

Patterns and Probabilities:

A Study in Algorithmic Randomness and
Computable Learning

Francesca Zaffora Blando

Patterns and Probabilities:

A Study in Algorithmic Randomness and
Computable Learning

ILLC Dissertation Series DS-2020-17



INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

For further information about ILLC-publications, please contact

Institute for Logic, Language and Computation

Universiteit van Amsterdam

Science Park 107

1098 XG Amsterdam

phone: +31-20-525 6051

e-mail: illc@uva.nl

homepage: <http://www.illc.uva.nl/>

Advisor: Johan van Benthem

Readers: Ray Briggs
Persi Diaconis
Thomas Icard
Brian Skyrms

Defense: July 24, 2020

Department of Philosophy
Stanford University
Stanford, CA
USA

PATTERNS AND PROBABILITIES:
A STUDY IN ALGORITHMIC RANDOMNESS
AND COMPUTABLE LEARNING

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF PHILOSOPHY
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Francesca Zaffora Blando
Stanford Version: August 2020
ILLC Version: September 2021

Abstract

This dissertation bridges the theory of algorithmic randomness—a branch of computability theory—and the foundations of inductive learning. Algorithmic randomness provides a mathematical analysis of the notion of an individual object (such as a string of bytes representing a computer file or a sequence of experimental outcomes) displaying no effective patterns or regularities. Here, we investigate the role that algorithmic randomness plays in inductive learning when randomness is taken to be a property of sequences of observations, or data streams, and the learners are computationally limited. Our results constitute a first step towards a systematic classification and analysis of the learning scenarios where algorithmic randomness is beneficial for inductive learning and the scenarios where it is instead detrimental for the learning process.

We focus on three main themes. Firstly, we explore the connections between algorithmic randomness and Bayesian merging-of-opinions theorems. In particular, we show that algorithmic randomness leads to merging of opinions in the following sense. When two computable Bayesian agents perform the same experiment, agreeing on which data streams are algorithmically random suffices to guarantee that they will eventually reach a consensus with probability one even if, at the beginning of the learning process, their beliefs are otherwise dissimilar. Secondly, we consider the role of algorithmic randomness in Bayesian convergence-to-the-truth results. More precisely, we show that there is a robust correspondence between algorithmic randomness and the collection of truth-conducive data streams. When a computable Bayesian agent is faced with an effective inductive problem, the algorithmically random data streams are in fact exactly the ones that ensure that their beliefs will asymptotically align with the truth. Lastly, we investigate a learning-theoretic approach—in the spirit of formal learning theory—for modelling algorithmic randomness itself. Building on the local irregularity and unruliness that is the hallmark of algorithmic randomness, we show that the algorithmically random data streams are, systematically, unlearnable: i.e., they coincide with the data streams from which no computable qualitative learning method can extrapolate any patterns.

Acknowledgments

First, I would like to thank the members of my reading committee: Johan van Benthem, Ray Briggs, Persi Diaconis, Thomas Icard, and Brian Skyrms. Their guidance throughout my time at Stanford has shaped not only this dissertation, but my entire intellectual trajectory as a Ph.D. student.

Johan has been a wonderful advisor, mentor, role model, and, more recently, collaborator. I am grateful for his unfaltering support, for gently steering me in the right direction while always letting me follow my own path, and for sharing his far-reaching and inspiring vision of logic with me. Johan's remarkable talent for revealing deep and illuminating connections in unexpected places, his knack for identifying open questions that end up giving rise to entire novel research programmes, and his ability to pay attention to the fine details while never losing track of the big picture has not only been a constant source of inspiration, but it also sets a high standard that I hope to be able to approximate in my own work (at least in the limit).

Ray has also been a mentor to me ever since the beginning of my studies at Stanford. I am extremely grateful for their generous advice and help, and for the countless thought-provoking conversations: their philosophical sharpness and impressive ability to always ask the right questions have been truly inspiring. I also thoroughly enjoyed our weekly group meetings, which greatly contributed to fostering a sense of community and belonging among the graduate students in the department working on formal philosophy.

Persi has a gift for seamlessly bringing abstract mathematical concepts down to earth, and his views on the foundations of probability and statistics permeate this dissertation. I am thankful for his encouragement, for the delightful Erdős anecdotes, and for his numerous insightful comments and questions. He is a well of invaluable knowledge, and many of the projects I intend to embark on next are fuelled by his suggestions. I very much hope that this dissertation marks but the beginnings of a fruitful conversation.

Thomas' help during these past five years, in matters both academic and personal, has been tremendous. I cannot thank him enough for his generosity and unwavering support,

for always being willing to engage in discussion, and for being a continual source of inspiring ideas. His breadth of expertise and depth of reflection, and his ability to drive straight to the heart of a problem are but some of the many qualities that make him an incredible mentor. Thank you for being such a brilliant teacher and friend to me and Krzys!

Brian's influence, palpable throughout this dissertation, stretches back to my undergraduate studies: my interest in Bayesian epistemology is in large part due to him and his pioneering work. Brian's striking ability to bridge technical results from diverse fields and unearth their profound philosophical ramifications drew me to the study of probability and its foundations, and his ideas provided the scaffolding for much of my research and current philosophical views. I am incredibly grateful for his being so supportive of my pursuits and for suggesting so many fascinating avenues for future work. I also hope to someday learn to write philosophy as pithily as he does.

The results in Chapter 3 were obtained as part of a joint—and still ongoing—research project with Simon Huttegger and Sean Walsh. Working with Sean and Simon has been a fantastic experience: I have learnt enormously from them, not only about logic, probability, and philosophy, but also about doing research. They have been, and continue to be, extraordinary role models. I would like to thank them both for generously letting me include results from our joint project in my dissertation. I am truly honoured to have the privilege of working with them.

I would also like to thank Tobias Gerstenberg for graciously agreeing to chair my dissertation defense on an incredibly short notice. I am also thankful for his interesting questions and for bringing to my attention some potential connections with recent work in psychology and Bayesian models of cognition.

A big thank you to Jill Covington, as well, for her patience and all her help with navigating the administrative maze of academia.

Many other people have helped me greatly along the way, both at Stanford and elsewhere. I have particularly benefited from conversations with Carlos Areces, Jeff Barrett, Gordon Belot, the late Sol Feferman, Michał Godziszewski, Nadeem Hussain, Jim Joyce, Hanti Lin, Anna-Sara Malmgren, Louis Narens, Jan-Willem Romeijn, Teddy Seidenfeld, Tom Sterkenburg, and Marta Sznajder. I would also like to thank the entire Stanford Philosophy Department, as well as the Logic and Philosophy of Science Department at UC Irvine (where I spent a terrific year before joining the Stanford Ph.D. programme). Both departments are exceptionally vibrant intellectual communities, where learning, doing research, and teaching have been unfailingly rewarding and an occasion for growth.

