selection rule for which Equation (✠) fails).

**Proposition 3.3.14** (Folklore)**.** *Let $X \in \{0,1\}^\omega$. If $X$ is Kolmogorov-Loveland stochastic, then it is Church stochastic.*

To nicely complete the picture in light of Vermeeren's inclusion of weak Church stochasticity, we then add the following notion.

**Definition 3.3.15** (Weak Kolmogorov-Loveland stochasticity)**.** A sequence $X \in \{0,1\}^\omega$ is *weakly Kolmogorov-Loveland stochastic* if and only if

$$\lim_{n\to\infty}\left(\frac{\#\mathsf{zeroes}(s[X]\upharpoonright n)}{n}\right) = \frac{1}{2}$$

for all computable non-monotonic selection functions $s$ such that $s[Y]$ is infinite for *all* $Y \in \{0,1\}^\omega$.

One can immediately see that weak Kolmogorov-Loveland stochasticity implies weak Church stochasticity, just like Kolmogorov-Loveland stochasticity implies Church stochasticity.

**Proposition 3.3.16.** *Let $X \in \{0,1\}^\omega$. If $X$ is weakly Kolmogorov-Loveland stochastic, then it is weakly Church stochastic.*

Moreover, the proof of Proposition 3.3.10 can be easily adapted to provide a characterisation of weak Kolmogorov-Loveland stochasticity in terms of non-monotonic betting strategies.

**Proposition 3.3.17.** *Let $X \in \{0,1\}^\omega$. Then, $X$ is weakly Kolmogorov-Loveland stochastic if and only if there is no computable simple non-monotonic betting strategy which always eventually bets that succeeds on $X$.*

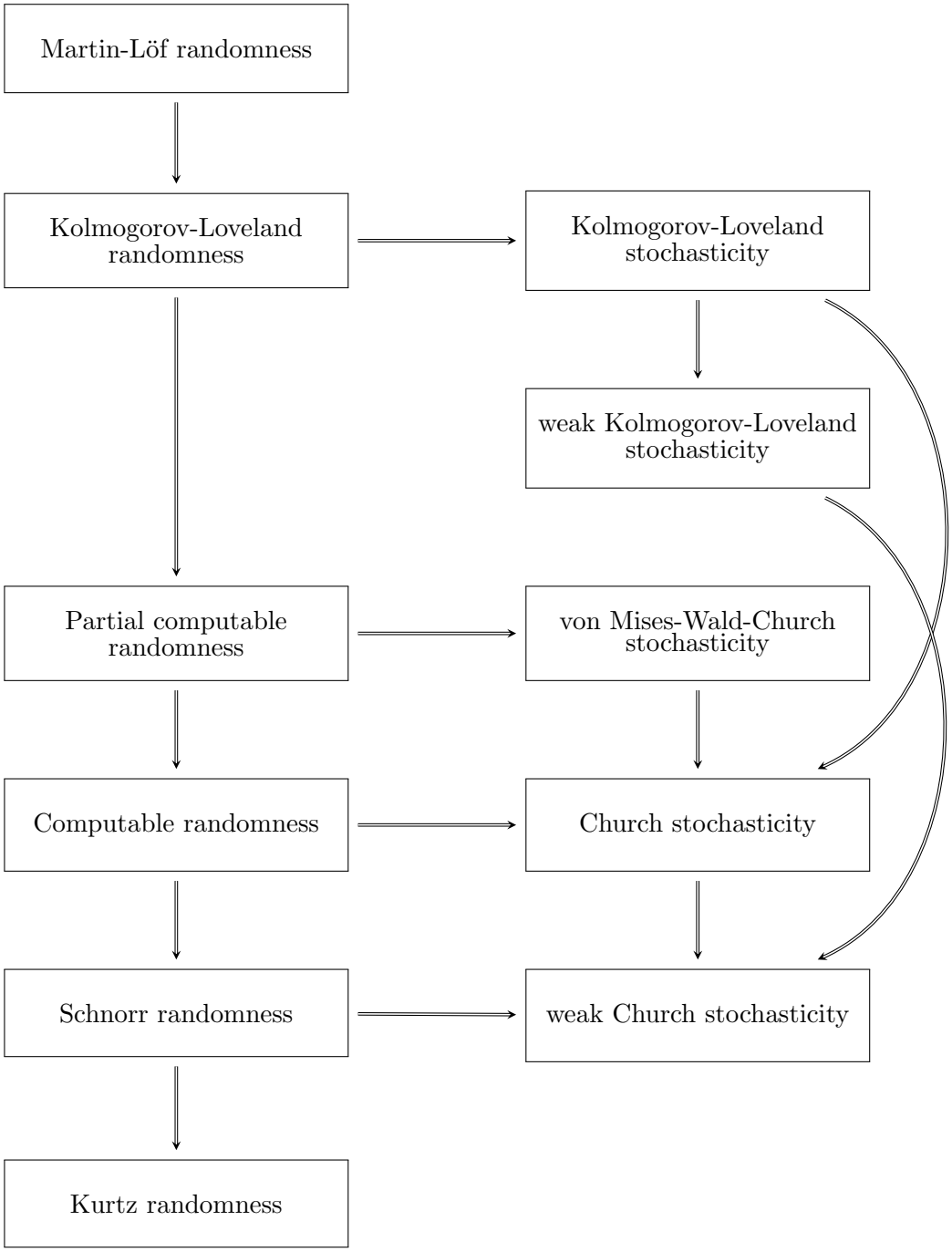



Figure 3.2: Known implications between randomness and stochasticity notions.

# Chapter 4

# Curbing the Algorithmic Randomness Zoo

The randomness notions illustrated in Chapter 3 are but a fraction of the myriad 'beasts' inhabiting the randomness zoo[42]. Recall that the measure-theoretic definition of Martin-Löf randomness presented in § 3.2.1 involves c.e. open subsets of $\{0,1\}^\omega$: these are also known as $\Sigma^0_1$ classes. The notion of a $\Sigma^0_1$ class can be generalised to any $n \in \mathbb{N}$, which gives rise to a hierarchy of $\Sigma^0_n$ (and their complements $\Pi^0_n$) classes analogous to the arithmetical hierarchy of sets. So, one can obtain randomness concepts stronger than Martin-Löf randomness by simply replacing $\Sigma^0_1$ classes by (open) $\Sigma^0_n$ classes ($n > 1$) in Martin-Löf's original definition. Each level of the hierarchy of $\Sigma^0_n$ classes then corresponds to a notion of $n$-randomness. The study of $n$-randomness can be further extended by increasing the logical complexity of the sets of real numbers under consideration: for instance, one can consider algorithmic randomness in the context of second-order arithmetic. In this setting, one investigates the properties of higher randomness notions such as $\Pi^1_1$ randomness, $\Pi^1_1$ Martin-Löf randomness, $\Delta^1_1$ randomness and $\Delta^1_1$ Martin-Löf randomness [Martin-Löf, 1970; Chong et al., 2008]. Naturally, the randomness zoo also extends below Martin-Löf randomness, Schnorr randomness and Kurtz randomness. Schnorr, for example, paved the way for the study of randomness concepts defined in terms of resource-bounded martingales (in the sense of computational complexity), which resulted in notions such as polynomial-time randomness and polynomial-time Schnorr randomness[43].

Given the humongous variety of specimens in the randomness zoo, it is only natural to wonder whether all of these notions are legitimate and useful. Perhaps, there is only one correct definition of algorithmic randomness and all other concepts are just outlandish mathematical abstractions. In light of these considerations, this chapter will be centred around the debate between *algorithmic randomness monism*—i.e., the

[42] See [Taveneaux, 2012] for a nice graphical representation of the randomness zoo.
[43] See [Wang, 1996].

view that there is only on true (algorithmic) randomness notion—and *algorithmic randomness pluralism*—namely, the position according to which there are several definitions of randomness that are correct, in the sense that they possess most of the properties mathematicians have taken to be crucial for randomness. We will begin by outlining three randomness theses that have been defended in the literature, each of which purports that a particular randomness notion is the only correct one. Then, we will appraise the arguments respectively used to support each of these theses and to refute the competitors. First, we will address Osherson and Weinstein's criticism of Martin-Löf randomness [2008]; then, we will discuss Schnorr's famous critique of Martin-Löf randomness [1971b]. As we will see, in spite of seemingly going in opposite directions, both criticisms actually rest on similarly epistemic interpretations of the concept of randomness. We will conclude by arguing that, given the available evidence, a pluralist viewpoint is the most reasonable position to adopt: many (though not all) randomness concepts on the market meet the criteria that mathematicians seem to (more or less) agree a natural notion of randomness should satisfy, so all of these notions are worthy of investigation.

## 4.1 Three randomness theses

Every introductory course in computability theory starts with the Church-Turing Thesis, according to which the intuitive concept of a function being 'effectively calculable in a finite number of steps by a human being following a finite set of rules' is captured by the notion of Turing-machine computability. The Church-Turing Thesis is not a mathematical result, so it cannot be proved; yet, it puts forward a very convincing identification, which derives most of its strength from the fact that every known formal specification of the class of calculable number-theoretic functions has the same extension. In particular, a function is Turing-machine computable if and only if it is general recursive if and only if it is $\lambda$-computable.

In the literature on algorithmic randomness, it is not uncommon to find theses that purport to be analogues of the Church-Turing Thesis for our pre-theoretic notion of randomness[44]. The most famous of these theses is the Martin-Löf Thesis[45], championed, among others, by Delahaye[46] [1993; 2011] and Dasgupta [2010]:

> **The Martin-Löf Thesis.** A sequence is intuitively random if and only if it is Martin-Löf random.

It should be immediately noted that the Martin-Löf Thesis is not meant as an explication of physical chance (it is generally agreed that algorithmic randomness

[44] We will clarify soon what this pre-theoretic notion might amount to.

[45] In the literature, this thesis is sometimes referred to as the Martin-Löf-Chaitin Thesis (see, for instance, [Delahaye, 2011]).

[46] Delahaye [2011, p. 132] also claims that the Martin-Löf Thesis is supported by Chaitin [1987], Kolmogorov and Uspenskii [1987], Gács [1986] and Levin [1973].

on the whole is not the right framework for modelling the kind of randomness we encounter in the physical world). It instead aims at capturing a purely mathematical notion: roughly, it proposes to identify everything that mathematicians have been taking to be significant about randomness with Martin-Löf's celebrated definition. Now, this claim of course raises the question of what it is that mathematicians have been taking to be significant about randomness. Is there a unique such list, or at least a collection of properties that everybody seems to agree upon? We will see that it is precisely by trying to make this idea more concrete by considering specific properties which are generally deemed desirable for a randomness notion that we will come to the conclusion that monism about algorithmic randomness is an unwarranted position.

The most common argument used to defend the Martin-Löf Thesis consists in claiming that Schnorr's characterisation results (Theorem 3.2.4 and Theorem 3.3.5) evince that Martin-Löf's definition is not only mathematically elegant, but also very robust, just like the definition of Turing-machine computability. The convergence of the measure-theoretic, incompressibility and martingale-based definitions must surely be an indication of the fact that we have managed to single out a truly significant notion—in fact, that we have managed to capture the one true concept of mathematical randomness.

The main problem with this reasoning is that, in the context of algorithmic randomness, the robustness argument is not nearly as strong as in the case of the Church-Turing Thesis. This is because, as we have seen, there are countless notions of algorithmic randomness in the literature, and most of them are similarly robust, in that they have formulations in all three different paradigms by now[47]. So, although Theorem 3.2.4 and Theorem 3.3.5 are altogether rather striking, so are their analogues for other randomness notions.

Ultimately, it seems that the main reason why this robustness argument has been deemed to provide strong evidence in favour of the Martin-Löf Thesis is that Martin-Löf randomness, originally defined in terms of statistical tests, was the first algorithmic randomness notion to be characterised in terms of martingales and Kolmogorov complexity, as well. Arguably, it was this historical advantage that made the analogy between the Martin-Löf Thesis and the Church-Turing Thesis appear so plausible at first.

This is not to say that the Martin-Löf Thesis enjoys no credibility whatsoever. As we will see, there are many reasons why Martin-Löf's definition can be said to correspond to a natural notion of randomness. For the time being, we only wish to point out that the robustness argument does not establish that Martin-Löf randomness is the *only* notion that can be said to capture mathematicians'

[47] For instance, we saw in § 3.2.2, § 3.2.3 and § 3.3.3 that Schnorr randomness and Kurtz randomness can be defined both in terms of statistical tests à la Martin-Löf and in terms of martingales. Both notions have also been given an equivalent machine characterisation via *computable measure machines* [Downey et al., 2004; Downey and Griffiths, 2004] and via *decidable machines* [Bienvenu and Merkle].

pre-theoretic intuitions about randomness.

In addition to the Martin-Löf Thesis, there are two other competing theses that have enjoyed some success in the literature, both of which were formulated as a reaction to the perceived inadequacies of Martin-Löf randomness. The first one is Schnorr's Thesis, advocated (unsurprisingly) by Schnorr [1971b]:

> **Schnorr's Thesis.** A sequence is intuitively random if and only if it is Schnorr random.

The second one is more recent and is due to Osherson and Weinstein [2008]:

> **The Weak 2-Randomness Thesis.** A sequence is intuitively random if and only if it is weakly 2-random.

The Weak 2-Randomness Thesis is based on a notion of randomness that is strictly stronger than Martin-Löf randomness[48]:

**Definition 4.1.1** (Weak 2-randomness)**.** (a) Let $(\mathcal{U}_n)_{n\in\mathbb{N}}$ be a computable sequence of c.e. open sets such that $\lim_{n\to\infty} \lambda(\mathcal{U}_n) = 0$. Then, $(\mathcal{U}_n)_{n\in\mathbb{N}}$ is said to be a *generalised Martin-Löf test*.

(b) For every generalised Martin-Löf test $(\mathcal{U}_n)_{n\in\mathbb{N}}$, a set $\mathcal{N} \in \wp(\bigcap_{n\in\mathbb{N}} \mathcal{U}_n)$ is called a *generalised Martin-Löf null set*.

(c) A sequence $X \in \{0,1\}^\omega$ is said to be *weakly 2-random* if and only if there is no generalised Martin-Löf null set $\mathcal{N}$ such that $X \in \mathcal{N}$.

Weak 2-randomness is a generalisation of Kurtz randomness: it can be equivalently characterised in terms of $\Sigma^0_2$ classes of measure one (in fact, Kurtz randomness can be generalised to higher levels as well, just like Martin-Löf randomness).

## 4.2 Critiques of Martin-Löf randomness

In what follows, we will discuss the rationale behind the Weak 2-Randomness Thesis and Schnorr's Thesis, respectively. Interestingly, these two alternatives to the Martin-Löf Thesis hinge on criticisms of Martin-Löf randomness which are, in one sense, diametrically opposed, but, in another respect, based on very similar intuitions.

[48] The first occurrence of the notion of weak 2-randomness in print dates back to [Gaifman and Snir, 1982]. However, weak 2-randomness is already discussed in Solovay's notes [1975]. It also appears in Kurtz' doctoral dissertation [1981].

### 4.2.1 Osherson and Weinstein's critique

Let us begin with a definition. A sequence $X \in \{0,1\}^\omega$ is said to be $\Delta^0_2$ if there is a computable function $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ such that, for each $n$,

(i) $X(n) = \lim_{m\to\infty} f(n,m)$, and

(ii) $f(n,i) \neq f(n,j)$ for only finitely many $i, j$.

So, a sequence is $\Delta^0_2$ if it is *decidable in the limit*.

The Weak 2-Randomness Thesis is mostly motivated by the following observation: there exist some Martin-Löf random sequences which are $\Delta^0_2$. However, being decidable in the limit is a property that, according to Osherson and Weinstein, is not "congruent with the intuition that random sequences lack structure" [2008, p. 2]. Hence, they argue, the collection of 'truly random' sequences cannot be identified with the set of Martin-Löf randoms. Moreover, as proven by Martin[49], there is no $\Delta^0_2$ weakly 2-random sequence[50]. So, weakly 2-random sequences display 'typical' random behaviour (in the sense of Martin-Löf) and, at the same time, they are not decidable in the limit (so, they are not 'close to being computable'). Thus, according to Osherson and Weinstein, weak 2-randomness, as opposed to Martin-Löf randomness, is the correct mathematical formalisation of the intuitive concept of randomness.

Although being decidable in the limit is a property that indeed appears to be somewhat at odds with randomness, there are other methods to gauge the extent to which a sequence is 'close to being computable'. The notion of *lowness* from computability theory, for example, can be viewed as addressing this question. First, note that several notions of algorithmic randomness can be relativised to an oracle. For instance, in the context of randomness concepts defined in terms of effective statistical tests, one could appeal to an oracle containing non-computable information so as to extend the class of available tests and allow for the identification of patterns which could not have been found with merely effective means. An example: given an oracle $A$, an $A$-Martin-Löf test is a nested sequence $(\mathcal{U}_n)_{n\in\mathbb{N}}$ of sets which are $A$-effectively open (i.e., their basic open subsets can be enumerated with oracle $A$), and whose measure is bounded by an $A$-computable sequence of rational numbers which goes to 0. Now, given a randomness concept $\mathcal{R}$ that is relativisable in this sense, we have that a sequence $X \in \{0,1\}^\omega$ is *low for $\mathcal{R}$-randomness* if every $\mathcal{R}$-random sequence is $\mathcal{R}$-random relative to $X$. Hence, $X$ is low for $\mathcal{R}$-randomness if it has no de-randomisation power with respect to it[51].

[49] See [Solovay, 1975].

[50] Note that this pattern actually persists throughout the algorithmic randomness hierarchy: for each $n \in \mathbb{N}$, there are $\Delta^0_{n+1}$ sequences which are $n$-random. However, no $\Delta^0_{n+1}$ sequence can be $(n+1)$-random [Downey and Hirschfeldt, 2010, Section 6.8].

[51] See [Downey and Hirschfeldt, 2010, Definition 11.2.1].

In [2013], Bienvenu et al. employ an altogether different method to measure 'how close a sequence $X$ is to being computable': they consider the size of the domain of any partial computable function that $X$ *extends*—where $X$ is said to extend a partial computable function $\varphi$ if, for all $n \in \mathrm{dom}(\varphi)$, $X(n) = \varphi(n)$. Intuitively, the larger the domain of $\varphi$ is, the closer $X$ is to being computable. Interestingly, in this context, it turns out that *bi-immune* sequences (i.e., sequences which do not contain any computable subsequences) are 'as far as possible from being computable', in the sense that a sequence is bi-immune if and only if it does not extend any partial computable function with infinite domain. Bi-immunity, however, is possibly the weakest randomness notion found in the literature: it is even implied by Kurtz randomness.

This variety of 'measures' that one may use to calibrate the 'distance' of a sequence from computability suggests that, in order to establish that the Weak 2-Randomness Thesis is correct, one would have to offer some further arguments to explain why avoiding $\Delta^0_2$ sequences indeed captures in the most convincing way the idea that a random sequence should be as uncomputable as possible.

To provide such an argument, Osherson and Weinstein offer a learning-theoretic characterisation of weak 2-randomness, creating an interesting bridge between algorithmic randomness and computational learning theory[52]. Their suggestion is that 'true' randomness should be equated with *anonymity*: in a nutshell, a random sequence should not be identifiable or recognisable by any *computable* process (or human mind, as Osherson and Weinstein more provocatively put it).

Osherson and Weinstein's learning-theoretic approach to randomness may be thought of as an infinite game played by a computable process, the learner, against Nature. The game begins with Nature showing the learner some infinite sequence $X \in \{0,1\}^\omega$, in its entire length. After the learner has had a chance to briefly observe $X$ and memorise as much of it as possible, the sequence is hidden from view. Then, Nature starts revealing to the learner some sequence $Y$ which may or may not coincide with $X$, this time one bit at a time, *ad infinitum*. The learner's goal is to determine whether $Y = X$. Whenever the learner believes to be confronted with an initial segment of the target sequence $X$, she has to answer "Yes!"; whenever she believes to be playing against a sequence different from $X$, she has to answer "No!". Now, suppose that $Y$ indeed coincides with the target sequence $X$. The basic intuition behind this game is that if $X$ is a random sequence, then a computable learner should not be able to figure out that $Y = X$, even after having had the opportunity to observe $X$ for a while. If $X$ is non-random, on the other hand, a clever learner might be able to exploit some of $X$'s identifying patterns to guess correctly.

During a run of the game, the learner has to make infinitely many yes/no guesses. In order to determine whether the learner is successful in identifying $X$ (that is, in recognising that $Y = X$), one first has to fix a success criterion against which

[52]For an overview of computational learning theory, see, for instance, [Jain et al., 1999].

to evaluate the learner's performance. The first such criterion that Osherson and Weinstein consider, called *sequence identification*, allows them to provide a learning-theoretic characterisation of weak 2-randomness, which they then use as an argument in favour of the Weak 2-Randomness Thesis.

First, note that a learner can be modelled as a function $\ell : \{0,1\}^* \to \{\mathsf{yes}, \mathsf{no}\}$, which takes as input finite binary strings and outputs either yes or no, depending on whether the input string is conjectured to be an initial segment of the target sequence or not. We will refer to $\ell$ as a *learning function*. Then, sequence identification is defined as follows.

**Definition 4.2.1** (Sequence identification)**.** Let $X \in \{0,1\}^\omega$. A learning function $\ell : \{0,1\}^* \to \{\mathsf{yes}, \mathsf{no}\}$ is said to *identify* $X$ if and only if the set

$$\mathcal{V}_\ell = \left\{Y \in \{0,1\}^\omega : \ell(Y \upharpoonright n) = \mathsf{yes} \text{ for infinitely many } n \in \mathbb{N}\right\}$$

is such that (i) $X \in \mathcal{V}_\ell$, and (ii) $\lambda(\mathcal{V}_\ell) = 0$—where $\lambda$ denotes the Lebesgue measure.

Intuitively, $\mathcal{V}_\ell$ ($\ell$'s 'success' set) contains all those sequences that the learner believes to coincide with the target sequence $X$. In order for $\ell$ to successfully identify $X$, $X$ has to be in the success set associated with $\ell$, and there cannot be too many sequences that $\ell$ mistakes for $X$—hence, the requirement that $\lambda(\mathcal{V}_\ell) = 0$. So, the learner is allowed to make mistakes, but only measure-zero many of them.

With Definition 4.2.1 at hand, Osherson and Weinstein are able to prove the following.

**Proposition 4.2.2** (Osherson and Weinstein [2008])**.** *Let $X \in \{0,1\}^\omega$. Then, $X$ is weakly 2-random if and only if no computable learning function identifies it.*

Note that, as a learning criterion, sequence identification is very permissive. Upon sequentially observing some sequence $X$, the learner may change her mind about whether she is indeed being fed bits from $X$ infinitely many times (i.e., $\ell(X \upharpoonright n) = \mathsf{no}$ may occur for infinitely many $n \in \mathbb{N}$) and yet count as having successfully identified $X$, so long as $\ell$ outputs yes infinitely often. Hence, if $X$ is weak 2-random—and so, by Proposition 4.2.2, no computable learner identifies $X$—then any computable learning function can output yes on at most finitely many initial segments of $X$.

In interpreting Proposition 4.2.2, Osherson and Weinstein adopt an epistemic perspective: weak 2-randomness, they argue, perfectly captures the intuition that a random sequence is one which no computable agent should be able to memorise, and which should not include any patterns discernible by computable agents that could set it apart from other random sequences. In this light, the authors conclude, weak 2-randomness can be seen to be more natural a notion than Martin-Löf randomness.

This line of reasoning—"at the confluence of measure-theoretic and epistemic perspectives on reals" [Osherson and Weinstein, 2008, p. 7]—is somewhat surprising because, as we will see in § 4.2.2, Schnorr's ultimate justification for his randomness thesis relies on a similarly epistemic, evidence-based approach. This is rather

unexpected, given that the Weak 2-Randomness Thesis and Schnorr's Thesis advocate the adoption of two completely different randomness concepts, on the opposite sides of the spectrum with respect to Martin-Löf randomness.

The 'agent-centred' argument delineated above might partially lose its bite in view of another characterisation result proven in the same paper. Osherson and Weinstein in fact consider a second learning criterion besides sequence identification: namely, *strong sequence identification*.

**Definition 4.2.3** (Strong sequence identification)**.** Let $X \in \{0,1\}^\omega$. A learning function $\ell$ is said to *strongly identify* $X$ if and only if the set

$$\mathcal{W}_\ell = \left\{Y \in \{0,1\}^\omega : \ell(Y \upharpoonright n) = \mathsf{yes} \text{ for cofinitely many } n \in \mathbb{N}\right\}$$

is such that (i) $X \in \mathcal{W}_\ell$, and (ii) $\lambda(\mathcal{W}_\ell) = 0$.

Interestingly, strong sequence identification can be used to characterise another familiar randomness concept:

**Proposition 4.2.4** (Osherson and Weinstein [2008])**.** *Let $X \in \{0,1\}^\omega$. Then, $X$ is Kurtz random if and only if no computable learning function strongly identifies it.*

Strong sequence identification is a much more demanding (and, perhaps, more natural) learning criterion than mere sequence identification: in order for her guesses to count as learning, a learner is allowed to change her mind at most a finite (although arbitrarily large) number of times. This means that, when a sequence $X$ is Kurtz random—and so, by Proposition 4.2.4, there is no computable learning function that strongly identifies $X$—there might still be a computable learner that, upon observing $X$, outputs yes infinitely often. So, weak 2-random sequences are indeed more difficult to recognise than Kurtz random sequences.

However, Proposition 4.2.4 may be taken to indicate that equating weak 2-randomness with the impossibility of successful learning by a computable agent is rather misleading. This is because *both* weak 2-randomness and Kurtz randomness can be characterised in terms of *computable* learning functions: the difference between the two characterisation results lies in the choice of identification criterion, not in the computational power of the underlying learner. Moreover, the fact that, when fed a Kurtz random sequence, a computable learner might still answer yes infinitely often is not necessarily problematic: after all, one might argue, answering yes infinitely often hardly counts as learning.

The problem is that, as explained in Chapter 3, Kurtz randomness is strictly weaker than Martin-Löf randomness (in fact, it is strictly weaker than Schnorr randomness, as well). Hence, there are Kurtz random sequences which are decidable in the limit (i.e., which are $\Delta^0_2$). So, it would hardly be a welcome conclusion if the notion of randomness that seems to best capture the learning-theoretic intuitions underlying Osherson and Weinstein's framework turned out to be Kurtz randomness.

Our goal here is not to countenance an argument of this sort. We believe, however, that for the Weak 2-Randomness Thesis to be vindicated, one would have to conclusively show that the above line of reasoning is amiss, and that sequence identification as per Definition 4.2.1 is the only learning criterion that makes intuitive sense in the context of algorithmic randomness. Moreover, any argument to this effect would have to appeal to independent evidence and not rely on the fact that there are Kurtz random sequences which are decidable in the limit. A cogent justification for a claim of this kind, however, does not seem to be readily available.

Setting aside this debate for a moment, it is worth pointing out that Osherson and Weinstein's learning-theoretic framework offers a very nice epistemic perspective on algorithmic randomness. As already noted, its underlying motivation seems to be akin to the rationale behind Schnorr's critique of Martin-Löf randomness (§ 4.2.2), and also to the 'behaviouristic' approach to pseudorandomness often adopted in cryptography. In this latter context, where one is mostly interested in defining the concept of a *pseudorandom generator*, a string is said to be pseudorandom if no efficient observer (where efficiency is cashed out in terms of feasible computation) can distinguish it from a uniformly chosen string of the same length [Wang, 2000].

Moreover, Osherson and Weinstein's work prompts the interesting question of whether other randomness notions may be amenable to a natural learning-theoretic characterisation, as well. For instance, is it possible to obtain Martin-Löf randomness or computable randomness or Schnorr randomness by modifying the amount of admissible mind changes in the definition of sequence identification? How could the requirement that, given a Martin-Löf test $(\mathcal{U}_n)_{n\in\mathbb{N}}$, the measures of the test sets $\mathcal{U}_n$ should converge to zero at a computable rate be captured in learning-theoretic terms? Osherson and Weinstein's results rely on the definition of weak 2-randomness and Kurtz randomness in terms of effective statistical tests. It would also be interesting to compare Osherson and Weinstein's learning-theoretic paradigm with the martingale-based approach to algorithmic randomness discussed in Section 3.3. What is the connection between identifiability and martingale success? Is there a natural correspondence between the sets $\mathcal{V}_\ell$ and $\mathcal{W}_\ell$ from Definition 4.2.1 and Definition 4.2.3 and the set of sequences against which a martingale gains unbounded capital?

#### A brief learning-theoretic digression

In view of the above considerations, in what follows we will have a first pass at further exploring Osherson and Weinstein's learning-theoretic framework (with an eye on the unpredictability paradigm).

Let us begin by introducing a new identification criterion.

**Definition 4.2.5** (Sequence identification with no mind changes)**.** Let $X \in \{0,1\}^\omega$. A learning function $\ell : \{0,1\}^* \to \{\mathsf{yes}, \mathsf{no}\}$ is said to *identify $X$ with no mind changes* if and only if the set

$$\mathcal{Z}_\ell = \left\{Y \in \{0,1\}^\omega : \ell(Y \upharpoonright n) = \mathsf{yes} \text{ for all } n \in \mathbb{N}\right\}$$

is such that (i) $X \in \mathcal{Z}_\ell$, and (ii) $\lambda(\mathcal{Z}_\ell) = 0$.

Definition 4.2.5 allows us to obtain the following, alternative characterisation result for Kurtz randomness.

**Proposition 4.2.6.** *Let $X \in \{0,1\}^\omega$. Then, $X$ is Kurtz random if and only if no computable learning function identifies $X$ with no mind changes.*

*Proof.* ($\Rightarrow$) Suppose that there is a computable learning function $\ell : \{0,1\}^* \to \{\mathsf{yes}, \mathsf{no}\}$ which identifies $X$ with no mind changes. Then, by Definition 4.2.5, the set $\mathcal{Z}_\ell = \left\{Y \in \{0,1\}^\omega : \ell(Y \upharpoonright n) = \mathsf{yes} \text{ for all } n \in \mathbb{N}\right\}$ is such that $X \in \mathcal{Z}_\ell$ and $\lambda(\mathcal{Z}_\ell) = 0$. The characterisation of $\mathcal{Z}_\ell$ in terms of $\ell$, which is by assumption computable, immediately allows us to conclude that $\mathcal{Z}_\ell$ is a $\Pi^0_1$ class. Hence, its complement $\overline{\mathcal{Z}_\ell}$ is a $\Sigma^0_1$ class (i.e., a c.e. open subset of $\{0,1\}^\omega$). Moreover, since $\lambda(\mathcal{Z}_\ell) = 0$, $\lambda(\overline{\mathcal{Z}_\ell}) = 1$. This means that there is a c.e. open subset of $\{0,1\}^\omega$ of measure one, $\overline{\mathcal{Z}_\ell}$, to which $X$ does not belong. Hence, by Definition 3.2.3, $X$ is not Kurtz random.

($\Leftarrow$) Suppose that $X$ is not Kurtz random. Then, by Theorem 3.3.7, there is a computable martingale $d : \{0,1\}^* \to \mathbb{R}^{\geq 0}$ and a computable, non-decreasing and unbounded function $h : \mathbb{N} \to \mathbb{N}$ such that $d(X \upharpoonright n) \geq h(n)$ for all $n$. Then, we define a learning function $\ell : \{0,1\}^* \to \{\mathsf{yes}, \mathsf{no}\}$ as follows: for all $\sigma \in \{0,1\}^*$, set

$$\ell(\sigma) = \begin{cases} \mathsf{yes} & \text{if } d(\sigma) \geq h(|\sigma|); \\ \mathsf{no} & \text{otherwise.} \end{cases}$$

Since $d$ and $h$ are by assumption computable, so is $\ell$. Now, consider the set

$$\mathcal{Z}_d = \{Y \in \{0,1\}^\omega : d(Y \upharpoonright n) \geq h(n) \text{ for all } n \in \mathbb{N}\}.$$

Clearly, $X \in \mathcal{Z}_d$. Moreover, $\mathcal{Z}_d$ is a subset of the collection of sequences on which $d$ manages to accrue unbounded capital. Recall that, by Theorem 2.2.6, a subset of $\{0,1\}^\omega$ has Lebesgue measure zero if and only if there exists a martingale which succeeds on it. By the effective version of Theorem 2.2.6 discussed in § 3.3.1, the set of sequences on which $d$ succeeds has effective Lebesgue measure 0; a fortiori, so does $\mathcal{Z}_d$. Finally, for any $Y \in \{0,1\}^\omega$, $Y \in \mathcal{Z}_d$ if and only if $\ell(Y \upharpoonright n) = \mathsf{yes}$ for all $n \in \mathbb{N}$, by our definition of $\ell$. Hence, $\ell$ is a computable learning function which identifies $X$ with no mind changes. □

**Corollary 4.2.7.** *Sequence identification with no mind changes is equivalent to strong sequence identification.*

An argument analogous to the right-to-left direction of the proof of Proposition 4.2.6 can be employed to show that, given some sequence $X$, if there is no computable learner which identifies $X$ (in the sense of Definition 4.2.1), then $X$ is Schnorr random.

Having indulged in this brief learning-theoretic detour, we now turn to Schnorr's Thesis and Schnorr's critique of Martin-Löf randomness.

### 4.2.2 Schnorr's critique

The rationale for Schnorr's Thesis is a well-known critique, advanced by Schnorr, of the notion of effective null sets introduced by Martin-Löf. Schnorr notes that, although we can know how fast a Martin-Löf test converges to zero, no such test is effectively given: the measures of the test sets $\mathcal{U}_n$ are not computable, they are only left-c.e.; hence, in general, we cannot computably decide whether a given cylinder belongs to the $n$-th level of some Martin-Löf test or not. However, according to Schnorr, the only statistical properties of randomness which have a 'physical meaning' are those properties the failure of which can be established by statistical experience using *computable* methods. Moreover, he argues, 'true' randomness should only include those properties of randomness that have a physical meaning. Hence, although Martin-Löf's definition captures all relevant statistical properties of randomness, it gives rise to a notion of algorithmic randomness which is too strong and devoid of physical significance.