Over the past five years, Stanford has truly become my home: the friendships I have

made here are of the kind in which “long years apart can make no breach a second cannot fill”, and leaving is bittersweet. I am particularly grateful to Robert Bassett, Monika Chao (and Zoe), J.T. Chipman, Michael Cohen, Phoebus Cotsapas, Lorenza D’Angelo, Nick DiBella, Huw Duffy, Jonathan Ettel, Dave Gottlieb, Nathan Hauthaler, Hannah Kim, Hy-oung Sung Kim, JJ Lang (and Raskol), Olga Lenczewska, Meica Magnani, Katy Meadows, Jerriey Mottley, Grace Paterson, Noe Pollack, Stephanie Roach, Patrick Ryan, Thomas Slabon, Shane Steinert-Threlkeld, Declan Thompson, John Turman, Yafeng Wang, Steve Woodworth, and Adam Zweber for being such wonderful people and for making my time at Stanford so special.

Reaching the finishing line would not have been possible without a caring family. There are no words big enough to describe how grateful I am to my parents, Elisabetta and Beppe: their unconditional love and support has kept me going all these years. Their incredible strength and integrity continue to set a luminous example, and their curiosity and inquisitive minds are what set me on a path towards philosophy. *Grazie per essere casa e radici.*

I would also like to thank my parents-in-law, Magda and Piotr, my sister-in-law Natalia and her soon-to-be husband Janek, as well as Andrzej and Ewa Witkowski (our California family). Thank you for reminding me and Krzyś what is truly important!

Thank you to Pringle for keeping my chair warm.

And, last but foremost, thank you to my husband Krzyś, who is necessary and makes it all possible. Going down stairs with you all these years has been the most wonderful of adventures, and you never cease to inspire me. I think I have now entirely forgotten where you begin and I end. As Pan Piękny once said, we are very similar (even in height), and therefore perfect for each other. I look forward to new beginnings together, which are always returns to where it all began.

Stanford, California

July 2020

Preface to the ILLC Version

I would like to thank Johan van Benthem for inviting me to publish my dissertation in the ILLC Dissertation Series. I am also grateful to Nick Bezhanishvili and Marco Vervoort for their patience and assistance with the publication of this work. The original version of this dissertation was submitted to Stanford University in August 2020. For the ILLC version, I have made some corrections and minor modifications, and I have included a few references to more recent work and results. Any remaining mistakes or inaccuracies are of course my past self's sole responsibility.

Pittsburgh, Pennsylvania

September 2021

Contents

Abstract	iii
Acknowledgments	iv
Preface to the ILLC Version	vii
Introduction	1
1 Algorithmic randomness	12
1.1 Notation and basic definitions	14
1.2 The measure-theoretic typicality paradigm	18
1.2.1 Sequential tests	18
1.2.2 Integral tests	26
1.3 The unpredictability paradigm	29
2 Algorithmic randomness and merging of opinions	35
2.1 Merging of opinions	39
2.1.1 Classical notions of compatibility and incompatibility	41
2.1.2 Merging and polarisation in the classical framework	45
2.2 Algorithmic randomness	48
2.2.1 Algorithmic randomness and merging	51
2.2.2 Algorithmic randomness and polarisation	64
2.3 Conclusion	65
Appendix 2.A Converse implications	68
3 Algorithmic randomness and convergence to the truth	72
3.1 Convergence to the truth	75

3.1.1	The classical version of Lévy’s Upward Theorem	76
3.1.2	The epistemic significance of Lévy’s Upward Theorem	78
3.2	Effectivising Lévy’s Upward Theorem	81
3.2.1	L^1 -computable functions	83
3.2.2	Integral tests for randomness	87
3.2.3	Weakly L^1 -computable functions	107
3.3	Discussion	117
3.3.1	Why is algorithmic randomness truth-conducive?	117
3.3.2	Probabilities over rich languages	119
3.3.3	In the long run we are all dead	121
3.3.4	Epistemic immodesty	123
3.4	Conclusion	126
4	Algorithmic randomness and unlearnability	131
4.1	Algorithmic randomness and detectability	134
4.1.1	Strong detectability	136
4.1.2	Weak detectability	138
4.2	Martin-Löf randomness and Schnorr randomness	140
4.2.1	A connection with integral tests	140
4.2.2	Uniform weak detectability and Martin-Löf randomness	143
4.2.3	Computably uniform weak detectability and Schnorr randomness	147
4.3	Conclusion	153
	Bibliography	156

List of Figures

1.1	The binary Cantor tree	17
1.2	The arithmetical hierarchy of subsets of Cantor space	20
1.3	Part of the algorithmic randomness hierarchy	25
2.1	Dependencies between notions of compatibility between measures	45
2.2	Dependencies between notions of (in)compatibility and of merging/polarisation	49
2.3	Dependencies between the notions of compatibility induced by algorithmic randomness and notions of merging	64
3.1	A left-c.e. Martin-Löf random sequence	97
4.1	Diagonal approximation of a lower semi-computable function	142

Introduction

We see how difficult it is to *not* behave inductively and to *not* believe in inductive success in the long run. We see how difficult it is to act as if one were *not* uncertain about chance. The interesting questions are not whether we do or should behave inductively, but about *how* we do or should behave inductively with respect to *what* conception of chance.

Skyrms, *Pragmatics and Empiricism*

Hume’s problem of induction and formal models of inductive learning. Induction is the hallmark of scientific thinking and is at the root of nearly all of our common-sense reasoning. Most of our knowledge of the world is inductive: we observe certain regularities in nature and exploit them to make successful predictions about what will be observed next, or to extrapolate general laws about the world. The miracle of induction lies in the fact that inductive inferences, which are by their very nature *ampliative*,¹ allow us to successfully extend our knowledge from the realm of the observed to that of the unobserved—and even to the unobservable.