Schnorr's reasoning can be illustrated from within the unpredictability paradigm, as well. His critique of Martin-Löf randomness consists in claiming that a satisfactory notion of randomness should be concerned with defeating computable betting strategies rather than computably enumerable ones, which are fundamentally asymmetric (why should we allow approximations from below, but not from above?). The concept of computable randomness, based on computable martingales, is expressly introduced to remedy this perceived problem. Yet, Schnorr considers computable randomness unsatisfactory, as well:

> Computability and the martingale property [i.e., the fairness condition] suffice to characterize effective tests. But which sequences are refused by an effective test? [...] One would define that a sequence $X$ does not withstand the test $d$ if and only if $\limsup_{n\to\infty} d(X \upharpoonright n) = \infty$. However, if the sequence $d(X \upharpoonright n)$ increases so slowly that no one working with effective methods only would observe its growth, then the sequence $X$ behaves as if it withstands the test $d$. The definition of [the set of sequences on which $d$ succeeds] has to reflect this fact. That is, we have to make *constructive*[53] the notion $\limsup_{n\to\infty} d(X \upharpoonright n) = \infty$ [1971b, p. 256].

To avoid counting as non-random sequences along which the capital of a computable martingale grows very slowly, Schnorr then proposes the notion of Schnorr randomness (as an alternative to computable randomness), where the corresponding martingales are required not only to be computable, but also to succeed in accruing unbounded capital at a *computable rate*.

The explication of Schnorr's critique from the perspective of betting strategies highlights the fact that Schnorr's conception of randomness is akin to that behind

[53] Emphasis added.

the definition of Church stochasticity (and, even more surprisingly, to that behind Osherson and Weinstein's critique): that is, Schnorr takes randomness to be an evidence-based notion. More precisely, in Schnorr's view, a sequence fails to be random if and only if there is a computable process in which this failure becomes apparent. If no such computable process exists, then the sequence behaves like a random sequence. A correct definition of randomness should thus ensure that sequences of this sort are classified as being random.

Such an epistemic perspective, although intriguing, raises the question of whether Schnorr's identification of the class of 'statistical properties with a physical meaning' with the collection of properties whose failure can be detected by a computable process is indeed legitimate. As already noted, Schnorr himself was dissatisfied with the notion of computable randomness and, in defining Schnorr randomness, added the requirement of 'convergence at a computable rate'. One may then take this line of reasoning one step further and argue as follows: even a computable martingale which accumulates unbounded capital at a computable rate may not correspond to a randomness test that is performable in any physically meaningful way (perhaps, because in order to become infinitely rich, a gambler is required to wait for too much time). So, from an epistemic viewpoint, one could very well contend that the only statistical tests that can be said to really have physical meaning are those which are feasibly given (in the sense of computational complexity), rather than those which are only effectively given (in the sense of computability theory). Perhaps, true (epistemic) randomness can only be captured in terms of, say, polynomial-time martingales.

This objection seems to indicate that Schnorr's case for the primacy of computable methods in the definition of randomness is not entirely convincing. If Schnorr's epistemic concerns are supposed to be taken very seriously, one may end up having to focus on pseudo-randomness notions from complexity theory instead of Schnorr randomness. On the other hand, if Schnorr's critique is not meant to be taken too literally, then one may wonder whether there are indeed enough grounds to dismiss computably enumerable martingales à la Martin-Löf as a useful idealisation.

This is not to say, however, that epistemic concerns should play no role when trying to evaluate the usefulness and legitimacy of a randomness notion. As seen in § 4.2.1, epistemic considerations can not only provide new insights, but even give rise to new paradigms within which to study the algorithmic randomness zoo. However, epistemic considerations should not be granted any priority and should be weighted against other possible types of considerations. For instance, we know that Schnorr randomness possesses some rather undesirable properties (such as lacking a universal test, as we already noted in § 3.2.2). So, these drawbacks should be taken into account when evaluating Schnorr's Thesis. This point, we believe, is also poignant in the context of randomness concepts that are weaker than Schnorr randomness (for instance, sub-computable randomness notions), which, although possibly more reasonable from an epistemic point of view, not only inherit the problems that

Schnorr randomness has, but also face other issues of their own.

## 4.3 From a pluralist point of view

In view of our earlier discussion, there does not seem to be any argument offered in favour of the Martin-Löf Thesis or Schnorr's Thesis or the Weak 2-Randomness Thesis which successfully establishes the correctness of the corresponding randomness notion and the inadequacy of all other randomness concepts. However, this could perhaps be the case because the arguments put forward by the proponents of these three randomness theses are not exhaustive enough: they do not consider all properties that a satisfactory randomness concept should possess and weight them against one another. Perhaps, if such a list were available, then it could be conclusively shown that there is only one algorithmic randomness notion which satisfies all of these properties and is therefore the correct formalisation of our intuitive conception of randomness. In what follows, we will see that, by explicitly considering a possible (non-exhaustive) list of desiderata for algorithmic randomness, pluralism, rather than monism, emerges as the most reasonable position to adopt.

In [2013; 2015], Rute discusses the following list of features that are generally acknowledged in the literature as being desirable for a randomness concept, and whose investigation has been fuelling much of the most recent work in algorithmic randomness (in particular, in the context of the interplay of algorithmic randomness with computable analysis and computable measure-theory):

(1) A natural randomness notion should be generalisable to other probability spaces (as opposed to being definable only on Cantor space equipped with the Lebesgue measure) and, ideally, to all computable probability spaces.

(2) It should be invariant under isomorphisms between probability spaces: that is, given two computable probability spaces $(\mathcal{X}, \mu)$ and $(\mathcal{Y}, \nu)$ and isomorphism $T : (\mathcal{X}, \mu) \to (\mathcal{Y}, \nu)$, we should have that if $X \in \mathcal{X}$ is random, so is $T(X) \in \mathcal{Y}$.

(3) Randomness should also be conserved under any computable measure-preserving map: that is, for any computable measure-preserving map $T : \{0,1\}^\omega \to \{0,1\}^\omega$ and $X \in \{0,1\}^\omega$, we should have that if $X$ is random, then so is $T(X)$. This criterion is known as *randomness preservation*.

(4) Measures that have the same null sets should count the same sequences as being random. When two measures (defined on the same space, of course) have the same null sets, they are in fact equivalent.

(5) The dual property of randomness preservation should hold as well: that is, for any effective measure-preserving map $T : \{0,1\}^\omega \to \{0,1\}^\omega$ and any random sequence $Y \in \{0,1\}^\omega$, there should be a random $X \in \{0,1\}^\omega$ such that $T(X) = Y$. This criterion is known as *no-randomness-from-nothing*.

(6) Finally, a natural randomness notion should satisfy van Lambalgen's Theorem[54] (or some appropriate variant thereof).

Now, how do the notions of randomness discussed in this dissertation—and, in particular, those furthered by the randomness theses discussed in § 4.1—fare with respect to the above criteria? Property (1) is satisfied, among others, by weak 2-randomness, Martin-Löf randomness, computable randomness, Schnorr randomness, and Kurtz randomness. It is not clear whether partial computable randomness satisfies it, as well. Property (2) obtains for all of the above notions, except for partial computable randomness. Randomness preservation (property (3)) holds for weak 2-randomness, Martin-Löf randomness, Schnorr randomness, and Kurtz randomness, but not for computable randomness. Weak 2-randomness, Martin-Löf randomness, computable randomness, Schnorr randomness, and Kurtz randomness all satisfy property (4). No-randomness-from-nothing (property (5)), on the other hand, holds for weak 2-randomness, Martin-Löf randomness and computable randomness, but not for Schnorr randomness[55]. Finally, Martin-Löf randomness satisfies the original formulation of van Lambalgen's Theorem, while neither Schnorr nor computable nor weak 2-randomness do. However, if the usual notion of relativised Schnorr randomness is replaced with the weaker notion of *uniformly relative* Schnorr randomness (see [Miyabe and Rute, 2013]), then van Lambalgen's Theorem can be shown to hold for Schnorr randomness, too. In the case of computable randomness, only one direction of van Lambalgen's Theorem for uniform relativisation holds.

Although Martin-Löf randomness still seems to have the upper hand with respect to Rute's list of desiderata (which is by no means exhaustive), there are many other notions of randomness which turn out to satisfy most of the above criteria. In particular, Schnorr randomness does nearly as well as Martin-Löf randomness. In view of these positive results, the moral of the controversy over monism vs. pluralism in the context of algorithmic randomness seems to be that adopting a pluralist perspective may indeed the most reasonable approach. By taking seriously and making precise the very criterion that Delahaye uses to argue for the Martin-Löf Thesis—namely, that a natural notion of randomness should satisfy as many of the properties mathematicians deem desirable for a randomness concept as possible—one

[54] In its most basic form, van Lambalgen's Theorem is the following result:

**Theorem** (van Lambalgen [1990]). The following are equivalent:

(a) $X \oplus Y$ is $n$-random (where $X \oplus Y$ is the *join* of $X$ and $Y$);

(b) $X$ is $n$-random and $Y$ is $n$-random relative to $X$;

(c) $Y$ is $n$-random and $X$ is $n$-random relative to $X$;

(d) $X$ and $Y$ are relatively $n$-random.

Intuitively, van Lambalgen's Theorem states that a random sequence should have the property that no information about its 'left part' should be obtainable from its 'right part' and vice versa.

[55] In fact, Martin-Löf randomness is the weakest algorithmic randomness notion which satisfies both no-randomness-from-nothing and randomness preservation.

can see that several randomness notions are, in a sense, correct and provide valuable insights into the workings of computable mathematics. So, this discussion suggests that there is not always a single answer to the question "is this sequence/set random?", and, rather than trying to dismantle the randomness zoo in order to single out one true randomness notion, the task ahead is more adequately seen as a taxonomic one: the randomness zoo can be organised in different ways, each of which could potentially reveal some new interesting connections between algorithmic randomness and other branches of mathematics. Our conclusion is thus consonant with Porter's:

> [W]e cannot justifiably say of any of these three randomness thesis candidates what Gödel said of the definition of computability, that we have an "absolute definition of an interesting epistemological notion, i.e., one not depending on the formalism chosen" [1946, p. 150]. None of these definitions of randomness captures everything that mathematicians have taken to be significant concerning the concept of randomness. Rather, we have a family of definitions of an interesting epistemological notion, many of which provide insight into certain mathematically significant notions of typicality [2015, p. 4].

To summarise our discussion: ultimately, the main issue with the randomness theses proposed in the literature is not that it does not make sense to ask the question of whether a given randomness notion is natural, or whether it satisfies the pre-theoretic intuitions about randomness held by the mathematical community. The problem is that, once these pre-theoretic intuitions are made precise and concrete, it turns out that various notions of randomness can be legitimately said to satisfy them, not just one.

# Chapter 5

# Randomness via Probabilistic Martingales

Most of the literature on Kolmogorov-Loveland randomness is motivated by the desire to obtain stronger notions of algorithmic randomness, closer to Martin-Löf's definition, while also complying with Schnorr's more stringent adequacy criteria. The very introduction of non-monotonic betting strategies was an attempt at extending Ville's definition of martingale functions in order to obtain more powerful, albeit still computable strategies [Merkle et al., 2006]. So, even though Schnorr's critique of Martin-Löf randomness is not entirely convincing, it has begun to be taken more seriously in the literature, and it does raise the interesting question of whether Martin-Löf randomness can be given an alternative characterisation in computable terms.

In trying to improve the performance of a betting strategy, it is natural to consider strategies where the gambler is allowed to randomise her bets. After all, probabilistic strategies or algorithms play a pivotal role in many fields (for instance, game theory and computational complexity), and probabilistic computation is known to be more powerful than its deterministic counterpart in several computational settings. Building on these intuitions, Buss and Minnes [2013] extend the unpredictability paradigm discussed in Section 3.3 precisely by allowing betting strategies to be probabilistic. This, as we will see, enables them to answer our latter question affirmatively: Martin-Löf randomness can indeed be characterised in terms of computable (probabilistic) martingales[56].

In this chapter, we will further explore the probabilistic framework introduced by Buss and Minnes. We will show that some natural modifications of their definitions give rise to notions of randomness that are in fact equivalent to Martin-Löf random-

[56]Hitchcock and Lutz [2006] independently proved that Martin-Löf randomness can be defined via computable *martingale processes*, as well. See [Downey and Hirschfeldt, 2010, § 6.3.4] for an overview of martingale processes.

ness, Schnorr randomness, Kurtz randomness and Kolmogorov-Loveland randomness, respectively. These results indicate that Buss and Minnes' framework can be successfully employed to provide a uniform characterisation of many randomness notions commonly found in the literature, as well as to explore non-standard randomness concepts. Moreover, this probabilistic setting offers a different, and hopefully fruitful, perspective from which to investigate the long-standing open question of whether Martin-Löf randomness and Kolmogorov-Loveland randomness coincide or not.

## 5.1 Probabilistic martingales

As in the classical martingale scenario illustrated in § 2.2.2 and Section 3.3, within Buss and Minnes' framework, a probabilistic betting strategy provides a gambler with a recipe for betting on the bits of an infinite binary sequence. The gambler is assumed to have a starting capital of 1—that is, the capital function is taken to be a normed martingale. Then, at each step, the strategy computes deterministically a *rational* probability value between 0 and 1 and a *rational* stake value. If the bet is placed and is correct, then the stake is added to the gambler's previous capital. If the bet is placed but is incorrect, then the stake is subtracted from the gambler's previous capital. Finally, if the gambler does not bet, the corresponding bit is not revealed and the capital stays the same[57]. In the next step, the gambler will again employ the same strategy to probabilistically determine (i) whether or not to bet on the same bit that was not revealed during the previous step, and (ii) with what stake to bet.

More formally: take a sequence $X \in \{0,1\}^\omega$ and let $\{\mathsf{b},\mathsf{w}\}^*$ denote the binary computation tree consisting of all possible deterministic betting strategies that a gambler could adopt against $X$. That is, each finite path $\pi \in \{\mathsf{b},\mathsf{w}\}^*$ will represent a gambler's possible sequence of decisions of whether to bet ($\mathsf{b}$) or to refrain from betting ('wait' or $\mathsf{w}$) on some initial segment of $X$. More precisely, each such $\pi$ will coincide with a possible sequence of moves against $X \upharpoonright \#\mathsf{bets}(\pi)$, where $\#\mathsf{bets}(\pi)$ denotes the number of bets among the choices encoded by $\pi$ (similarly, we will let $\#\mathsf{waits}(\pi)$ denote the number of wait moves among the choices encoded by $\pi$). We consider $X \upharpoonright \#\mathsf{bets}(\pi)$ rather than $X \upharpoonright |\pi|$ because, as already mentioned, if the gambler does not bet, then the next bit of $X$ is not revealed.

We can now formally characterise the notion of a probabilistic betting strategy or martingale (since, as we shall see in Definition 5.1.3, the capital accumulated by a probabilistic betting strategy is a martingale, the terms 'probabilistic strategy' and 'probabilistic martingale' will be used interchangeably).

**Definition 5.1.1** (Probabilistic strategy)**.** A *probabilistic strategy* $A$ is a pair $\langle p_A, q_A\rangle$, where $p_A : \{\mathsf{b},\mathsf{w}\}^* \times \{0,1\}^* \to \mathbb{Q} \cap [0,1]$ and $q_A : \{\mathsf{b},\mathsf{w}\}^* \times \{0,1\}^* \to \mathbb{Q}$ are two

[57]The reason behind this condition will be clarified once the framework has been formally illustrated.

*computable* rational-valued functions.

Suppose that the gambler is playing against some sequence $X \in \{0,1\}^\omega$. Intuitively, given (i) some $\sigma \sqsubset X$ that the gambler has observed up to now and (ii) the gambler's moves so far, as encoded by some string $\pi \in \{\mathsf{b},\mathsf{w}\}^*$, the value $p_A(\pi,\sigma)$ corresponds to the probability that the gambler will place a bet on the next bit of $X$—namely, $X \upharpoonright (|\sigma|+1)$. The value $q_A(\pi,\sigma)$, on the other hand, is the stake associated with this bet, if the bet is placed. If $q_A(\pi,\sigma) > 0$, then the gambler is betting that $X(|\sigma|+1) = 0$; if $q_A(\pi,\sigma) < 0$, on the other hand, the gambler is betting that $X(|\sigma|+1) = 1$. It is also possible for $q_A(\pi,\sigma)$ to be equal to 0, in which case the gambler's move is still counted as a bet, albeit with a null stake. To discriminate between this situation and the one in which the gambler opts for a wait move upon seeing $\sigma$, the former case is set to prompt the uncovering of the next bit of $X$, while the latter is not. This distinction may at first seem redundant; however, it allows one to consider strategies with a capital which grows at a given rate in the number of bets, but not necessarily in the number of moves.