Yet, in spite of being the “the glory of Science” and everyday life, inductive reasoning is still widely regarded as “the scandal of Philosophy” [Broad, 1952, p. 143]. This is because of what has come to be known as *Hume’s problem of induction*, which purports to show that it is simply not possible to provide a rational justification for our ubiquitous reliance

¹That is, their conclusions go beyond their premises: they are not truth-preserving (i.e., they can take us from true premises to a false conclusion) and typically involve generalisations.

on inductive reasoning. Hume's famous argument proceeds as follows (cf. [Hume, 1978, Book 1, Part iii, Section 6]). There are two possible routes for supplying a rational justification of a mode of reasoning: a deductively valid argument or an inductive argument. Since inductive inferences are not truth-preserving, induction cannot be justified deductively: that is, via a deductively valid (and, thus, truth-preserving) argument relying on the premise that induction has worked well in the past. Therefore, our only hope is that induction may itself be justifiable inductively. An inductive justification of induction would amount to arguing that, since our inductive inferences have worked well in the past, we can reasonably expect that they will continue to do so in the future. However, as persuasively argued by Hume, any justification of this type is irremediably circular: by taking the past successes of induction as a reason to infer that induction will be successful again in the future, the argument presupposes exactly what it was meant to establish—namely, that inductive inferences can be rationally relied upon. It then appears that we have reached an insurmountable impasse: neither a deductive nor an inductive justification of induction is available, and so there can be no principled justification for our reliance on inductive reasoning, either in the sciences or in everyday life.

While there has been a plethora of attempts at circumventing Hume's negative conclusion, none of the proposed solutions to the problem of induction is fully general and assumption-free. Many have thus decided to bite the bullet and accept that Hume's argument is simply correct. Savage, for instance, asserts the following:

Hume's arguments, and modern variants of them such as Goodman's discussion of 'bleen' and 'grue', appeal to me as correct and realistic. That all my beliefs are but my personal opinions, no matter how well some of them may coincide with opinions of others, seems to me not a paradox but a truism. [Savage, 1967, p. 602]

Dreary as it might seem at first, embracing the conclusion of Hume's argument need not amount to a debacle and to the abandonment of the project of building a theory of rational inductive inference. If a fully general justification of induction cannot be obtained, perhaps it is nonetheless possible to offer a *local*, or *conditional*, justification:

[J]ust as the force of Hume’s argument is usually underestimated, so the devastation it is supposed to bring in its wake is usually exaggerated, and my supplementary thesis is that there is none the less a positive solution to Hume’s problem. Indeed, I will argue the apparently paradoxical claim that there are nevertheless demonstrably sound inductive inferences! The resolution of the paradox is that inductive inference arises as a necessary feature of consistent reasoning, given the sorts of initial plausibility assumptions scientists habitually make. [...] I believe that Hume genuinely has solved the problem of induction. He solved it by showing, in general terms, that a sound inductive inference must possess, in addition to whatever observational or experimental data is specified, at least one independent assumption (an inductive assumption) that in effect weights some of the possibilities consistent with that evidence more than others. I take this to be a great logical discovery, comparable to that of deductive inference itself, and with consequences of as much practical as theoretical importance. [Howson, 2000, p. 2]

The basic idea behind this approach is that, even though Hume’s problem of induction leaves little to no room for the possibility of establishing the rationality of inductive reasoning in absolute terms (by either deductive or inductive means), we may nevertheless be able to evaluate the reasonableness of an inductive inference relative to specific collections of inductive assumptions. By singling out a family of rational learning methods for each particular learning situation and class of inductive assumptions, we may still be able to provide local, rather than global, justifications of induction: that is, justifications which, by virtue of not being assumption-free, do not work in every case—and therefore fail to be global—but which nonetheless do succeed, provided that the inductive assumptions on which they rely hold.

This attitude towards the problem of induction fuels much work in statistics, machine learning, and formal epistemology (as remarked by Skyrms, perhaps “[t]he scandal of philosophy is not that the logic of induction does not exist, but rather that philosophy has paid so little attention to it” [Skyrms, 2012, p. 250]). Formal theories and models of inductive learning—from probability and statistics to logic and computer science—are particularly

well-suited for advancing this programme, as they allow to unambiguously spell out what the underlying inductive assumptions are, and to elucidate how the power and reliability of an inductive method vary as a function of said assumptions. In each case, the adopted inductive assumptions may of course be called into question; however, a leading objective of this approach is systematisation: that is, the goal is to gain a clear and comprehensive understanding of the *conditional* rationality of inductive reasoning and learning.

The present work follows in this tradition: the results in this dissertation target the problem of induction from the conditional perspective outlined above. In particular, the key aim of this work is to clarify the extent to which inductive learning is affected (positively or negatively) by the presence of randomness in the data—where randomness is understood as irregularity and lack of discernible patterns. To address this question, we focus on two central paradigms for modelling inductive learning: the theory of Bayesian learning and formal (or computational) learning theory. Both paradigms tackle the problem of induction through the prism of specific classes of inductive assumptions. In return, they provide precise, yet very general frameworks within which one can represent different inductive problems and gauge the performance of various inductive methods. This, in turn, is exactly what allows to rigorously study the role of randomness in inductive learning. To model the presence of randomness in the data, on the other hand, we draw on the theory of algorithmic randomness, which, as we shall see, relies on the tools of mathematical logic and measure theory to elucidate the concepts of patternlessness and irregularity.

Randomness and learning. Randomness is generally taken to be a property of processes that are governed by chance (or, at the very least, that are taken to be best modelled in probabilistic terms). Phenomena that typically get classified as random processes in this sense include games of chance—such as rolling a die, tossing a coin, or spinning a roulette wheel—as well as various processes studied in the natural sciences—such as the growth of a bacterial population, the dynamics of gas molecules, or radioactive decay—and the social sciences—such as the fluctuations of the stock market.

However, this process conception of randomness is not the only one available. A second, complementary way of thinking about randomness is as a property of outcomes, rather than processes. This second type of randomness, usually called *product randomness* (as opposed

to *process randomness*), plays a significant role in a variety of fields, including information theory, cryptography, the foundations of probability and statistics, and computability theory. Intuitively, product randomness amounts to lack of structure or patternlessness: a sequence of events, observations, experimental outcomes, or symbols from some alphabet is product random if it does not display any discernible patterns or regularities.

To build some intuition about the notion of product randomness, consider the two binary strings below, each of which consists of fifty digits:

01101110110011000100110101011111001011110011100011
10

The first string is ostensibly more random-looking than the second. This is because, by contrast to the first string, the second string displays an obvious pattern that is very easy to describe (the pair “10” is repeated twenty-five times). Moreover, this pattern appears to be easily exploitable to make successful predictions: observing the first, e.g., twenty digits of the second string seems to provide useful information for predicting what digits will be observed next. On the other hand, it is not at all obvious that the same can be said about the first string.

Random processes can be successfully modelled using the tools of probability theory and statistics. It is then natural to wonder whether the notion of product randomness just delineated is amenable to a rigorous conceptual and mathematical analysis, as well. Can the above intuitions be made precise and combined into a coherent account of product randomness as lack of patterns?

The most well-developed mathematical treatment of the notion of product randomness is the theory of algorithmic randomness, a branch of computability theory which equates randomness with the absence of any regularities discernible using algorithmic means. More precisely, the theory of algorithmic randomness specifies under what conditions an *individual* mathematical object (such as a computer file, encoded as a string of bytes, an infinite binary sequence, or a graph) can be said not to possess any effectively discernible patterns.

Algorithmic randomness has two fundamental features. First, due to its reliance on computability theory, it is not a theory of “absolute randomness”. Rather, it yields an infinite hierarchy of randomness notions, each of which corresponds to a specific “level

of randomness”—where each level is determined by the effectivity constraints used to demarcate the random objects from the non-random ones. The more restricted the kind of effectivity involved, the logically weaker the resulting algorithmic randomness notion, in the sense that strictly more objects are classified as random relative to said notion.

Second, algorithmic randomness notions are always defined with respect to some underlying probability measure, so that an object that is classified as random relative to a certain measure may be classified as non-random relative to a different measure (for instance, sequences that are algorithmically random relative to the uniform measure fail to be so relative to any computable biased Bernoulli measure). As we will see, both of these features will play an important role in the results to come.

The goal of this dissertation is to bridge the theory of algorithmic randomness and the foundations of inductive learning. The driving question behind this work is the following: what are the effects of product randomness on inductive learning—where randomness is taken to be a property of sequences of observations (data streams)? More precisely, to strike a better balance between generality and realism, we tackle this question in the context of computationally limited learners. In other words, we focus on how observing a random data stream affects the learning performance of a computable (and, thus, less-than-ideal) agent, and investigate to what extent the effects of randomness in the data depend on the particular learning task to be solved.

We appeal to the theory of algorithmic randomness because it is singularly well-suited for investigating the relationship between product randomness and learning for computable agents. Not only does algorithmic randomness provide a coherent formal framework within which the above questions can be made precise, but, due to its reliance on computability theory, it is by its very nature congenial to the study of learners whose resources do not exceed those of a Turing machine.

The results in this dissertation constitute a first step towards a systematic classification of the learning scenarios where algorithmic randomness is beneficial for inductive learning and the learning scenarios where it is instead detrimental for the learning process. Each chapter explores these issues from a different angle. Chapter 2 and Chapter 3 focus on the connections between algorithmic randomness and probabilistic learning—specifically, on the learning performance of computable Bayesian agents. In both chapters, we show

that, in the different learning scenarios under consideration, algorithmic randomness is not only beneficial for learning: it is conducive to it. Chapter 4, on the other hand, is centred around the issue of learnability by computable learning functions—traditionally investigated within the field of formal learning theory—and on how learnability relates to randomness. In particular, we show that, in that context, algorithmic randomness acts against learning: in fact, it corresponds to a specific type of unlearnability. Crucially, this result does not contradict the findings from Chapter 2 and Chapter 3: we will see that the opposite effects that algorithmic randomness turns out to have on inductive learning in these two settings can be traced back to the differences between the types of inductive problems considered in each case.

What this dissertation is not about. Before discussing in some detail the content and structure of this dissertation, we briefly consider some topics that, although related to the subject matter of the present work, fall outside its scope.

As mentioned above, Chapter 2 and Chapter 3 bridge algorithmic randomness with Bayesian learning and Bayesian probability theory. However, there is another interpretation of probability that is more often discussed in connection with algorithmic randomness: the frequency interpretation. This is because the theory of algorithmic randomness has a famous forerunner: von Mises' theory of *collectives* [1919; 1981], which is steeped in the frequency interpretation. According to von Mises, an account of product randomness is necessary to define the concept of probability: he in fact believed that probabilities are limiting relative frequencies within collectives—i.e., within infinite random sequences of events or experimental outcomes. In turn, in von Mises' theory, a sequence is random if it is not possible to devise a strategy for selecting from it an infinite subsequence that allows odds for gambling different from those allowed by the initial sequence—that is, such that the limiting relative frequencies of events in the selected subsequence differ from the ones displayed by the initial sequence. Von Mises' original definition of an admissible strategy for selecting subsequences was left informal. Church [1940] later made it precise in computability-theoretic terms by means of identifying admissibility with computability. This computability-theoretic turn is what paved the way for the development of the theory of algorithmic randomness.

Nowadays, von Mises’ definition of product randomness is by and large considered too weak: as shown by Ville [1939], random sequences in the sense of von Mises can display regularities that render them highly predictable. Moreover, unlike von Mises’ theory of collectives, the theory of algorithmic randomness is not based on the tenet that an account of randomness is required to define the concept of probability. In fact, as mentioned earlier, algorithmic randomness is heavily measure-theoretic: the notion of a probability measure is necessary to be able to define randomness. So, while algorithmic randomness does not depend on any specific interpretation of probability, it is at odds with von Mises’ brand of frequentism, according to which probabilities are to be defined in terms of limiting relative frequencies. We will not further discuss von Mises’ approach and how it differs from the modern theory of algorithmic randomness in this dissertation. The reader interested in an in-depth analysis of von Mises’ ideas, as well as in a defense of his definition of randomness, is invited to consult [van Lambalgen, 1987a,b, 1996].

The first author to bridge Bayesian learning and algorithmic randomness—and, in particular, the theory of Kolmogorov complexity, which equates randomness with incompressibility (see [Li and Vitányi, 2019])—was Solomonoff [1964], with his theory of inductive inference. Solomonoff’s induction (as his account is often referred to in the literature) is a theory of prediction: given an infinite sequence of outcomes or observations generated probabilistically, the task is to repeatedly estimate the value of the next outcome on the basis of the values of the previous outcomes. Because of the nature of the task at hand, the best possible predictor is of course the true distribution generating the data. However, the true distribution may be unknown, so Solomonoff’s idea was to search for a universally reliable prior: in a nutshell, one guaranteed to converge to the true distribution no matter what the true distribution is. Solomonoff used the theory of Kolmogorov complexity to define a prior that assigns greater probability to hypotheses that have simpler descriptions (relative to the complexity measure known as Kolmogorov complexity): i.e., that are highly compressible. Then, he proved that, provided that the true distribution is effective (an assumption which he took to be realistic, while also allowing to retain a high degree of generality), this prior is indeed reliable in the above sense.