From now on, we will use $\pi$ both to denote a finite path in $\{\mathsf{b},\mathsf{w}\}^*$ and as a label for the node in the computation tree that can be reached from the root $\varepsilon$ by following path $\pi$. The probability that strategy $A$ reaches any particular node in the computation tree will of course depend on the values of $p_A$, which, in turn, hinge on the so-far revealed bits of the infinite sequence being played against.

**Definition 5.1.2** (Cumulative probability)**.** Given $\pi \in \{\mathsf{b},\mathsf{w}\}^*$ and $\sigma \in \{0,1\}^*$, the *cumulative probability* $P_A(\pi,\sigma)$ of $\pi$ relative to $\sigma$ is the probability that strategy $A$ reaches node $\pi$ in the computation tree when playing against a sequence $X$ of which $\sigma$ is an initial segment. More precisely, $P_A(\varepsilon,\varepsilon) = 1$, while, for any other pair $(\pi,\sigma)$, we have that

$$P_A(\pi,\sigma) = \begin{cases} P_A(\pi^-,\sigma^-)\cdot p_A(\pi^-,\sigma^-) & \text{if } \pi = (\pi^-)\mathsf{b}; \\ P_A(\pi^-,\sigma)\cdot(1-p_A(\pi^-,\sigma)) & \text{if } \pi = (\pi^-)\mathsf{w}. \end{cases}$$

Intuitively, if the gambler's latest move was a bet move—i.e., if $\pi = (\pi^-)\mathsf{b}$—then, during the previous stage of the game corresponding to node $\pi^-$, the gambler had only seen string $\sigma^-$. Hence, $P_A(\pi,\sigma)$ is simply the cumulative probability $P_A(\pi^-,\sigma^-)$ of reaching node $\pi^-$, multiplied by the probability $p_A(\pi^-,\sigma^-)$ of betting, given that the gambler's previous moves are encoded by $\pi^-$ and the string observed so far is $\sigma^-$. If, on the other hand, the gambler's latest move was a wait move—i.e., if $\pi = (\pi^-)\mathsf{w}$—then the gambler had already observed string $\sigma$ during the round of the game corresponding to node $\pi^-$. Hence, $P_A(\pi,\sigma)$ is the cumulative probability $P_A(\pi^-,\sigma)$ of reaching node $\pi^-$, multiplied by the probability $(1-p_A(\pi^-,\sigma))$ of *not* betting when confronted with $\pi^-$ and $\sigma$.

We can then define the capital process induced by the probabilistic strategy $A$.

**Definition 5.1.3** (Capital)**.** Given $\pi \in \{\mathsf{b},\mathsf{w}\}^*$ and $\sigma \in \{0,1\}^*$, the *capital* $C_A(\pi,\sigma)$ at $\pi$ relative to $\sigma$ is the amount of money available at node $\pi$ after having played against

the initial segment $\sigma$ of some sequence $X \in \{0,1\}^\omega$. More precisely, the starting capital $C_A(\varepsilon,\varepsilon)$ is 1, while, for any other pair $(\pi,\sigma)$, the value $C_A(\pi,\sigma)$ depends on whether $\pi = (\pi^-)\mathsf{b}$ or $\pi = (\pi^-)\mathsf{w}$. If $\pi = (\pi^-)\mathsf{w}$, then $C_A(\pi,\sigma) = C_A(\pi^-,\sigma)$, because, when no bet is placed, no money is gained or lost, and no new bit is revealed. If, on the other hand, $\pi = (\pi^-)\mathsf{b}$, then

$$C_A(\pi,\sigma) = \begin{cases} C_A(\pi^-,\sigma^-) + q_A(\pi^-,\sigma^-) & \text{if } \sigma = (\sigma^-)0 \\ C_A(\pi^-,\sigma^-) - q_A(\pi^-,\sigma^-) & \text{if } \sigma = (\sigma^-)1, \end{cases}$$

where $\dfrac{q_A(\pi^-,\sigma^-)}{C_A(\pi^-,\sigma^-)}$ is required to be in $[-1,1]$ to ensure that, at each step, the stake does not exceed the available capital.

Definition 5.1.3 entails that, for each $j \in \mathbb{N}$, $2 \cdot C_A(\pi,\sigma) = C_A(\pi\mathsf{w}^j\mathsf{b},\sigma 0) + C_A(\pi\mathsf{w}^j\mathsf{b},\sigma 1)$: i.e., $C_A$ satisfies the fairness condition illustrated in Definition 2.2.4(a). Hence, $C_A$ is a normed rational-valued martingale.

For ease of notation, from now on we shall adopt the following abbreviations: given $X \in \{0,1\}^\omega$, let

— $p_A^X(\pi)$ denote $p_A(\pi, X \upharpoonright \#\mathsf{bets}(\pi))$,

— $q_A^X(\pi)$ denote $q_A(\pi, X \upharpoonright \#\mathsf{bets}(\pi))$,

— $P_A^X(\pi)$ denote $P_A(\pi, X \upharpoonright \#\mathsf{bets}(\pi))$, and

— $C_A^X(\pi)$ denote $C_A(\pi, X \upharpoonright \#\mathsf{bets}(\pi))$.

Now, the set of all infinite sequences of bet/wait moves $\{\mathsf{b},\mathsf{w}\}^\omega$, together with the product topology, is homeomorphic to Cantor space. The base which generates the product topology is given by all sets of the form $[\![\pi]\!] = \{\Pi \in \{\mathsf{b},\mathsf{w}\}^\omega : \pi \sqsubset \Pi\}$, which will be referred to as cylinders. Let $\mathcal{C}$ denote the collection of all such cylinders. Since $\mathcal{C} \subseteq \wp(\{\mathsf{b},\mathsf{w}\}^\omega)$, take the sigma-algebra $\sigma(\mathcal{C})$ generated by $\mathcal{C}$ (namely, the smallest sigma-algebra containing $\mathcal{C}$). Then, $(\{\mathsf{b},\mathsf{w}\}^\omega, \sigma(\mathcal{C}))$ is a measurable space. Given a probabilistic strategy $A$ and a sequence $X \in \{0,1\}^\omega$, one can then define a probability measure $\mu_A^X : \sigma(\mathcal{C}) \to [0,1]$ over $(\{\mathsf{b},\mathsf{w}\}^\omega, \sigma(\mathcal{C}))$ by setting $\mu_A^X(\varnothing) = 0$ and, for every $\pi \in \{\mathsf{b},\mathsf{w}\}^*$, by setting $\mu_A^X([\![\pi]\!]) = P_A^X(\pi)$.

A probabilistic strategy is said to be *successful* against an infinite sequence $X \in \{0,1\}^\omega$ along a fixed infinite path $\Pi \in \{\mathsf{b},\mathsf{w}\}^\omega$ if it allows a gambler to become infinitely rich by betting against $X$ in accordance with the moves encoded by $\Pi$: that is, if $\lim_{n\to\infty} C_A^X(\Pi \upharpoonright n) = \infty$. With the notion of a probability measure at hand, we can also talk about the probability with which a probabilistic strategy succeeds, when played against some sequence.

**Definition 5.1.4** (Probability of success)**.** The *probability of success* of a probabilistic strategy $A$ when playing against $X \in \{0,1\}^\omega$ is given by

$$\mu_A^X\left(\{\Pi \in \{\mathsf{b},\mathsf{w}\}^\omega : \lim_{n\to\infty} C_A^X(\Pi \upharpoonright n) = \infty\}\right).$$

## 5.2 P1-randomness and Ex-randomness

The first algorithmic randomness concept introduced by Buss and Minnes is P1-*randomness*, according to which an infinite sequence $X$ is random if and only if no probabilistic betting strategy, when played against $X$, succeeds on it with probability one.

**Definition 5.2.1** (P1-randomness)**.** Let $X \in \{0,1\}^\omega$.

(a) A probabilistic strategy $A$ is said to be a P1-strategy for $X$ if and only if

$$\mu_A^X\left(\{\Pi \in \{\mathsf{b},\mathsf{w}\}^\omega : \lim_{n\to\infty} C_A^X(\Pi \upharpoonright n) = \infty\}\right) = 1;$$

(b) $X$ is said to be P1-*random* if and only if no probabilistic strategy is a P1-strategy for $X$.

The second notion of randomness proposed by Buss and Minnes, Ex-*randomness*, instead relies on the concept of 'infinite capital in expectation'. For each $n \in \mathbb{N}$, we first define the set of computation nodes in $\{\mathsf{b},\mathsf{w}\}^*$ that are reachable with $n$ bets, and such that the last move consisted in betting[58]. We do so by induction:

(i) $R(0) = \{\varepsilon\}$;

(ii) $R(n+1) = \bigcup_{\pi \in R(n)} \{\pi \mathsf{w}^j \mathsf{b} \in \{\mathsf{b},\mathsf{w}\}^* : j \in \mathbb{N}\}$.

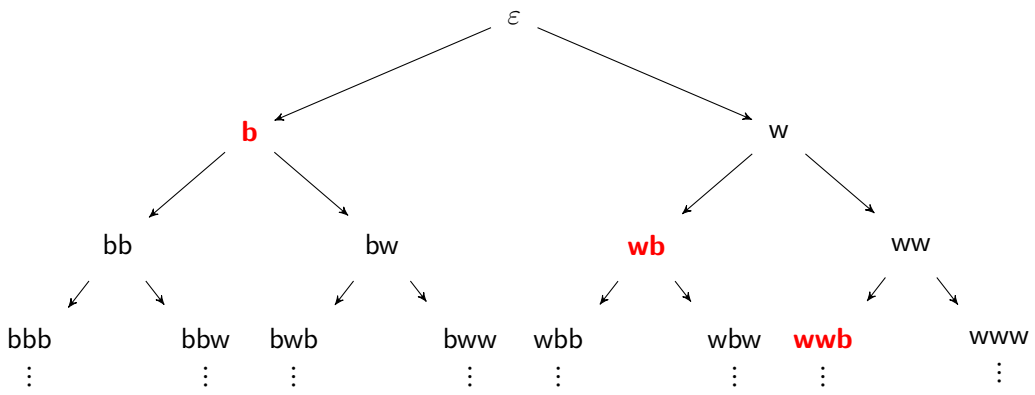



Figure 5.1: Partial representation of the set of computation nodes (red nodes) reachable with one bet, and such that the last move consisted in placing a bet.

[58] See [Buss and Minnes, 2013, § 7] for a thorough explanation of why a more general way of defining Ex-randomness fails.

**Definition 5.2.2** (Expected capital)**.** Let $A$ be a probabilistic strategy and $X \in \{0,1\}^\omega$. The *expected capital* of $A$ over $X$ after $n$ bets is defined as

$$\mathsf{Ex}_A^X(n) = \sum_{\pi \in R(n)} P_A^X(\pi) \cdot C_A^X(\pi).$$

For ease of notation, the expected capital $\mathsf{Ex}_A^X(|\sigma|)$ after seeing $\sigma \in \{0,1\}^*$, where $X$ is any infinite sequence extending $\sigma$, will be denoted by $\mathsf{Ex}_A^\sigma$.

We are now ready to formally characterise the notion of Ex-randomness introduced by Buss and Minnes.

**Definition 5.2.3** (Ex-randomness)**.** Let $X \in \{0,1\}^\omega$.

(a) A probabilistic strategy $A$ is an Ex-strategy for $X$ if and only if

$$\lim_{n\to\infty} \mathsf{Ex}_A^X(n) = \infty;$$

(b) $X$ is said to be Ex-*random* if and only if no probabilistic strategy is an Ex-strategy for $X$.

The average capital accumulated by a probabilistic strategy can be shown to be a supermartingale[59]: i.e., given a probabilistic strategy $A$ and any $\sigma \in \{0,1\}^*$, $2 \cdot \mathsf{Ex}_A^\sigma \geq \mathsf{Ex}_A^{\sigma 0} + \mathsf{Ex}_A^{\sigma 1}$. The reason why it is a supermartingale rather than a martingale is that the underlying probabilistic strategy, $A$, may refrain from placing any further bets from some point onwards.

As proven by Buss and Minnes, Ex-randomness implies P1-randomness. In fact, it is sufficient to have a positive-measure fraction of the computation paths with capital tending to infinity to ensure that the expected capital does so, as well[60].

A simple argument shows that the collection of Ex-random sequences and that of P1-random sequences both have measure one. Take some $\mathcal{N} \subseteq \{0,1\}^\omega$ and suppose that $A$ is a probabilistic strategy such that $\lim_{n\to\infty} \mathsf{Ex}_A^X(n) = \infty$ for all $X \in \mathcal{N}$. We can then construct a Martin-Löf test $(\mathcal{U}_n)_{n\in\mathbb{N}}$ that each $X \in \mathcal{N}$ fails, which suffices to establish that $\mathcal{N}$ is a Martin-Löf null set (see Definition 3.2.3). For each $n \in \mathbb{N}$, define $\mathcal{S}_n = \{\sigma \in \{0,1\}^* : \mathsf{Ex}_A^\sigma > 2^n\}$. Then, set $\mathcal{U}_n = \bigcup \{[\![\sigma]\!] \subseteq \{0,1\}^\omega : \sigma \in \mathcal{S}_n\}$. Clearly, $\mathcal{N} \subseteq \bigcap_{n\in\mathbb{N}} \mathcal{U}_n$. Moreover, $(\mathcal{U}_n)_{n\in\mathbb{N}}$ is a computable sequence of c.e. open sets, for the values of $\mathsf{Ex}_A^\sigma$ are uniformly computably approximable from below. Now take, for each $n \in \mathbb{N}$, a prefix-free subset $\mathcal{T}_n$ of $\mathcal{S}_n$ such that the cylinders generated by its elements cover $\mathcal{U}_n$. Then,

$$\lambda(\mathcal{U}_n) = \sum_{\sigma\in\mathcal{T}_n} 2^{-|\sigma|} < \sum_{\sigma\in\mathcal{T}_n} 2^{-|\sigma|} \cdot \mathsf{Ex}_A^\sigma \leq 2^{-n} \cdot \mathsf{Ex}_A^\varepsilon,$$

[59] For a proof of this result, see [Buss and Minnes, 2013, Lemma 4.6, p. 12].

[60] For a proof, see [Buss and Minnes, 2013, Lemma 2.20, pp. 8-9].

where the last inequality holds because of a classical theorem by Ville [1939] (known as Kolmogorov's Inequality), which states that, for any (super)martingale $d$, any string $\sigma$ and any prefix-free set $\mathcal{S} \subseteq \{0,1\}^*$ of extensions of $\sigma$,

$$\sum_{\tau \in \mathcal{S}} 2^{-|\tau|} \cdot d(\tau) \leq 2^{-|\sigma|} \cdot d(\sigma).$$

This shows that no $X \in \mathcal{N}$ is Martin-Löf random. Hence, Martin-Löf randomness implies Ex-randomness, which entails that the collection of non-Ex-random sequences has (effective) measure zero, while the collection of Ex-random sequences has (effective) measure one. Since every P1-strategy can be converted into an Ex-strategy, the same holds for P1-random sequences.

As observed earlier, both Definition 5.2.1 and Definition 5.2.3 admit of probabilistic strategies that might reach a stage after which no more bets are made or, more generally, after which the probability of never placing another bet is positive. This can be disallowed by only considering probabilistic strategies which *always eventually bet with probability one*: i.e., strategies $A$ such that, for all $\pi \in \{\mathsf{b},\mathsf{w}\}^*$ and $\sigma \in \{0,1\}^{\#\mathsf{bets}(\pi)}$,

$$P_A(\pi,\sigma) \cdot \prod_{j \in \mathbb{N}} \big(1 - p_A(\pi \mathsf{w}^j, \sigma)\big) = 0.$$

Such a restriction gives rise to weaker versions of P1-randomness and Ex-randomness.

**Definition 5.2.4** (Weak P1-randomness and Ex-randomness)**.** Let $X \in \{0,1\}^\omega$. Then,

(a) $X$ is said to be *weakly* P1*-random* if and only if no probabilistic strategy which always eventually bets with probability one is a P1-strategy for $X$;

(b) $X$ is said to be *weakly* Ex*-random* if and only if no probabilistic strategy which always eventually bets with probability one is an Ex-strategy for $X$.

We can then state the following characterisation results for Martin-Löf randomness, partial computable randomness and computable randomness (which constitute the core of Buss and Minnes' paper on probabilistic algorithmic randomness).

**Theorem 5.2.5** (Buss and Minnes [2013])**.** *Let $X \in \{0,1\}^\omega$. Then,*

(a) *$X$ is* Ex*-random if and only if it is Martin-Löf random;*

(b) *$X$ is* P1*-random if and only if it is partial computably random;*

(c) *$X$ is weakly* P1*-random if and only if it is weakly* Ex*-random if and only if it is computably random.*

Theorem 5.2.5 provides the sought-after characterisation of Martin-Löf randomness in terms of computable betting strategies, thus offering a viable response to Schnorr's critique. However, the usefulness of Buss and Minnes' probabilistic framework does not end here, as we will see in the remainder of this chapter.

## 5.3 KP1-randomness and KEx-randomness

In this section, we will introduce some variants of the notions of probabilistic algorithmic randomness proposed by Buss and Minnes by appealing to Schnorr's concept of *success at a computable rate* (§ 3.3.3). We will then prove that these 'more effective' versions of (weak) P1-randomness and (weak) Ex-randomness each fit into the standard hierarchy of randomness concepts.

**Definition 5.3.1** (Success at a computable rate)**.** Let $X \in \{0,1\}^\omega$ and $\Pi \in \{\mathsf{b},\mathsf{w}\}^\omega$. A probabilistic strategy $A$ is said to *succeed against* $X$ *along* $\Pi$ *at a computable rate* if and only if there is a computable unbounded non-decreasing function $h : \mathbb{N} \to \mathbb{N}$ such that

$$\exists^\infty n \in \mathbb{N} : C_A^X(\Pi \upharpoonright n) \geq h(n).$$

We can now formalise the notion of a strategy which allows a gambler to become infinitely rich at a computable rate with probability one.

**Definition 5.3.2** (KP1-strategy)**.** Let $X \in \{0,1\}^\omega$. A probabilistic strategy $A$ is said to be a KP1-*strategy* for $X$ if and only if there is a computable unbounded non-decreasing function $h : \mathbb{N} \to \mathbb{N}$ such that

$$\mu_A^X\left(\{\Pi \in \{\mathsf{b},\mathsf{w}\}^\omega : \exists^\infty n \in \mathbb{N} \text{ with } C_A^X(\Pi \upharpoonright n) \geq h(n)\}\right) = 1.$$

With these definitions at hand, we can characterise the following randomness notions.

**Definition 5.3.3** (KP1-randomness)**.** Let $X \in \{0,1\}^\omega$. Then,

(a) $X$ is said to be KP1-*random* if and only if there is no probabilistic strategy $A$ such that $A$ is a KP1-strategy for $X$;

(b) $X$ is said to be *weakly* KP1-*random* if and only if there is no probabilistic strategy $A$ which always eventually bets with probability one such that $A$ is a KP1-strategy for $X$.

Clearly, P1-randomness, as defined in Section 5.2, implies KP1-randomness, which, in turn, implies weak KP1-randomness. Since the set of P1-random sequences was shown to have (effective) measure one, so do the collections of KP1-random sequences and of weakly KP1-random sequences.

Next, we formalise the notion of a strategy gaining unbounded expected capital at a computable rate.

**Definition 5.3.4** (KEx-strategy)**.** Let $X \in \{0,1\}^\omega$. A probabilistic strategy $A$ is said to be a KEx-strategy for $X$ if and only if there is a computable unbounded non-decreasing function $h : \mathbb{N} \to \mathbb{N}$ such that

$$\exists^\infty n \in \mathbb{N} : \mathsf{Ex}_A^X(n) \geq h(n).$$

This allows to define two more randomness concepts, akin to Ex-randomness and weak Ex-randomness, respectively.

**Definition 5.3.5** (KEx-randomness)**.** Let $X \in \{0,1\}^\omega$. Then,

(a) $X$ is said to be KEx-*random* if and only if there is no probabilistic strategy $A$ such that $A$ is a KEx-strategy for $X$;

(b) $X$ is said to be *weakly* KEx-*random* if and only if there is no probabilistic strategy $A$ which always eventually bets with probability one such that $A$ is a KEx-strategy for $X$.

Again, we have that Ex-randomness, as defined in Section 5.2, implies KEx-randomness, which, in turn, implies weak KEx-randomness. As the set of Ex-random sequences was shown to have (effective) measure one, so do the collections of KEx-random sequences and of weakly KEx-random sequences.

We now show that weak KP1-randomness (and, *a fortiori*, KP1-randomness) implies Schnorr randomness[61].

**Proposition 5.3.6.** *Let $X \in \{0,1\}^\omega$. If $X$ is weakly* KP1*-random, then it is Schnorr random.*

*Proof.* Suppose that $X$ is not Schnorr random. Then, by Theorem 3.3.6, there are a computable rational-valued martingale $d$ and a computable unbounded non-decreasing function $h : \mathbb{N} \to \mathbb{N}$ such that $d(X \upharpoonright n) \geq h(n)$ for infinitely many $n \in \mathbb{N}$. W.l.o.g., we can assume that $d$ is normed. Martingale $d$ can then be readily employed to construct a probabilistic strategy $A_d$. For all $\pi \in \{\mathsf{b},\mathsf{w}\}^*$ and $\sigma \in \{0,1\}^{\#\mathsf{bets}(\pi)}$, set $p_{A_d}(\pi,\sigma) = 1$; moreover, for each $n \in \mathbb{N}$ and $\sigma \in \{0,1\}^n$, set $q_{A_d}(\mathsf{b}^n,\sigma) = d(\sigma 0) - d(\sigma)$. All other values of $p_{A_d}$ and $q_{A_d}$ can be set arbitrarily. Then, there is exactly one path $\Pi$ through the computation tree $\{\mathsf{b},\mathsf{w}\}^\omega$ with non-zero probability: namely, $\Pi = \varepsilon\mathsf{bbbb}...$ In particular, $\mu^X_{A_d}(\{\Pi\}) = 1$. Moreover, along $\Pi$, the capital accumulated by $A_d$ equals the capital accrued by martingale $d$. This can be seen via a simple argument by induction. First of all, we have that $C_{A_d}(\varepsilon,\varepsilon) = 1 = d(\varepsilon)$ by the definition of a probabilistic strategy and because $d$ is a normed martingale. Now, for the inductive step, suppose that $C^X_{A_d}(\Pi \upharpoonright n) = d\big(X \upharpoonright \#\mathsf{bets}(\Pi \upharpoonright n)\big)$. For each $\sigma \in \{0,1\}^*$, let $\mathsf{stake}(d,\sigma)$ denote the quantity $d(\sigma 0) - d(\sigma)$. Then, if $X\big(\#\mathsf{bets}(\Pi \upharpoonright n+1)\big) = 0$,

$$d\big(X \upharpoonright \#\mathsf{bets}(\Pi \upharpoonright n+1)\big) = d\big(X \upharpoonright \#\mathsf{bets}(\Pi \upharpoonright n)\big) + \mathsf{stake}\big(d, X \upharpoonright \#\mathsf{bets}(\Pi \upharpoonright n)\big),$$

while if $X\big(\#\mathsf{bets}(\Pi \upharpoonright n+1)\big) = 1$,

$$d\big(X \upharpoonright \#\mathsf{bets}(\Pi \upharpoonright n+1)\big) = d\big(X \upharpoonright \#\mathsf{bets}(\Pi \upharpoonright n)\big) - \mathsf{stake}\big(d, X \upharpoonright \#\mathsf{bets}(\Pi \upharpoonright n)\big).$$

Moreover, since $A_d$ always bets, we have that if $X\big(\#\mathsf{bets}(\Pi \upharpoonright n+1)\big) = 0$, then

[61] The argument used in this proof is analogous to the left-to-right direction of the proof of Theorem 5.1 from [Buss and Minnes, 2013].

$$C^X_{A_d}(\Pi \upharpoonright n+1) = C^X_{A_d}(\Pi \upharpoonright n) + \mathsf{stake}\big(d, X \upharpoonright \#\mathsf{bets}(\Pi \upharpoonright n)\big)$$

because of the definition of $q_{A_d}$; on the other hand, if $X\big(\#\mathsf{bets}(\Pi \upharpoonright n+1)\big) = 1$, then

$$C^X_{A_d}(\Pi \upharpoonright n+1) = C^X_{A_d}(\Pi \upharpoonright n) - \mathsf{stake}\big(d, X \upharpoonright \#\mathsf{bets}(\Pi \upharpoonright n)\big),$$

again because of the definition of $q_{A_d}$. Since, by the induction hypothesis, $C^X_{A_d}(\Pi \upharpoonright n) = d\big(X \upharpoonright \#\mathsf{bets}(\Pi \upharpoonright n)\big)$, we can conclude that the values of $C_{A_d}$ and $d$ always coincide along $\Pi$. This implies that there are infinitely many $n \in \mathbb{N}$ such that $C^X_{A_d}(\Pi \upharpoonright n) \geq h(n)$. Thus, $A_d$ is a $\mathsf{KP1}$-strategy for $X$ which always eventually bets with probability one: hence, $X$ is not weakly $\mathsf{KP1}$-random. □

Next, we show that Schnorr randomness is equivalent to weak $\mathsf{KEx}$-randomness.

**Theorem 5.3.7.** *Let $X \in \{0,1\}^\omega$. Then, $X$ is weakly $\mathsf{KEx}$-random if and only if it is Schnorr random.*

*Proof.* ($\Rightarrow$) Suppose that $X$ is not Schnorr random. Then, there are a computable rational-valued martingale $d$ and a computable unbounded non-decreasing function $h : \mathbb{N} \to \mathbb{N}$ such that $d(X \upharpoonright n) \geq h(n)$ for infinitely many $n \in \mathbb{N}$. From $d$, construct the same probabilistic strategy $A_d$ described in the proof of Proposition 5.3.6. Since $d$ is a total function and all bets are placed with probability one, $A_d$ always eventually bets with probability one. Moreover, for all $n \in \mathbb{N}$, we have that

$$\mathsf{Ex}^X_{A_d}(n) = \sum_{\pi \in R(n)} P^X_{A_d}(\pi) \cdot C^X_{A_d}(\pi) = C^X_{A_d}(\mathsf{b}^n) = d(X \upharpoonright n).$$

Hence, there are infinitely many $n \in \mathbb{N}$ such that $\mathsf{Ex}^X_{A_d}(n) \geq h(n)$, which means that $A_d$ is a $\mathsf{KEx}$-strategy for $X$. Thus, $X$ is not weakly $\mathsf{KEx}$-random.

($\Leftarrow$) Suppose that $X$ is not weakly $\mathsf{KEx}$-random. Then, there is a probabilistic strategy $A$ which always eventually bets with probability one that is a $\mathsf{KEx}$-strategy for $X$. Define $d_A : \{0,1\}^* \to \mathbb{R}^{\geq 0}$ as $d_A(\sigma) = \mathsf{Ex}^\sigma_A$ for all $\sigma \in \{0,1\}^*$. Then, we get that $d_A(X \upharpoonright n) \geq h(n)$ for infinitely many $n \in \mathbb{N}$. Lemma 4.6 from [Buss and Minnes, 2013] gives us that $d_A$ is a supermartingale. Since $A$ always eventually bets with probability one, however, $d_A$ is in fact a martingale, as shown by Lemma 5.4 from [Buss and Minnes, 2013]. Now, all that is left to do is showing that $d_A$ is computable. This can be done by employing the same argument used in the proof of Theorem 5.2 in [Buss and Minnes, 2013]. Since $d_A$ is a normed martingale, it can be easily shown by induction that, for all $n \in \mathbb{N}$,

$$\sum_{\tau \in \{0,1\}^n} d_A(\tau) = 2^n. \qquad (\text{❖})$$

Then, define the approximation to $d_A$ at level $M > 0$ to be

$$d^M_A(\tau) = \sum_{\pi \in R(|\tau|) \colon \#\mathsf{waits}(\pi) < M} P_A(\pi, \tau) \cdot C_A(\pi, \tau).$$

We have that $d_A^M(\tau)$ is a finite sum of computable terms which approaches the value $d_A(\tau)$ from below. For $d_A$ to be computable, we must be able to compute $d_A(\sigma)$ to within $\varepsilon$ of the true value, for each $\sigma \in \{0,1\}^*$ and $\varepsilon > 0$. To this end, compute $\sum_{\tau \in \{0,1\}^{|\sigma|}} d_A^M(\tau)$ for increasingly large values of $M$, until an $M'$ is found which renders the previous sum greater than $2^{|\sigma|} - \varepsilon$. By (❖), we then have that $M'$ puts $d_A^{M'}(\sigma)$ within $\varepsilon$ of $d_A(\sigma)$. Hence, $d_A$ is indeed computable and $X$ is not Schnorr random. □

So, we now know that KEx-randomness is at most as strong as Martin-Löf randomness and at least as strong as Schnorr randomness. Next, we will show that KEx-randomness is in fact equivalent to Martin-Löf randomness[62].

**Theorem 5.3.8.** *Let $X \in \{0,1\}^\omega$. If $X$ is KEx-random, then it is Martin-Löf random.*

*Proof.* Suppose that $X$ is not Martin-Löf random. Then, w.l.o.g., there is a nested Martin-Löf test $(\mathcal{U}_n)_{n\in\mathbb{N}}$ such that $X \in \bigcap_{n\in\mathbb{N}} \mathcal{U}_n$. For simplicity, we consider only the $\mathcal{U}_n$'s with $n \geq 1$. Then, there exists a sequence $(\mathcal{S}_n)_{n\in\mathbb{N}^{>0}}$ of infinite prefix-free subsets of $\{0,1\}^*$ such that, for each $n > 0$, $\mathcal{S}_n = (\sigma_{n,i})_{i\in\mathbb{N}}$ and $\mathcal{U}_n = \bigcup \{[\![\sigma_{n,i}]\!] : i \in \mathbb{N}\}$. Now, consider the following probabilistic strategy $A$. We run the algorithm that computably enumerates the $\sigma_{n,i}$'s. Suppose that we have already observed some prefix $X \upharpoonright \ell$ of $X$ ($\ell \in \mathbb{N}$), and that we are now enumerating the $\sigma_{\ell+1,i}$'s ($i \in \mathbb{N}$). When $\sigma_{\ell+1,i'}$ is enumerated, strategy $A$ bets all-or-nothing with probability $2^{\ell+1-|\sigma_{\ell+1,i'}|}$ that $X(k) = \sigma_{\ell+1,i'}(k)$ for $\ell < k \leq |\sigma_{\ell+1,i'}|$. More precisely, suppose that $\pi$ is a minimal node for which $p_A(\pi,\sigma)$ and $q_A(\pi,\sigma)$ have not been defined yet and such that $\#\mathsf{bets}(\pi) = \ell$. Let $p_A(\pi,\sigma) = 2^{\ell+1-|\sigma_{\ell+1,0}|}$. Since $|\sigma_{\ell+1,0}| \geq \ell+1$, $p_A(\pi,\sigma) \leq 1$. For $i \geq 1$, on the other hand, let

$$p_A(\pi\mathsf{w}^i,\sigma) = \frac{2^{\ell+1-|\sigma_{\ell+1,i}|}}{\prod_{k=0}^{i-1}\big(1 - p_A(\pi\mathsf{w}^k,\sigma)\big)} \leq 1.$$

Then, for all $i \in \mathbb{N}$ and $1 \leq k \leq |\sigma_{\ell+1,i}| - \ell$, let $p_A(\pi\mathsf{w}^i\mathsf{b}^k,\sigma) = 1$. Moreover, for $i \in \mathbb{N}$ and $0 \leq k \leq |\sigma_{\ell+1,i}| - \ell$, set

$$q_A(\pi\mathsf{w}^i\mathsf{b}^k,\sigma) = \begin{cases} C_A(\pi\mathsf{w}^i\mathsf{b}^k,\sigma) & \text{if } \sigma_{\ell+1,i}(\ell+k) = 0; \\ -C_A(\pi\mathsf{w}^i\mathsf{b}^k,\sigma) & \text{if } \sigma_{\ell+1,i}(\ell+k) = 1. \end{cases}$$

Clearly, both $p_A$ and $q_A$ are computable, so $A$ is indeed a computable probabilistic strategy. Since $X \in \bigcap_{n\in\mathbb{N}} \mathcal{U}_n$, there is a unique c.e. sequence $(\sigma_{n,i_n})_{n\in\mathbb{N}}$ such that $\sigma_{n,i_n} \sqsubset X$ for each $n \in \mathbb{N}$. Given an arbitrary $n$, we do not necessarily have that $|\sigma_{n,i_n}| < |\sigma_{n+1,i_{n+1}}|$. So, we extract from $(\sigma_{n,i_n})_{n\in\mathbb{N}}$ an infinite subsequence $\sigma_{n_1,i_1}, \sigma_{n_2,i_2}, \ldots$ with this property in the following way: set $n_1 = 1$ and, for each $j \geq 1$,

[62] The first part of the proof of Theorem 5.3.8 is analogous to the proof of Theorem 4.2 from [Buss and Minnes, 2013].

set $n_{j+1} = |\sigma_{n_j,i_j}| + 1$. Then, $\sigma_{1,i_1} \sqsubset \sigma_{2,i_2} \sqsubset \sigma_{3,i_3} \sqsubset ... \sqsubset X$. Now, define $\ell_0 = 0$ and $\ell_j = |\sigma_{n_j,i_j}|$; then, $n_{j+1} = \ell_j + 1$ and $\ell_j \geq n_j$. Consider the following computation paths $\pi_k$, for each $k \geq 1$,

$$\pi_k = \mathsf{w}^{i_1}\mathsf{b}^{\ell_1}\mathsf{w}^{i_2}\mathsf{b}^{\ell_2-\ell_1}...\,\mathsf{w}^{i_k}\mathsf{b}^{\ell_k-\ell_{k-1}}.$$

When playing against $X$, each computation path $\pi_k$ is such that every bet placed is successful. Since each $\pi_k$ involves $\ell_k$ bets, we then have that $C_A^X(\pi_k) = 2^{\ell_k}$. So,

$$\begin{aligned}
\mathsf{Ex}_A^X(\ell_k) &= \sum_{\pi \in R(\ell_k)} P_A^X(\pi) \cdot C_A^X(\pi) && \text{by Definition 5.2.2} \\
&\geq P_A^X(\pi_k) \cdot C_A^X(\pi_k) \\
&= 2^{\ell_k} \cdot \prod_{j=1}^{k} 2^{n_j - \ell_j} && \text{by the definition of } p_A \\
&= 2^{\ell_{n_1}} \cdot \prod_{j=1}^{k-1} 2^{n_{j+1} - \ell_j} \\
&= 2^k && \text{because } n_1 = 1 \text{ and } n_{j+1} = \ell_j + 1.
\end{aligned}$$

So, $\mathsf{Ex}_A^X(\ell_k) \geq 2^k$ for infinitely many $\ell_k \in \mathbb{N}$. We then define the following function $h : \mathbb{N} \to \mathbb{N}$:

$$h(0) = 0$$
$$h(m) = \begin{cases} 2^k & \text{if there is some } k \text{ such that } m = \ell_k; \\ h(m-1) & \text{otherwise.} \end{cases}$$

We argue that $h$ is computable. Take $m \geq 1$. There are $2^m$ strings of length $m$. For each such string $\sigma_m$, we know that we can only have that $\sigma_m = \sigma_{n_j,i_j}$ for $n_j \leq m$. So, for each of them, we check whether $\sigma_m = \sigma_{n_j,i_j}$ starting with $n_j = 1$ and up to $n_j = m$. If, during any of these comparisons, we find that one of the $\sigma_m$'s is in fact a $\sigma_{n_j,i_j}$, we set $h(m) = 2^j$. Else, we set $h(m) = h(m-1)$. Since we only ever have to check finitely many strings, this procedure can be carried out computably. We also have that $h$ is non-decreasing and unbounded. Hence, $A$ is a $\mathsf{KEx}$-strategy for $X$, which means that $X$ is not $\mathsf{KEx}$-random. □

Since $\mathsf{Ex}$-randomness implies $\mathsf{KEx}$-randomness, we then get the following corollary.