Solomonoff’s result continues to stir much debate and controversy, since it is often taken

to offer a justification of *Occam's razor*:² the principle according to which one should always favour the simplest available hypothesis that fits the observational data. For a thorough discussion of Solomonoff's theory, as well as a critical appraisal of its philosophical import, the reader is invited to consult [Sterkenburg, 2016, 2017, 2018]. For the purpose of this dissertation, suffice it to say that Solomonoff's framework is quite different from ours. In Solomonoff's setting, randomness—or, rather, a lack thereof—is used to define a prior “biased towards simplicity”. Here, on the other hand, we take randomness to be a property of data streams and investigate the effects that random data have on learning. In addition, Solomonoff's work is usually taken to fall squarely within the objective Bayesianism camp, since the data-generating probabilistic source can be naturally interpreted as an objective chance distribution. By contrast, the results in this dissertation are particularly well-suited for a subjectivist reading. In fact, as we will see shortly, the theorems that we focus on, in spite of being amenable to an objectivist interpretation, are generally taken to corroborate the subjectivist viewpoint.

What this dissertation is about. This dissertation comprises four chapters. Chapter 1 provides a concise introduction to the theory of algorithmic randomness. The focus is on two standard approaches to modelling algorithmic randomness, respectively called the *unpredictability paradigm* and the *measure-theoretic typicality paradigm*. These two frameworks highlight two distinct, yet complementary aspects of algorithmic randomness, both of which play an important role in our results. The unpredictability paradigm formalises the intuition that, locally, the algorithmically random data streams are patternless and irregular—and, as a consequence, unpredictable. Because of the absence of any (effective) patterns, the algorithmically random data streams do not possess any distinctive features that could set them apart and that could be exploited for making successful predictions about future observations. The measure-theoretic typicality paradigm, on the other hand, reveals that algorithmic randomness does not amount to absolute lawlessness. As pithily put by Boëthius [1999], “Chance, too, which seems to rush along with slack reins, is bridled and governed by law.” A similar moral can be drawn about randomness: as established by

²For instance, Li and Vitányi state that Solomonoff's induction rests on “an objective and absolute definition of ‘simplicity’ as ‘low Kolmogorov complexity’”. Consequently, one obtains an objective and absolute version of the classic maxim of William of Ockham.” [Li and Vitányi, 2019, p. 262]

the measure-theoretic typicality paradigm, by their very nature, the algorithmically random data streams, while locally irregular and lawless, must obey certain global, statistical laws. It is precisely by virtue of satisfying said statistical laws that the algorithmically random data streams are measure-theoretically typical.

In Chapter 2, Chapter 3, and Chapter 4, we then turn to exploring the ramifications of algorithmic randomness for inductive learning.

Chapter 2 focuses on the connections between algorithmic randomness and Bayesian merging-of-opinions theorems: a collection of results of foundational import for Bayesian epistemology (especially for subjective Bayesianism) that establish under what conditions the posterior probability distributions of two Bayesian agents become arbitrarily close as the shared evidence accumulates (i.e., under what conditions their respective beliefs merge). We argue that two computable Bayesian agents beginning the learning process with different subjective priors, but who nonetheless agree on which data streams are algorithmically random can be viewed as sharing certain inductive assumptions: in particular, as having compatible beliefs about the uniformity of nature. This is because, by virtue of concurring on which data streams are algorithmically random, the two agents are in agreement about which global, effectively specifiable laws they expect the data to satisfy. We then show that the types of doxastic compatibility yielded by algorithmic randomness suffice to ensure that any two computable Bayesian agents will eventually reach a consensus. In other words, they guarantee that the agents' respective probability assignments will almost surely become arbitrarily close as the number of observations increases. Thus, when shared by computable Bayesian learners with different subjective priors, the beliefs about uniformity encoded by algorithmic randomness notions provably lead to merging of opinions. We conclude the chapter by examining the phenomenon of polarisation of opinions—the conditions under which, as the evidence increases, the beliefs of two Bayesian agents grow arbitrarily far apart—from the perspective of the types of doxastic compatibility (and incompatibility) induced by algorithmic randomness.

Chapter 3, on the other hand, investigates Bayesian convergence-to-the-truth theorems from the perspective of algorithmic randomness. Bayesian convergence-to-the-truth results are another staple of Bayesian epistemology: they show under what conditions a Bayesian agent expects their beliefs to align with the truth as the available evidence increases. In this

chapter, we study effectivised versions of a fundamental convergence-to-the-truth result: Lévy’s Upward Theorem [1937]. This means that we consider Lévy’s Upward Theorem in the context of computable Bayesian agents and effective inductive problems—namely, inductive problems whose complexity can be characterised in computability-theoretic terms. We then show that, in this effective setting, the algorithmically random data streams are exactly the ones that ensure that a computable Bayesian agent’s beliefs will asymptotically converge to the truth. In other words, we prove that the collections of truth-conducive data streams systematically correspond to standard algorithmic randomness notions, where the type of randomness that emerges from Lévy’s Upward Theorem crucially depends on the complexity of the inductive problems faced by the Bayesian agent. The results in this chapter were obtained as part of a joint ongoing research project with Simon Huttegger and Sean Walsh, and they also appear in a joint working manuscript titled “Algorithmic randomness and Lévy’s Upward Theorem” [Huttegger et al., 2021]. However, the only characterisation results whose proofs are included in this chapter are the ones that were the author’s individual contribution to the project.

In Chapter 4, rather than exploring the applications of algorithmic randomness in formal models of learning, we study algorithmic randomness itself from a learning-theoretic perspective. Chapter 2 and Chapter 3 rely on the fact that the algorithmically random data streams have to be globally regular (in that they have to satisfy certain statistical laws) and are the most typical outcomes of the underlying probability measure. Chapter 4, on the other hand, hinges on the local irregularity of random sequences. In particular, since algorithmic randomness amounts to local patternlessness, there is a sense in which it is natural to regard the algorithmically random data streams as incompatible with learning (after all, they are unpredictable). In this chapter, we vindicate this intuition by exploring an approach for modelling algorithmic randomness that relies on the conceptual apparatus of formal learning theory. In particular, we show that the algorithmically random data streams are, in a precise sense, unlearnable: they coincide with the sequences that do not display any local patterns that can be extrapolated by computable qualitative learning methods. Our main results are novel learning-theoretic characterisation of two central algorithmic randomness notions: Martin-Löf randomness and Schnorr randomness. A version of this chapter appears in *The Review of Symbolic Logic* [Zaffora Blando, 2021].

Chapter 1

Algorithmic randomness

He deals the cards to find the answer
The sacred geometry of chance
The hidden law of a probable outcome
The numbers lead a dance

Sting, *Shape Of My Heart*

Chance, too, which seems to rush along with
slack reins, is bridled and governed by law.

Boëthius, *The Consolation of Philosophy*

In this chapter we provide a brief overview of the theory of algorithmic randomness,¹ which combines computability theory and measure theory to define what it means for an individual mathematical object to be (product) random.