**Corollary 5.3.9.** *Let $X \in \{0,1\}^\omega$. Then, $X$ is Martin-Löf random if and only if it is $\mathsf{Ex}$-random if and only if it is $\mathsf{KEx}$-random.*

We can modify Definition 5.3.1 by imposing the stronger requirement that $C_A^X(\Pi \upharpoonright n) \geq h(n)$ for *all* $n \in \mathbb{N}$. Then, Definition 5.3.3 and Definition 5.3.5 can be modified accordingly. We call the resulting notions of randomness (weak) $\mathsf{WP1}$-randomness and (weak) $\mathsf{WEx}$-randomness. We can then characterise Kurtz randomness in terms of probabilistic strategies.

**Theorem 5.3.10.** *Let $X \in \{0,1\}^\omega$. Then, $X$ is weakly* WEx*-random if and only if it is Kurtz random.*

The proof of this result is essentially the same as that of Theorem 5.3.7: the only difference is that, in this case, we exploit the martingale-based characterisation of Kurtz randomness provided by Theorem 3.3.7.

We can also show that Kurtz randomness coincides with weak WP1-randomness.

**Theorem 5.3.11.** *Let $X \in \{0,1\}^\omega$. Then, $X$ is weakly* WP1*-random if and only if it is Kurtz random.*

*Proof.* ($\Rightarrow$) The proof of this direction is analogous to that of Proposition 5.3.6.

($\Leftarrow$) Suppose that $X$ is not weakly WP1-random. Then, there is a probabilistic strategy $A$ which always eventually bets that is a WP1-strategy for $X$: i.e., there is a computable unbounded non-decreasing function $h : \mathbb{N} \to \mathbb{N}$ such that

$$\mu_A^X\left(\{\Pi \in \{\mathsf{b},\mathsf{w}\}^\omega : C_A^X(\Pi \upharpoonright n) \geq h(n) \text{ for all } n \in \mathbb{N}\}\right) = 1.$$

For ease of notation, we will denote the set

$$\{\Pi \in \{\mathsf{b},\mathsf{w}\}^\omega : C_A^X(\Pi \upharpoonright n) \geq h(n) \text{ for all } n \in \mathbb{N}\}$$

as $\mathcal{S}_A^X$. Now, fix some arbitrary $m \in \mathbb{N}$. We then have that

$$\begin{aligned}\mathcal{S}_A^X &\subseteq \bigcup\left\{[\![\pi]\!] \subseteq \{\mathsf{b},\mathsf{w}\}^\omega : \pi \in R(m) \text{ and there is some } \Pi \in \mathcal{S}_A^X \text{ with } \pi \sqsubset \Pi\right\}\\ &\subseteq \bigcup\left\{[\![\pi]\!] \subseteq \{\mathsf{b},\mathsf{w}\}^\omega : \pi \in R(m)\right\}\\ &= [\![R(m)]\!],\end{aligned}$$

where $R(m) \subseteq \{\mathsf{b},\mathsf{w}\}^*$ is the set of all finite computation paths that involve exactly $m$ bets and such that the last move was a bet. Now, we have that

$$\begin{aligned}\mathsf{Ex}_A^X(m) &= \sum_{\pi \in R(m)} P_A^X(\pi) \cdot C_A^X(\pi)\\ &= \sum_{\pi \in R(m)} \mu_A^X([\![\pi]\!]) \cdot C_A^X(\pi) && \text{by the def. of } \mu_A^X\\ &\geq \sum_{\substack{\pi \in R(m):\\ \exists \Pi \in \mathcal{S}_A^X\\ \text{with } \pi \sqsubset \Pi}} \mu_A^X([\![\pi]\!]) \cdot C_A^X(\pi)\\ &\geq \sum_{\substack{\pi \in R(m):\\ \exists \Pi \in \mathcal{S}_A^X\\ \text{with } \pi \sqsubset \Pi}} \mu_A^X([\![\pi]\!]) \cdot h(m) && \text{by the def. of } \mathcal{S}_A^X \text{ and because } h \text{ is non-decreasing}\end{aligned}$$

$$
\begin{aligned}
&\geq h(m) \cdot \sum_{\substack{\pi \in R(m): \\ \exists \Pi \in \mathcal{S}_A^X \\ \text{with } \pi \sqsubset \Pi}} \mu_A^X([\![\pi]\!]) && \text{because } h(m) \text{ does not depend on } \pi \\
&\geq h(m) \cdot \mu_A^X(\mathcal{S}_A^X) \\
&= h(m) && \text{because } \mu_A^X(\mathcal{S}_A^X) = 1 \text{ by assumption}
\end{aligned}
$$

Since $m$ was chosen arbitrarily, we can conclude that $\mathsf{Ex}_A^X(n) \geq h(n)$ for all $n \in \mathbb{N}$. So, by the right-to-left direction of Theorem 5.3.10, we have that $X$ is not Kurtz random. □

**Corollary 5.3.12.** *Let $X \in \{0,1\}^\omega$. Then, $X$ is Kurtz random if and only if it is weakly* WEx*-random if and only if it is weakly* WP1*-random.*

## 5.4 Non-monotonic P1-randomness and Ex-randomness

In this section, we further extend Buss and Minnes' framework by studying probabilistic non-monotonic betting strategies. We propose non-monotonic variants of P1-randomness and Ex-randomness, and we show that they are equivalent to Kolmogorov-Loveland randomness and Martin-Löf randomness, respectively.

Recall that, in Definition 3.3.2, we are confronted with a 'multitasking' betting strategy. In the probabilistic setting, we split the work between three different functions.

**Definition 5.4.1** (Probabilistic non-monotonic strategy)**.** A *probabilistic non-monotonic strategy* $A$ is a triple $\langle n_A, p_A, q_A \rangle$, where (i) $n_A : \{\mathsf{b},\mathsf{w}\}^* \times \{0,1\}^\omega \to \mathbb{N}$ is a computable position-selection function, (ii) $p_A : \{\mathsf{b},\mathsf{w}\}^* \times \{0,1\}^\omega \to \mathbb{Q} \cap [0,1]$ a computable probability function, and (iii) $q_A : \{\mathsf{b},\mathsf{w}\}^* \times \{0,1\}^* \to \mathbb{Q}$ a computable stake function.

We require that function $n_A$ be to some extent oblivious, in the sense that its decision of what position to bet on next should depend solely on the bits previously observed and not on the particular series of bet/wait moves chosen so far. Formally, for any $\sigma \in \{0,1\}^*$ and $\pi, \pi' \in \{\mathsf{b},\mathsf{w}\}^*$ such that $\#\mathsf{bets}(\pi) = \#\mathsf{bets}(\pi') = |\sigma|$, $n_A(\pi, \sigma) = n_A(\pi', \sigma)$. This restriction ensures that, given some sequence $X \in \{0,1\}^\omega$, strategy $A$ induces a unique transformation of $X$ along all computation paths which include infinitely many bet moves: that is, the infinite sequence corresponding to all of the bits from $X$ sequentially selected by $n_A$ is the same for all computation paths in which a bet is always eventually placed. We denote as $X_A$ the sequence resulting from applying $n_A$ to $X$, whenever such sequence is defined.

One can see that the definitions of cumulative probability (Definition 5.1.2) and capital (Definition 5.1.3) from Section 5.1 carry over to this probabilistic non-monotonic context.

Then, the probability of success of a probabilistic non-monotonic strategy $A$ when playing against a sequence $X$ is given by

$$\mu_A^{X_A}\left(\{\Pi \in \{\mathsf{b},\mathsf{w}\}^\omega : \lim_{n\to\infty} C_A^{X_A}(\Pi \upharpoonright n) = \infty\}\right).$$

We can then define the non-monotonic counterpart of P1-randomness.

**Definition 5.4.2** (Non-monotonic P1-randomness)**.** Let $X \in \{0,1\}^\omega$.

(a) A probabilistic non-monotonic strategy $A$ is said to be a P1-strategy for $X$ if and only if

$$\mu_A^{X_A}\left(\{\Pi \in \{\mathsf{b},\mathsf{w}\}^\omega : \lim_{n\to\infty} C_A^{X_A}(\Pi \upharpoonright n) = \infty\}\right) = 1.$$

(b) $X$ is said to be *non-monotonically* P1*-random* if and only if no probabilistic non-monotonic strategy is a P1-strategy for $X$.

Since non-monotonic betting strategies are a generalisation of monotonic betting strategies, we have that non-monotonic P1-randomness implies P1-randomness.

**Proposition 5.4.3.** *Let $X \in \{0,1\}^\omega$. If $X$ is non-monotonically* P1*-random, then it is* P1*-random.*

Given our definition of probabilistic non-monotonic betting strategies—where the values of function $n_A$ are independent of the particular computation path followed so far—the non-monotonic variant of Ex-randomness can be characterised as follows.

**Definition 5.4.4** (Non-monotonic Ex-randomness)**.** Let $X \in \{0,1\}^\omega$.

(a) A probabilistic non-monotonic strategy $A$ is an Ex-strategy for $X$ if and only if

$$\lim_{n\to\infty} \mathsf{Ex}_A^{X_A}(n) = \infty;$$

(b) $X$ is said to be non-monotonically Ex*-random* if and only if no probabilistic non-monotonic strategy is an Ex-strategy for $X$.

Once again, this definition immediately gives us that non-monotonic Ex-randomness implies Ex-randomness.

**Proposition 5.4.5.** *Let $X \in \{0,1\}^\omega$. If $X$ is non-monotonically* Ex*-random, then it is* Ex*-random.*

We now show that non-monotonic P1-randomness coincides with Kolmogorov-Loveland randomness.

**Theorem 5.4.6.** *Let $X \in \{0,1\}^\omega$. Then, $X$ is non-monotonically* P1*-random if and only if it is Kolmogorov-Loveland random.*

*Proof.* ($\Rightarrow$) Suppose that $X$ is not Kolmogorov-Loveland random. By Definition 3.3.3, there is a computable non-monotonic betting strategy $b$ such that $\lim_{n\to\infty} C_b^X(\chi^{(n)}) = \infty$. This holds even though Definition 3.3.3 is technically given in terms of the limsup, rather than the limit of $C_b^X(\chi^{(n)})$: this is because, as already mentioned in § 3.3.1, any martingale for which the limsup equals infinity can be transformed into another martingale which succeeds on exactly the same sequences, but whose *limit* equals infinity. Now, strategy $b$ provides us with a probabilistic non-monotonic strategy $B$. For each $\sigma \in \{0,1\}^*$ and $\pi \in \{\mathsf{b},\mathsf{w}\}^*$ with $\#\mathsf{bets}(\pi) = |\sigma|$, we proceed as follows:

(i) if $b(\sigma) = (k, \mathsf{scan})$, then we set $n_B(\pi,\sigma) = k$, $p_B(\pi,\sigma) = 1$ and $q_B(\pi,\sigma) = 0$;

(ii) if $b(\sigma) = (k,\rho)$ (with $\rho \in [-1,1]$), then we set $n_B(\pi,\sigma) = k$, $p_B(\pi,\sigma) = 1$ and $q_B(\pi,\sigma) = \rho \cdot C_b(\sigma)$.

Then, there is exactly one path $\Pi'$ through the computation tree $\{\mathsf{b},\mathsf{w}\}^\omega$ that gets assigned non-zero probability: namely, $\Pi' = \varepsilon\mathsf{bbbb}...$ In particular, $\mu_B^{X_B}(\{\Pi'\}) = 1$. Moreover, along $\Pi'$, the capital hoarded by $B$ is equal to the capital accumulated by martingale $C_b$. So, we have that $\lim_{n\to\infty} C_B^{X_B}(\Pi' \upharpoonright n) = \infty$. Hence, $\mu_B^{X_B}\Big(\{\Pi \in \{\mathsf{b},\mathsf{w}\}^\omega : \lim_{n\to\infty} C_B^{X_B}(\Pi \upharpoonright n) = \infty\}\Big) = 1$. Thus, $B$ is a non-monotonic $\mathsf{P1}$-strategy for $X$, which means that $X$ is not non-monotonically $\mathsf{P1}$-random.

($\Leftarrow$) Suppose that $X$ is not non-monotonically $\mathsf{P1}$-random. Then, there is a probabilistic non-monotonic strategy $A = \langle n_A, p_A, q_A\rangle$ that is a $\mathsf{P1}$-strategy for $X$: i.e.,

$$\mu_A^{X_A}\Big(\{\Pi \in \{\mathsf{b},\mathsf{w}\}^\omega : \lim_{n\to\infty} C_A^{X_A}(\Pi \upharpoonright n) = \infty\}\Big) = 1.$$

We then have that $\lim_{n\to\infty} \mathsf{Ex}_A^{X_A}(n) = \infty$ and that $A$ eventually bets on sequence $X_A$ with probability one. We also know that, for any $\sigma \in \{0,1\}^*$ and $\pi,\pi' \in \{\mathsf{b},\mathsf{w}\}^*$ with $\#\mathsf{bets}(\pi) = \#\mathsf{bets}(\pi') = |\sigma|$, $n_A(\pi,\sigma) = n_A(\pi',\sigma)$. So, the argument used in the proof of Theorem 6.2 from [Buss and Minnes, 2013] can be adapted to our case to construct a partial computable deterministic non-monotonic betting strategy $a : \{0,1\}^* \to \mathbb{N} \times \big(\{\mathsf{scan}\} \cup [-1,1]\big)$ which is defined for all initial segments of $X_A$, and which also succeeds on $X_A$—in the sense that $\lim_{n\to\infty} C_a^X(\chi^{(n)}) = \infty$. For each $\sigma \in \{0,1\}^*$, simply set $a(\sigma) = \Big(n_A(\mathsf{b}^{|\sigma|},\sigma), \dfrac{C_a(\sigma 0) - C_a(\sigma)}{C_a(\sigma)}\Big)$, where $C_a$ is defined as in the proof Theorem 6.2 (equation (17)). Then, $C_a$ is a partial computable *supermartingale* which non-monotonically succeeds on $X$, and which can be converted into a partial computable *martingale* which non-monotonically succeeds on $X$. This can be done via the *savings trick* discussed in § 3.3.1. Since the notion of Kolmogorov-Loveland randomness is left unaltered if one replaces computable non-monotonic betting strategies with partial computable non-monotonic betting strategies in Definition 3.3.3, we can then conclude that $X$ is not Kolmogorov-Loveland random. □

It is also worth noting that the proof of Theorem 4.1 from [Buss and Minnes, 2013] can be adapted to show that Martin-Löf randomness implies non-monotonic Ex-randomness.

**Proposition 5.4.7.** *Let $X \in \{0,1\}^\omega$. If $X$ is Martin-Löf random, then it is non-monotonically* Ex*-random.*

The proof of Proposition 5.4.7 is again by contraposition: a probabilistic non-monotonic strategy whose expected capital is unbounded can be exploited to construct a Martin-Löf test.

**Corollary 5.4.8.** *Let $X \in \{0,1\}^\omega$. The following are equivalent:*

(1) *$X$ is Martin-Löf random;*

(2) *$X$ is* Ex*-random;*

(3) *$X$ is non-monotonically* Ex*-random;*

(4) *$X$ is* KEx*-random.*

The equivalence of Martin-Löf randomness and non-monotonic Ex-randomness is reminiscent of the fact that when one replaces computable non-monotonic betting strategies by merely computably enumerable non-monotonic betting strategies in the definition of Kolmogorov-Loveland randomness, the resulting notion is once again Martin-Löf randomness.

We conclude this section with some remarks on our definitions of non-monotonic P1-randomness and Ex-randomness, and on possible variations thereof. An interesting question, for example, is what happens if we lift the restriction that the values of the position-selection function $n_A$ depend uniquely on the binary strings previously observed and allow $n_A$ to make decisions on the basis of previous bet/wait moves, too.

Relaxing Definition 5.4.1 clearly gives rise to a more general notion of probabilistic non-monotonic betting strategies. However, a probabilistic non-monotonic strategy $A$ defined in this way does not play against one unique sequence $X_A$ or string $\chi_A^X$ (depending on whether the number of bets along a given computation path is infinite or finite), determined by the positions selected by $n_A$ from some $X \in \{0,1\}^\omega$, no matter what computation path ends up being chosen during a run of the strategy. Now, each computation path $\Pi \in \{\mathsf{b},\mathsf{w}\}^\omega$ determines a potentially different transformation of $X$ into either an infinite sequence $X_A(\Pi)$ or a finite string $\sigma_A^X(\Pi)$. This is because, for pairs of the form $(\pi,\sigma)$ and $(\pi',\sigma)$, with $\pi \neq \pi'$, $\#\mathsf{bets}(\pi) = \#\mathsf{bets}(\pi') = |\sigma|$ and $\sigma \sqsubset X$, the function $n_A$ may decide to bet on the on the $k$-th bit of $X$ upon seeing $(\pi,\sigma)$ and on the $k'$-th bit of $X$ after observing $(\pi',\sigma)$, where $k \neq k'$. So, each run of the strategy induces a transformation of $X$ into a different sequence/string.

In this scenario, it is not immediately clear how to define the probability of success of a probabilistic non-monotonic strategy $A$. The measure $\mu_A$ is defined

with respect to a fixed sequence in $\{0,1\}^\omega$, whereas here we are facing a potentially uncountable number of sequences induced by the various computation paths along which the capital values tend to infinity. So, providing a meaningful definition of non-monotonic P1-randomness in this more general setting is not as straightforward as before.

The situation appears to be somewhat simpler when one considers the notion of expected capital. Let $A$ be a probabilistic non-monotonic strategy in the more general sense discussed above, and let $X \in \{0,1\}^\omega$. Then, let $\sigma(\pi)$ denote the string of bits sequentially selected by $A$ from $X$ along the finite sequence of moves encoded by $\pi \in \{\mathsf{b},\mathsf{w}\}^*$. The expected capital of $A$ over $X$ after $n$ bets may be defined as

$$\mathsf{Ex}_A^X(n) = \sum_{\pi \in R(n)} P_A(\pi, \sigma(\pi)) \cdot C_A(\pi, \sigma(\pi)).$$

We can then define a more general analogue of non-monotonic Ex-randomness.

**Definition 5.4.9** (Non-monotonic GEx-randomness)**.** Let $X \in \{0,1\}^\omega$.

(a) A non-monotonic probabilistic strategy $A$ is a *generalised* Ex-*strategy* for $X$ if and only if

$$\lim_{n\to\infty} \mathsf{Ex}_A^X(n) = \infty;$$

(b) $X$ is said to be *non-monotonically* GEx-*random* if and only if no generalised non-monotonic probabilistic strategy is an Ex-strategy for $X$.

Now, is GEx-randomness a meaningful randomness notion? Does the class of sequences which are random in the sense of Definition 5.4.9 form a measure-one subset of $\{0,1\}^\omega$? How does GEx-randomness compare with non-monotonic Ex-randomness and (monotonic) Ex-randomness? These questions remain to be answered.

# Chapter 6

# Conclusion

Writing a thesis is a bit like playing the hydra game [Kirby and Paris, 1982]: for any question that one feels to have (at least partially) managed to answer, there are myriad new questions that suddenly pop up. If this analogy is indeed appropriate, though, we should be optimistic, for we are guaranteed to be able to address all these new questions in a finite amount of time! Embracing this hopeful attitude, we conclude this thesis by briefly summarising our results and by discussing some issues that we would like to further explore in the future, arranged in accordance with the chapter that they would be the natural continuation of.

## 6.1 Summary

The first part of this thesis centred around the following methodological question: what counts as a 'good' formalisation of our intuitive notion of randomness? First, we considered von Mises' pioneering work on randomness [1919], which constitutes the first attempt at providing a rigorous definition of randomness for infinite binary sequences. We discussed several arguments aimed at establishing the inadequacy of von Mises' theory, and we concluded, in agreement with van Lambalgen [1987a], that the demise of von Mises' approach is best understood in the context of a more general rejection of strict frequentism. We also noted that the tenability and usefulness of the theory of collectives crucially rests on the correctness of von Mises' objectivist interpretation of probability, for adopting a subjectivist interpretation renders his definition of randomness superfluous.

Having introduced the alternative (and, by now, orthodox) *algorithmic* approach to randomness, we then addressed the question of whether any of the definitions of randomness offered within this paradigm can be said to be more legitimate than the others. In particular, we focused on the notion of Martin-Löf randomness—arguably,

the most popular concept of algorithmic randomness in the literature—and we considered two main objections that have been levelled against it. First, we discussed a recent criticism due to Osherson and Weinstein [2008], which relies on a learning-theoretic argument. Then, we examined a well-known critique of Martin-Löf randomness due to Schnorr [1971a]. We pointed out the inconclusiveness of these criticisms and advocated a pluralistic approach to algorithmic randomness. While appraising Osherson and Weinstein's critique, we also allowed ourselves a brief learning-theoretic digression and proved a characterisation result for Kurtz randomness in learning-theoretic terms.

In the second part of this thesis, we considered some of the technical implications of taking Schnorr's critique of Martin-Löf randomness seriously, by investigating a probabilistic framework for algorithmic randomness introduced by Buss and Minnes [2013]. First, we reviewed Buss and Minnes' paper, where the authors countenance Schnorr's critique by offering a characterisation of Martin-Löf randomness in terms of *computable* probabilistic martingales. Then, we addressed a question that Buss and Minnes ask at the end of their paper: are there any natural conditions on the class of probabilistic martingales that can be used to characterise other common algorithmic randomness notions? We answered Buss and Minnes' question in the affirmative both in the monotonic and the non-monotonic setting by providing probabilistic characterisations of Martin-Löf randomness, Schnorr randomness, Kurtz randomness and Kolmogorov-Loveland randomness in terms of computable probabilistic martingales.

## 6.2 Future research

### Chapter 2

All definitions of randomness found in the algorithmic randomness literature are explicit or operational: a sequence is categorised as being random if and only if it satisfies a series of well-defined properties. However, other approaches are possible, as well. Van Lambalgen [1990], for instance, in trying to vindicate von Mises' frequentist theory, advances an axiomatic approach to randomness. He expands the language of set theory by adding a new primitive *independence* relation $R(X, \overline{Y})$, which expresses that $X \in \{0,1\}^\omega$ is uniformly random relative to some known data $\overline{Y}$—where $\overline{Y}$ is a finite tuple of sets. He then provides axioms for this relation and investigates various models of the proposed axioms. Interestingly, his analysis shows that randomness à la von Mises is at odds with both the axiom of choice and the extensionality axiom.

In [2014], Simpson provides an alternative realisation of van Lambalgen's programme, where randomness is defined via a series of axioms aimed at capturing the notion of *information independence*. Just like van Lambalgen, Simpson adds a new primitive independence relation to the language of set theory—more precisely, to $\mathsf{ZF}$ together with the axiom of dependent choice. The axioms for Simpson's independence relation (and the randomness axioms which hinge on this independence relation),

however, are based on ideas from *independence logic* (see, for instance, [Grädel and Väänänen, 2013]). Simpson is then able to show that his framework can be used to characterise randomness for general probability spaces.

Finally, in a recent talk [2015], Rute has also put forward a series of tentative axioms for algorithmic randomness. It is then natural to ask how these different axiomatisations relate to each other, and what their main differences are. Pursuing these questions could also provide a deeper understanding of the differences between von Mises' theory of collectives and algorithmic randomness.

### Chapter 3

The rejection of von Mises' foundational project in favour of the algorithmic paradigm inspired by Ville's measure-theoretic approach raises the fascinating question of whether there are any meaningful connections between algorithmic randomness and the foundations/interpretations of probability.

In their book "Probability and Finance. It's Only a Game!" [2001], Shafer and Vovk propose a new foundation for probability theory based on 'game theory' instead of measure theory—where the games in question involve the martingale functions that we discussed in Section 3.3.

Shafer and Vovk's work relies on a betting interpretation of probability which generalises *Cournot's Principle* [1843] (namely, the principle according to which an event of small or zero probability singled out in advance will not happen), which the authors take to be the only bridge between probability theory and the empirical world[63] [Shafer, 2015]. In view of Theorem 2.2.6 (proved by Ville), Shafer and Vovk reinterpret Cournot's Principle as saying that an event of small or zero probability is one for which a betting strategy whose capital never becomes negative will not multiply the gambler's capital by a large or infinite factor. They call this *Ville's Principle* and use it to advance an interpretation of probability based on forecasting.

In a series of papers, the authors (either individually or jointly) prove game-theoretic versions of classical measure-theoretic results: for instance, the strong law of large numbers, Lévy's zero-one law, and the law of calibration [Shafer et al., 2010]. Interestingly, they also provide game-theoretic variants of several results on 'merging of opinions' obtained within measure-theoretic probability (see, for instance, [Blackwell and Dubins, 1962]) and algorithmic randomness (see [Dawid, 1985]).

It would be interesting to further investigate Shafer and Vovk's betting interpretation of probability from both a philosophical and a technical point of view. For instance, to what extent is Shafer and Vovk's framework related to the subjective interpretation of probability of, say, de Finetti, which also hinges on betting intuitions? Can it itself be interpreted from a subjectivist viewpoint?

[63] Vovk also argues that Kolmogorov's finite frequentism is a combination of Cournot's Principle and a modified version of von Mises' theory of collectives [2001].

### Chapter 4

As already mentioned at the end of Chapter 4, we would like to further probe the connections between algorithmic randomness and computational learning theory.

The two notions of identification proposed by Osherson and Weinstein [2008] allow infinitely many (Definition 4.2.1) and finitely many (Definition 4.2.3) mind changes, respectively. So, a natural question is whether there are any reasonable identification criteria which

(i) are intermediate between strong sequence identification and sequence identification, in the sense that the corresponding success set is defined by a clause that is weaker than requiring cofinitely many yes's, but stronger than only requiring infinitely many yes's;

(ii) could be used to characterise learning-theoretically any of the randomness notions in between weak 2-randomness and Kurtz randomness.

To this end, one possible idea would be to use the notion of *asymptotic density*—where, given a set $D \subseteq \mathbb{N}$, its asymptotic density is the quantity

$$\rho(D) = \lim_{n\to\infty} \frac{|D \cap \{0, ..., n-1\}|}{n}.$$

Then, we could perhaps define the success set of a learning function $\ell$ as follows:

$$\mathcal{U}_\ell = \Big\{ Y \in \{0,1\}^\omega : \{n \in \mathbb{N} : \ell(Y \upharpoonright n) = \mathsf{yes}\} \text{ has positive asymptotic density} \Big\},$$

while still requiring that $\lambda(\mathcal{U}_\ell) = 0$. Now, if a sequence is such that no computable learning function identifies it in the sense of the above density-based criterion, is this sequence also random, as per some standard algorithmic randomness notion? If not, what is the relationship between this new randomness concept and the other randomness notions living within the algorithmic randomness hierarchy?

Another interesting question is what would happen if we modified the notion of 'allowable mistakes' in the definition of the success set of a learning function. For example, we could require the success set not only to have Lebesgue measure zero, but also to contain only sequences which lie within a small distance of the target sequence. For this purpose, we could perhaps employ the *asymptotic Hamming distance*, or some other natural measure of distance for infinite binary sequences. If a restriction along these lines were added to Definition 4.2.1 and Definition 4.2.3, would the concepts of sequence identification and strong sequence identification change or remain the same?

### Chapter 5

The first questions that we would like to answer are (i) whether KP1-randomness coincides with any standard randomness notion, and (ii) whether Schnorr randomness implies weak KP1-randomness (if this were indeed the case, then we would have that Schnorr randomness is equivalent to both weak KEx-randomness and weak KP1-randomness).

Now, Definition 5.3.5 from Section 5.3 could be extended by adding the intermediate notion of *locally weak* KEx-*randomness*:

> Let $X \in \{0,1\}^\omega$. Then, $X$ is said to be *locally weakly* KEx-*random* if and only if there is no probabilistic strategy $A$ which always eventually bets on $X$ with probability one such that $A$ is a KEx-strategy for $X$.

Then, is this new notion strictly weaker than Martin-Löf randomness (which coincides with KEx-randomness)? Is it strictly stronger than Schnorr randomness (which coincides with weak KEx-randomness)? Does it correspond to any well-studied randomness notion in between Martin-Löf randomness and Schnorr randomness?

We concluded Section 5.3 by defining the concepts of (weak) WP1-randomness and (weak) WEx-randomness, and by showing that Kurtz randomness is equivalent to both weak WEx-randomness and weak WP1-randomness. We did not further investigate the notion of WEx-randomness and WP1-randomness. Do they also correspond to well-known randomness concepts? Are they strictly stronger than their weak counterparts?

As already mentioned at the end of Section 5.4, it would also be interesting to determine whether there are meaningful analogues of the notions of non-monotonic P1-randomness and non-monotonic Ex-randomness once the position selection function $n_A$ is allowed to make its decisions not only on the basis of the binary string observed so far, but also according to the previously made bet/wait moves.

Finally, we would like to investigate notions of probabilistic randomness in the context of resource bounded probabilistic martingales. Schnorr's critique of Martin-Löf randomness has in fact prompted a number of researchers (including Schnorr himself) to study notions of resource bounded (pseudo)randomness in an attempt to further a more evidence-based approach to algorithmic randomness [Wang, 2000]. For instance, Schnorr [1971a] and Ko [1986] introduced resource bounded versions of Martin-Löf, computable and Schnorr randomness, while Lutz [1990; 1992] developed a resource bounded measure theory based on the notion of resource bounded martingales. In the future, we would like to generalise the notions of P1-randomness, Ex-randomness, KP1-randomness and KEx-randomness to this resource bounded setting. A natural question, then, would be whether Buss and Minnes' and our results extend to this framework, as well. For instance, does weak polynomial-time P1-randomness still coincide with polynomial-time randomness? Is weak polynomial-time KEx-randomness again equivalent to polynomial-time Schnorr randomness? What about notions of probabilistic randomness based on primitive

recursive betting strategies [Cenzer and Remmel, 2013; Buss et al., 2014], rather than computable ones? Lastly, it would be interesting to check to what extent Buss and Minnes' framework differs from Regan and Sivakumar's definition of probabilistic martingales based on randomised approximation schemes [1998], which is explicitly inspired by Lutz' paradigm [1990; 1992].