As traditionally done in the algorithmic randomness literature, we focus on algorithmic randomness notions defined over the *Cantor space* of infinite binary sequences. Algorithmic randomness can be defined in the context of many other spaces;² however, for the purpose

¹For an in-depth treatment of the theory, the reader may consult the textbook by Nies [2009] or the textbook by Downey and Hirschfeldt [2010]. The monograph by Li and Vitányi [2019] is the standard reference for the Kolmogorov complexity approach to algorithmic randomness.

²It is equally standard to define algorithmic randomness in the context of the real numbers (cf. [Downey and Hirschfeldt, 2010]). For more general spaces, such as computable topological spaces and computable metric spaces, see, for instance, [Levin and Zvonkin, 1970], [Hertling and Weihrauch, 2003], [Gács, 2005], [Hoyrup and Rojas, 2009], and [Miyabe, 2014].

of this dissertation, focusing on infinite binary sequences will suffice. In particular, many of the learning scenarios investigated in this dissertation are naturally modelled in this setting, as well.

In spite of their simplicity, infinite binary sequences can in fact be used to capture a variety of learning situations. For instance, each such sequence may be thought of as representing the successive outcomes of an idealised, infinitely long coin tossing experiment, where 1 marks a heads outcome and 0 a tails outcome. Alternatively, infinite binary sequences could be seen as recording the outcomes of a binary test whose goal is repeatedly checking whether a certain property holds or not. In this case, 1 indicates that the property in question has been observed, while 0 indicates that it has not. Yet another possible interpretation is that binary sequences provide a complete description of the world, so that a 1 (respectively, a 0) in position n means that property P_n holds (respectively, does not hold), given some countable list P_0, P_1, P_2, \dots that exhausts all possible properties.

Within the theory of algorithmic randomness, an infinite binary sequence is classified as random if it does not exhibit any “effectively detectable” regularities. This idea is standardly formalised in the context of either one of three different (yet closely connected) paradigms:

- (a) the *incompressibility paradigm*—pioneered by Kolmogorov [1965] and Solomonoff [1964], as well as Chaitin [1966] and Levin [1973]—according to which a sequence is random if none of its initial segments can be produced by a program much shorter than that initial segment itself;
- (b) the *measure-theoretic typicality paradigm*, initiated by Martin-Löf [1966], which identifies randomness with the satisfaction of all statistical laws that can be *effectively specified* (e.g., the Law of Large Numbers and the Law of the Iterated Logarithm);
- (c) the *unpredictability paradigm*—sparked by Ville’s early work on *martingales* [1939] and then further developed by Schnorr [1971b,a] and others in a computability-theoretic setting—which is based on the intuition that no effective gambling strategy can gain unbounded capital by betting against a random sequence.

In what follows, we will discuss in some detail both the measure-theoretic typicality paradigm and the unpredictability paradigm. The focus will be on these two paradigms

because the proofs presented in this dissertation only rely on characterisations of algorithmic randomness notions in terms of either measure-theoretic typicality or unpredictability. For a thorough discussion of the incompressibility paradigm, the reader may consult [Li and Vitányi, 2019].

The structure of this chapter is as follows. We begin with some notational conventions and basic definitions in §1.1. In §1.2, we then introduce the measure-theoretic typicality paradigm. We present two types of statistical tests for randomness used within such paradigm: sequential tests (§1.2.1) and integral tests (§1.2.2). Lastly, in §1.3, we discuss the unpredictability paradigm.

1.1 Notation and basic definitions

Strings and sequences. The set of finite binary strings is denoted by $2^{<\mathbb{N}}$. We use lowercase Greek letters such as σ , τ , and ρ to represent finite strings. The empty string is denoted by ε . Given a string $\sigma \in 2^{<\mathbb{N}}$, $|\sigma|$ stands for its length, $\sigma(n)$ for the $(n+1)$ -st bit of σ , and $\sigma \upharpoonright n$ for the initial segment of σ consisting of its first n bits $\sigma(0)\dots\sigma(n-1)$. If $|\sigma| < n$, $\sigma \upharpoonright n = \sigma$, and if $n = 0$, $\sigma \upharpoonright 0 = \varepsilon$. By $\sigma\tau$, we mean the concatenation of σ and τ . If σ is an initial segment of τ , we write $\sigma \sqsubseteq \tau$; $\sigma \sqsubset \tau$ indicates that the relation is strict. A set $\mathcal{S} \subseteq 2^{<\mathbb{N}}$ is said to be *prefix-free* if and only if, for any two distinct strings $\sigma, \tau \in \mathcal{S}$, neither $\sigma \sqsubseteq \tau$ nor $\tau \sqsubseteq \sigma$ holds. The set of infinite binary sequences is denoted by $2^{\mathbb{N}}$. The elements of this set are denoted by lowercase Greek letters such as ω or α . The terms $|\omega|$, $\omega(n)$, $\omega \upharpoonright n$, $\sigma\omega$, and the strict relation $\sigma \sqsubset \omega$ are defined analogously to the case of finite strings. Note that infinite binary sequences can be interpreted both as sets of natural numbers and as real numbers in the unit interval: a sequence $\omega \in 2^{\mathbb{N}}$ naturally corresponds to the set $S_\omega = \{n \in \mathbb{N} : \omega(n) = 1\}$ (in other words, ω is the characteristic function of S_ω) and to the real number $r_\omega = \sum_{n \in \mathbb{N}} \omega(n) \cdot 2^{-n} \in [0, 1]$.

Computability. We assume some familiarity with the basic notions of computability theory over the natural numbers \mathbb{N} , such as computable functions, computable sets, and computably enumerable sets. The notion of a computable function from \mathbb{N} to \mathbb{N} can be extended to functions from D to D' , as long as the sets D and D' can be identified in a