# Bibliography


K. Ambos-Spies, E. Mayordomo, Y. Wang, and X. Zheng. Resource-bounded balanced genericity, stochasticity and weak randomness. In C. Puech and R. Reischuk (eds.), *STACS '96: Proceedings of the 13th Annual Symposium on Theoretical Aspects of Computer Science,* Grenoble, Feb. 22-24, 1996, 1046: 63–74, 1996. (page 46)

L. Bienvenu and W. Merkle. Reconciling data compression and Kolmogorov complexity. In L. Arge et al. (eds.), *ICALP 2007*, Lecture Notes in Computer Science 4596, Springer-Verlag. (page 55)

L. Bienvenu, G. Shafer, and A. Shen. On the history of martingales in the study of randomness. *Journal Électronique d'Histoire des Probabilités et de la Statistique*, 5(1) (available at http://www.jehps.net), 2009. (page 22)

L. Bienvenu, A. R. Day, and R. Hölzl. From bi-immunity to absolute undecidability. *Journal of Symbolic Logic*, 78(4): 1218–1228., 2013. (page 58)

D. Blackwell and L. Dubins. Merging of opinions with increasing information. *The Annals of Mathematical Statistics*, 33: 882–886, 1962. (page 88)

L. E. J. Brouwer. Begründung der Mengenlehre unabhängig vom logischen Satz vom ausgeschlossenen Dritten I: Allgemeine Mengenlehre. *Ned. Acad. Wetensch. Verh. Tweede Afd. Nat.*, 12(5): 1–33, 1918. (page 15)

S. R. Buss and M. Minnes. Probabilistic Algorithmic Randomness. *Journal of Symbolic Logic*, 78(2): 578–601, 2013. (page iii, 6, 68, 72, 73, 74, 76, 77, 78, 83, 84, 87)

S. R. Buss, D. Cenzer, and J. B. Remmel. Sub-computable Bounded Pseudorandomness. Forthcoming: 1–30, 2014. (page 91)

D. Cenzer and J. B. Remmel. Sub-computable bounded pseudo-randomness. In *Proc. Logical Foundations of Computer Science*, Lecture Notes in Computer Science 7734, Springer-Verlag: 104–118, 2013. (page 91)

G. J. Chaitin. On the length of programs for computing finite binary sequences: statistical considerations. *Journal of the Association for Computing Machinery*,

16: 145–159, 1969. (page 31, 32)

G. J. Chaitin. A theory of program size formally identical to information theory. *Journal of the Association for Computing Machinery*, 22: 329–340, 1975. (page 39)

G. J. Chaitin. *Algorithmic Information Theory*. Cambridge University Press, Cambridge, 1987. (page 54)

T. Childers. *Philosophy & Probability*. Oxford University Press, Oxford, 2013. (page 17)

C. Chong, A. Nies, and L. Yu. The theory of higher randomness. *Israel J. Math.*, 166(1): 39–60, 2008. (page 53)

A. Church. On the Concept of a Random Sequence. *Bulletin of the American Mathematical Society*, 46: 130–135, 1940. (page 19, 20)

J. Conway. The Weird and Wonderful Chemistry of Audioactive Decay. *Eureka*, 46: 5–16, 1986. (page 32)

A. A. Cournot. *Exposition de la théorie des chances et des probabilités*. Hachette, Paris, 1843. (page 88)

H. Cramér. *Mathematical Methods of Statistics*. Princeton University Press, Princeton, New Jersey, 1957. (page 13)

V. Crupi. Confirmation. *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), Edward N. Zalta (ed.) (available at http://plato.stanford.edu/archives/spr2014/entries/confirmation/), 2014. (page 12)

A. Dasgupta. Mathematical Foundations of Randomness. In P. Bandyopadhyay and M. Forster (eds.), *Philosophy of Statistics,* Handbook of the Philosophy of Science, Vol. 7, Amsterdam, Elsevier (available at http://dasgupab.faculty.udmercy.edu/Dasgupta-JSfinal.pdf), 2010. (page 54)

M. David. The Correspondence Theory of Truth. *The Stanford Encyclopedia of Philosophy* (Summer 2015 Edition), Edward N. Zalta (ed.) (available at http://plato.stanford.edu/archives/sum2015/entries/truth-correspondence/), 2015. (page 27)

A. P. Dawid. Calibration-based empirical probability (with discussion). *Annals of Statistics*, 13: 1251–1285, 1985. (page 88)

J.-P. Delahaye. Randomness, Unpredictability and Absence of Order: The Identification by the Theory of Recursivity of the Mathematical Notion of Random Sequence. In J. Dubucs (ed.), *Philosophy of Probability*, Kluwer Academic Publishers, Dordrecht: 145–167, 1993. (page 6, 54)

J.-P. Delahaye. The Martin-Löf-Chaitin Thesis: The Identification by Recursion Theory of the Mathematical Notion of Random Sequence. In H. Zenil (ed.), *Randomness Through Computation: Some Answers, More Questions*: 121–140, 2011. (page 54)

O. Demut. The differentiability of constructive functions of weakly bounded variation on pseudo numbers. *Comment. Math. Univ. Carolinae*, 16(3): 583–599, 1975.

(page 31)

R. G. Downey and E. J. Griffiths. Schnorr randomness. *The Journal of Symbolic Logic*, 69: 533–55, 2004. (page 55)

R. G. Downey and D. R. Hirschfeldt. *Algorithmic Randomness and Complexity*. Springer, New York, 2010. (page 22, 43, 45, 57, 68)

R. G. Downey, E. J. Griffiths, and S. Reid. On Kurtz randomness. *Theoretical Computer Science*, 321: 249–270, 2004. (page 55)

B. de Finetti. La prévision: Ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, 7: 1–68, 1937. (page 11)

B. de Finetti. *Probability, Induction and Statistics*. Wiley, New York, 1972. (page 11)

M. Fisz. *Probability Theory and Mathematical Statistics*. Wiley, New York, 1963. (page 13)

P. Gács. Every Sequence Is Reducible to a Random One. *Information and Control*, 70: 186–192, 1986. (page 54)

H. Gaifman and M. Snir. Probabilities over rich languages, testing and randomness. *The Journal of Symbolic Logic*, 47: 495–548, 1982. (page 56)

D. Gillies. *Philosophical Theories of Probability*. Routledge, New York, 2000. (page 11, 12, 14)

M. Glanzberg. Truth. *The Stanford Encyclopedia of Philosophy* (Fall 2014 Edition), Edward N. Zalta (ed.) (available at http://plato.stanford.edu/archives/fall2014/entries/truth/), 2014. (page 27)

B. Gnedenko. *Theory of Probability*. Chelsea, Bronx, New York, 1968. (page 13)

K. Gödel. Remarks before the Princeton Bicentennial Conference on Problems in Mathematics. *Collected works*, 2: 144–153, 1946. (page 67)

E. Grädel and J. Väänänen. Dependence and independence. *Studia Logica*, 101(2): 399–410, 2013. (page 88)

I. Hacking. *Logic of Statistical Inference*. Cambridge University Press, Cambridge, 1965. (page 26, 28, 29)

A. Hájek. 'Mises Redux'-Redux: Fifteen arguments against finite frequentism. *Erkenntnis*, 45: 209–227, 1997. (page 30)

A. Hájek. The reference class problem is your problem too. *Synthese*, 156(3): 563–585, 2007. (page 28)

A. Hájek. Fifteen Arguments Against Hypothetical Frequentism. *Erkenntnis*, 70: 211–235, 2009. (page 29, 30)

D. Hilbert. Mathematical problems. *Bulletin of the American Mathematical Society*, 2(8): 437–479, 1902. (page 10)

J. M. Hitchcock and J. H. Lutz. Why computational complexity requires stricter martingales. *Theory of Computing Systems*, 39: 277–296, 2006. (page 22, 68)

S. Jain, D. Osherson, J. S. Royer, and A. Sharma. *Systems that Learn*. MIT Press, Chicago, 1999. (page 58)

E. Kamke. Über neuere Begründungen der Wahrscheinlichkeitsrechnung. *Jahres-*

*bericht Deutsche Mathematiker Vereinigung*, 42: 14–27, 1933. (page 18)

J. M. Keynes. *A Treatise on Probability*. Macmillan, 1921/1963. (page 13)

L. Kirby and J. Paris. Accessible Independence Results for Peano Arithmetic. *Bulletin of the London Mathematical Society*, 14: 285–293, 1982. (page 86)

K. Ko. On the notion of infinite pseudorandom sequences. *Theoretical Compuer Science*, 48: 9–33, 1986. (page 90)

A. N. Kolmogorov. On Tables of Random Numbers. *Sankhya: The Indian Journal of Statistics, Series A*, 25(4): 369–376, 1963. (page 28, 43)

A. N. Kolmogorov. Three approaches to the quantitative definition of information. *Problems of Information Transmission*, 1: 1–7, 1965. (page 28, 31, 32, 34, 35)

A. N. Kolmogorov and V. A. Uspenskii. Algorithms and Randomness. *Theory Probab. Appl.*, 32(3): 389–412, 1987. (page 54)

S. Kurtz. Randomness and genericity in the degrees of unsolvability. Ph.D. Thesis, University of Illinois at Urbana, 1981. (page 38, 40, 46, 56)

M. van Lambalgen. Random Sequences. Ph.D. Thesis, Universiteit van Amsterdam, 1987a. (page 4, 5, 15, 17, 18, 25, 86)

M. van Lambalgen. Von Mises' definition of random sequences reconsidered. *Journal of Symbolic Logic*, 52: 725–755, 1987b. (page 10)

M. van Lambalgen. The axiomatization of randomness. *The Journal of Symbolic Logic*, 55: 1143–1167, 1990. (page 66, 87)

M. van Lambalgen. Randomness and foundations of probability: von Mises' axiomatisation of random sequences. In T. S. Ferguson, L. S. Shapley, and J. B. MacQueen (eds.), *Statistics, probability and game theory: Papers in honor of David Blackwell* (available at http://projecteuclid.org/euclid.lnms/1215453582): 347–367, 1996. (page 25)

P.-S. Laplace. *A Philosophical Essay on Probabilities*. Dover, Translated from 6th French edition, 1819/1952. (page 4)

L. Levin and A. K. Zvonkin. The complexity of finite objects and the develpment of the concepts of information and randomness by means of the theory of algorithms. *Uspekhi Mat. Nauk*, 25: 85–127, 1970. (page 31)

L. A. Levin. On the notion of a random sequence. *Soviet Mathematics Doklady*, 14: 1413–1416, 1973. (page 8, 31, 54)

L. A. Levin. Uniform tests of randomness. *Soviet Mathematics Doklady*, 17(2): 337–340, 1976. (page 8)

L. A. Levin. Randomness conservation inequalities: Information and independence in mathematical theories. *Information and Control*, 61(1): 15–37, 1984. (page 8)

D. Lewis. A Subjectivist's Guide to Objective Chance. *Philosophical Papers, Volume II*, Oxford University Press, 1980. (page 29)

M. Li and P. M. B. Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications*. Third Edition, Springer, New York, 1993/1997/2008. (page 3, 4, 15, 16, 28, 31, 34, 35)

D. Loveland. A new interpretation of the von Mises' concept of random sequence. *Z. Math. Logik Grundlagen Math*, 12: 279–294, 1966. (page 43)

J. H. Lutz. Category and measure in complexity classes. *SIAM J. Comput.*, 19: 1100–1131, 1990. (page 90, 91)

J. H. Lutz. Almost everywhere high nonuniform complexity. *J. Comput. System Sci.*, 44: 220–258, 1992. (page 90, 91)

R. Mansuy. The Origins of the Word "Martingale". *Journal Électronique d'Histoire des Probabilités et de la Statistique*, 5(1) (available at http://www.jehps.net/juin2009/Mansuy.pdf), 2009. (page 22)

P. Martin-Löf. The Definition of a Random Sequence. *Information and Control*, 9: 602–619, 1966. (page iii, 8, 31, 36, 37)

P. Martin-Löf. On the notion of randomness. In *Intuitionism and Proof Theory*, (*Proc. Conf., Buffalo, N.Y., 1968)*: 73–78, 1970. (page 53)

W. Merkle, J. S. Miller, A. Nies, J. Reimann, and F. Stephan. Kolmogorov-Loveland randomness and stochasticity. *Annals of Pure and Applied Logic*, 138(1-3): 183–210, 2006. (page 45, 68)

J. S. Miller and L. Yu. On initial segment complexity and degrees of randomness. *Transaction of the American Mathematical Society*, 360: 3193–3210, 2008. (page 36)

R. von Mises. Grundlagen der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 5: 52–99, 1919. (page iii, 4, 10, 15, 86)

R. von Mises. *Probability, Statistics and Truth.* 2nd revised English edition, Allen and Unwin, 1928/1961. (page 11, 16, 26)

R. von Mises. On the Foundations of Probability and Statistics. *The Annals of Mathematical Statistics*, 12(2): 191–205, 1941. (page 14)

R. von Mises. *Mathematical Theory of Probability and Statistics.* Academic Press, New York, 1964. (page 13)

K. Miyabe and J. Rute. Van Lambalgen's theorem for uniformly relative Schnorr and computable randomness. *Proceedings of the 12th Asian Logic Conference*: 251–270, 2013. (page 66)

A. A. Muchnik, A. L. Semenov, and V. A. Uspensky. Mathematical metaphysics of randomness. *Theoretical Computer Science*, 207: 263–317, 1998. (page 41, 43, 45)

D. Osherson and S. Weinstein. Recognizing strong random reals. *The Review of Symbolic Logic*, 1(01): 56–63, 2008. (page iii, 6, 7, 54, 56, 57, 59, 60, 87, 89)

C. P. Porter. The Algorithmic Approach to Randomness. Talk given at the *Groupe de Travail Mathématiques et Philosophie Contemporaines,* Institut de Mathématiques de Toulouse, Université Paul Sabatier, Toulouse, France, March 20, 2014, (available at http://www.cpporter.com/wp-content/uploads/2013/08/PorterToulouse2014.pdf), 2014. (page 39)

C. P. Porter. On analogues of the Church-Turing thesis in algorithmic randomness. Submitted for publication: 1–25, 2015. (page 67)

F. P. Ramsey. Truth and probability. In R. B. Braithwaite (ed.), *Foundations of*

*Mathematics and other Essays*, London, Kegan, Paul, Trench, Trubner & Co.: 156–198, 1931. (page 11)

K. W. Regan and D. Sivakumar. Probabilistic Martingales and BPTIME Classes. *Proceedings of the 13th Annual IEEE Conference on Computational Complexity*: 186–200, 1998. (page 91)

J. Richard. Les Principes des Mathématiques et le Problème des Ensembles. *Revue Générale des Sciences Pures et Appliquées,* translated in van Heijenoort, J. (1964), *Source Book in Mathematical Logic 1879-1931,* Cambridge, MA: Harvard University Press, 1905. (page 20)

J. Rute. Topics in algorithmic randomness and computable analysis. PhD Thesis, Carnegie Mellon University, 2013. (page 65)

J. Rute. New directions in randomness. Talk given at *Computability, Complexity, and Randomness*, June 22-26, 2015. (page 65, 88)

L. J. Savage. *The Foundations of Statistics*. Wiley, 1954. (page 11)

C. P. Schnorr. Zufälligkeit und Wahrscheinlichkeit. *Lecture Notes in Mathematics*, 218, Springer-Verlag, Berlin-Heidelberg-New York, 1971a. (page iii, 21, 31, 38, 40, 41, 43, 45, 46, 87, 90)

C. P. Schnorr. A unified approach to the definition of a random sequence. *Mathematical Systems Theory*, 5: 246–258, 1971b. (page 6, 31, 38, 54, 56, 63)

G. Shafer. A betting interpretation for probabilities and Dempster-Shafer degrees of belief. Forthcoming, to appear in the *International Journal of Approximate Reasoning*: 1–18, 2015. (page 88)

G. Shafer and V. Vovk. *Probability and Finance: It's Only a Game!* Wiley, New York, 2001. (page 88)

G. Shafer, V. Vovk, and R. Chychyla. How to base probability theory on perfect-information games. *BEATCS*, 100: 115–148, 2010. (page 88)

C. Shannon. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3) (available at http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf): 379–423, 1948. (page 32)

A. Simpson. Independence set theory and randomness. Talk given at the *Logic Seminar*, Department of Pure Mathematics, University of Leeds, 28th January, 2014. (page 87)

R. J. Solomonoff. A preliminary report on a general theory of inductive inference. Technical Report ZTB-138, Zator Company, Cambridge, Mass., 1960. (page 31, 32)

R. J. Solomonoff. A formal theory of inductive inference, I and II. *Information and Control*, 7: 1–22 and 224–254, 1964. (page 31, 32)

R. Solovay. Draft of a paper (or a series of papers) on Chaitin's work. Unpublished notes: 1–215, 1975. (page 56, 57)

A. Taveneaux. The Randomness Zoo. In A. Nies (ed.), *Logic Blog 2012* (available at https://dl.dropboxusercontent.com/u/370127/Blog/Blog2012.pdf): 15–19, 2012.

(page 53)

A. M. Turing. On Computable Numbers, with an Application to the Entscheidungsproblem. *Proc. London Math. Soc.*, S2-42(1):230–265, 1936. (page 32)

S. Vermeeren. Notions and applications of algorithmic randomness. Ph.D. Thesis, The University of Leeds, School of Mathematics, 2013. (page 6, 7, 21, 31, 43, 48)

J. Ville. Étude critique de la notion de collectif. *Monographies des probabilités*, Gauthiers-Villars, Paris, 1939. (page 22, 24, 31, 74)

J. Ville. A Counterexample to Richard von Mises's Theory of Collectives. Translation and Introduction by Glenn Shafer (available at http://www.probabilityandfinance.com/misc/ville1939.pdf): 1–13, 1939/2005. (page 19)

P. M. B. Vitányi. Turing machine. *Scholarpedia*, 4(3): 6240, 2009. (page 32)

V. Vovk. Kolmogorov's complexity conception of probability. In V. F. Hendricks, S. A. Pedersen, and K. F. Jørgensen (eds.), *Probability Theory: Philosophy, Recent History and Relations to Science*, Kluwer Academic Publishers: 51–70, 2001. (page 88)

A. Wald. Die Wiederspruchsfreiheit des Kollektivbegriffes der Wahrscheinlichkeitsrechnung. *Ergebnisse eines Mathematischen Kolloquiums*, 8: 38–72, 1937. (page 19)

Y. Wang. Randomness and Complexity. Ph.D. Thesis, Univerity of Heidelberg, 1996. (page 40, 41, 48, 53)

Y. Wang. Resource bounded randomness and computational complexity. *Theoretical Computer Science*, 237(1-2): 33–55, 2000. (page 61, 90)

A. N. Whitehead and B. Russell. *Principia Mathematica*. Cambridge University Press, Cambridge, 1910/1912/1913. (page 35)