computable way with \mathbb{N} or a subset thereof. This includes, for instance, $\mathbb{N} \times \mathbb{N}$, the rationals \mathbb{Q} , $\{0, 1\}$, and $2^{<\mathbb{N}}$. A (total) function $f : D \rightarrow \mathbb{R}$ is *computable* if there is a two-place computable function $h : D \times \mathbb{N} \rightarrow \mathbb{Q}$ such that, for all $(x, n) \in D \times \mathbb{N}$, $|h(x, n) - f(x)| \leq 2^{-n}$ (that is, if f can be effectively approximated by a rational-valued function with any given precision). A partial function $f : \subseteq D \rightarrow \mathbb{R}$ is *partial computable* if there is a two-place partial computable function $h : \subseteq D \times \mathbb{N} \rightarrow \mathbb{Q}$ such that $x \in \text{dom}(f)$ if and only if $(x, n) \in \text{dom}(h)$ for all $n \in \mathbb{N}$, and, for all $(x, n) \in \text{dom}(h)$, $|h(x, n) - f(x)| \leq 2^{-n}$. A (total) function $f : D \rightarrow \mathbb{R}$ is *left-computably enumerable* (left-c.e., from now on) if there is a computable function $h : D \times \mathbb{N} \rightarrow \mathbb{Q}$ such that, for all $(x, n) \in D \times \mathbb{N}$, the sequence $\{h(x, n)\}_{n \in \mathbb{N}}$ is non-decreasing and converges to $f(x)$: i.e., $\lim_{n \rightarrow \infty} h(x, n) = f(x)$. If $f : D \rightarrow \mathbb{R}$ is left-c.e., the set $\{(x, q) \in D \times \mathbb{Q} : f(x) > q\}$ is computably enumerable. A partial left-c.e. function $f : \subseteq D \rightarrow \mathbb{R}$ is defined in the obvious way. Similarly, a (total) function $f : D \rightarrow \mathbb{R}$ is *right-computably enumerable* (right-c.e.) if there is a computable function $h : D \times \mathbb{N} \rightarrow \mathbb{Q}$ such that, for all $(x, n) \in D \times \mathbb{N}$, the sequence $\{h(x, n)\}_{n \in \mathbb{N}}$ is non-increasing and converges to $f(x)$. If $f : D \rightarrow \mathbb{R}$ is right-c.e., the set $\{(x, q) \in D \times \mathbb{Q} : f(x) \geq q\}$ is co-computably enumerable. A function is both left-c.e. and right-c.e. if and only if it is computable. Again, the notion of a partial right-c.e. function is defined in the obvious way. A sequence of rational numbers q_0, q_1, q_2, \dots is computable if there is a computable function $f : \mathbb{N} \rightarrow \mathbb{Q}$ which, on input n , outputs the n -th rational number in the sequence. A real number $r \in \mathbb{R}$ is *computable* if there is a computable sequence q_0, q_1, q_2, \dots of rationals such that $|r - q_n| \leq 2^{-n}$ for all $n \in \mathbb{N}$. In other words, the q_n 's get closer and closer to r at a computable rate, uniformly in n . A sequence r_0, r_1, r_2, \dots of computable reals is a sequence of *uniformly computable* reals if there is a computable function $f : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{Q}$ which, on input (n, m) , outputs the m -th rational number in the approximation of r_n . If $r \in \mathbb{R}$ is such that there is a computable sequence r_0, r_1, r_2, \dots of uniformly computable reals with $|r - r_n| \leq 2^{-n}$ for all $n \in \mathbb{N}$, then r is computable. A real number r is *left-computably enumerable* (left-c.e.) if there is a computable monotonically increasing sequence q_0, q_1, q_2, \dots of rationals that converges to r in the limit (equivalently, a computable non-decreasing sequence q_0, q_1, q_2, \dots of rationals that converges to r in the limit). Similarly, a real number r is *right-computably enumerable* (right-c.e.) if there is a computable monotonically decreasing (equivalently, non-increasing) sequence q_0, q_1, q_2, \dots

of rationals that converges to r in the limit. A real is both left-c.e. and right-c.e. if and only if it is computable. An infinite binary sequence $\omega \in 2^{\mathbb{N}}$ is said to be computable if it is computable when seen as a function from \mathbb{N} to $\{0, 1\}$. It is left-c.e. if it is the binary expansion of a left-c.e. real $r \in [0, 1)$, and it is right-c.e. if it is the binary expansion of a right-c.e. real $r \in [0, 1)$.

Cantor space: topology and measure. Recall that a *topological space* consists of a set Ω , together with a *topology* \mathcal{T} on it: namely, a family of subsets of Ω that contains both the empty set \emptyset and Ω itself, and that is closed under arbitrary unions and finite intersections. The elements of the topology \mathcal{T} are called the *open* sets in Ω , while a subset of Ω is said to be *closed* if its complement is open in Ω . If a subset of Ω is both open and closed, it is called *clopen*. A *base* \mathcal{B} for a topological space (Ω, \mathcal{T}) is a collection of subsets of Ω such that every open set in Ω can be written as a union of elements of \mathcal{B} . Then, \mathcal{B} is said to generate \mathcal{T} .

The Cantor space of infinite binary sequences is the topological space consisting of $2^{\mathbb{N}}$ together with the *topology of pointwise convergence*. A base for this topology is given by the clopen *cylinders* $[\sigma]$, where $\sigma \in 2^{<\mathbb{N}}$ is a finite binary string and $[\sigma] = \{\omega \in 2^{\mathbb{N}} : \sigma \sqsubset \omega\}$ is the set of all infinite binary sequences that extend σ . Without loss of generality, every open set \mathcal{U} in $2^{\mathbb{N}}$ can be written as a disjoint union of cylinders: that is, $\mathcal{U} = \bigcup \{[\sigma] \subseteq 2^{\mathbb{N}} : \sigma \in \mathcal{S}\}$ for some prefix-free set of strings $\mathcal{S} \subseteq 2^{<\mathbb{N}}$. For ease of notation, $\bigcup \{[\sigma] \subseteq 2^{\mathbb{N}} : \sigma \in \mathcal{S}\}$ will be abbreviated as $[\mathcal{S}]$.

As mentioned in the introduction, every notion of algorithmic randomness hinges on a probability measure fixed in advance. Each measure determines a specific set of algorithmically random sequences, and a sequence which is random relative to a given measure may not be random relative to a different measure.

Recall that a σ -algebra \mathcal{F} on a set Ω is a collection of subsets of Ω that includes Ω itself, and which is closed under complement and countable unions. A probability measure $\mu : \mathcal{F} \rightarrow \mathbb{R}$ on \mathcal{F} is a function whose values lie in the unit interval $[0, 1]$, that returns 0 for the empty set \emptyset and 1 for Ω , and that is *countably additive*: namely, for all countable collections $\{A_n\}_{n \in \mathbb{N}}$ of pairwise disjoint elements of \mathcal{F} , $\mu(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mu(A_n)$. A triple of the form $(\Omega, \mathcal{F}, \mu)$ —with Ω a set, \mathcal{F} a σ -algebra on Ω , and μ a probability measure

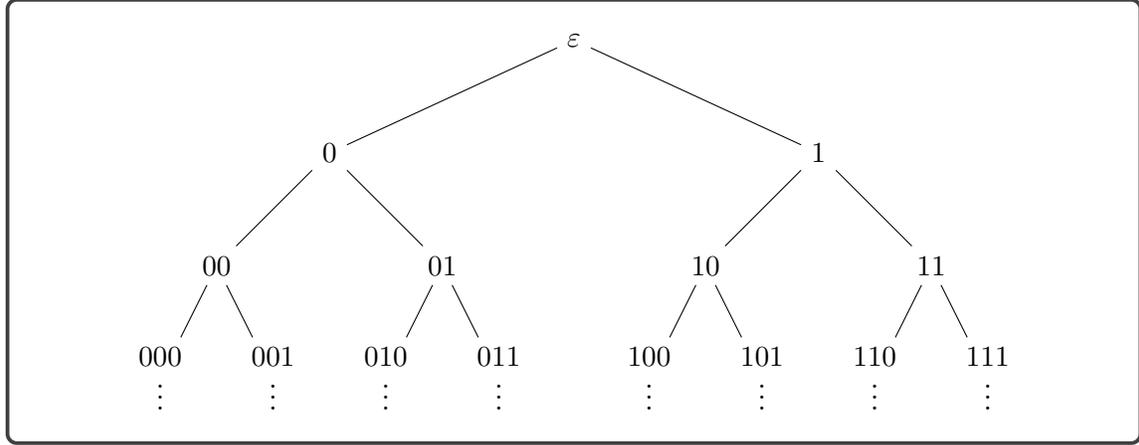


Figure 1.1: The binary Cantor tree of height ω_0 , where each node corresponds to an initial segment of some sequence in $2^{\mathbb{N}}$. The cylinder $[\sigma] \subseteq 2^{\mathbb{N}}$ generated by $\sigma \in 2^{<\mathbb{N}}$ corresponds to the collection of all paths in this tree that go through the node labelled by σ .

on \mathcal{F} —is called a *probability space*. In what follows, we will focus on probability spaces of the form $(2^{\mathbb{N}}, \mathcal{B}(2^{\mathbb{N}}), \mu)$, where $\mathcal{B}(2^{\mathbb{N}})$ is the *Borel σ -algebra* on $2^{\mathbb{N}}$ —i.e., the smallest σ -algebra containing all open sets of Cantor space—and $\mu : \mathcal{B}(2^{\mathbb{N}}) \rightarrow [0, 1]$ is a probability measure on $\mathcal{B}(2^{\mathbb{N}})$.

By *Caratheodory's Extension Theorem*, any function m defined on cylinders that takes values in $[0, 1]$, and such that $m([\varepsilon]) = 1$ and, for all $\sigma \in 2^{<\mathbb{N}}$, $m([\sigma]) = m([\sigma 0]) + m([\sigma 1])$ can be uniquely extended to a probability measure on $\mathcal{B}(2^{\mathbb{N}})$. Hence, from now on, probability measures on $\mathcal{B}(2^{\mathbb{N}})$ will be identified with their restriction to cylinders. A canonical probability measure on $\mathcal{B}(2^{\mathbb{N}})$ is the *uniform* (or *Lebsegue*) *measure* λ , given by $\lambda([\sigma]) = 2^{-|\sigma|}$ for all $\sigma \in 2^{<\mathbb{N}}$ —where, as defined earlier, $|\sigma|$ denotes the length of σ . We will also make use of the notion of a *semi-measure*: namely, a function δ defined on cylinders and taking values in $[0, 1]$ such that $\delta([\varepsilon]) \leq 1$ and, for all $\sigma \in 2^{<\mathbb{N}}$, $\delta([\sigma]) \geq \delta([\sigma 0]) + \delta([\sigma 1])$.

Since the only type of measures considered here are probability measures, from now on we will simply refer to them as measures. Moreover, throughout this dissertation, we shall restrict attention to *computable*, *lower semi-computable*, and *upper semi-computable* measures (and semi-measures), which are defined as follows:

Definition 1.1.1 (Computable, lower semi-computable, and upper semi-computable measures). *A measure μ on $\mathcal{B}(2^{\mathbb{N}})$ is computable if the function $\sigma \mapsto \mu([\sigma])$ is computable: that*

is, if $\mu([\sigma])$ is a computable real, uniformly in σ . This means that there exists a computable two-place function $f : 2^{<\mathbb{N}} \times \mathbb{N} \rightarrow \mathbb{Q}$ which, on inputs σ and n , outputs the n -th rational in the approximation that witnesses the computability of $\mu([\sigma])$. Analogously, μ is a lower semi-computable measure if $\mu([\sigma])$ is a left-c.e. real, uniformly in σ , and it is an upper semi-computable measure if $\mu([\sigma])$ is a right-c.e. real, uniformly in σ . The notions of computable, lower semi-computable, and upper semi-computable semi-measure are defined in an analogous way.

The uniform measure λ on $\mathcal{B}(2^{\mathbb{N}})$ defined above is an obvious example of a computable measure, as is any Bernoulli measure with a computable bias.

1.2 The measure-theoretic typicality paradigm

According to the *measure-theoretic typicality* paradigm, a sequence is algorithmically random if it is an effectively typical outcome of the underlying measure: namely, if it satisfies certain “effectively specifiable” properties that hold with probability one (certain effectively specifiable statistical laws) relative to the underlying measure. Equivalently, a sequence is algorithmically random if it cannot be effectively discovered to be atypical: i.e., if it cannot be determined to violate any relevant effectively specifiable statistical law via suitable randomness tests.

1.2.1 Sequential tests

A property is effectively specifiable if it coincides with a subset of Cantor space that is definable by an arithmetical formula. Within the arithmetical hierarchy, a subset of Cantor space is assigned classifications of the form Π_n^0 , Σ_n^0 , or Δ_n^0 , with n a natural number, as follows:

Definition 1.2.1 (Π_n^0 , Σ_n^0 , and Δ_n^0 classes). *A set $\mathcal{C} \subseteq 2^{\mathbb{N}}$ is a Π_n^0 class if it is definable by a Π_n^0 formula: that is, if there is a computable relation R such that*

$$\mathcal{C} = \{\omega \in 2^{\mathbb{N}} : (\forall k_1)(\exists k_2)\dots(Qk_n) R(\omega \upharpoonright k_1, \omega \upharpoonright k_2, \dots, \omega \upharpoonright k_n)\},$$

where $Q = \forall$ if n is odd and $Q = \exists$ if n is even.

A Σ_n^0 class, on the other hand, is the complement of a Π_n^0 class. Equivalently, it is a set $C \subseteq 2^{\mathbb{N}}$ definable by a Σ_n^0 formula, which means that there exists a computable relation R such that

$$C = \{\omega \in 2^{\mathbb{N}} : (\exists k_1)(\forall k_2)\dots(Qk_n) R(\omega \upharpoonright k_1, \omega \upharpoonright k_2, \dots, \omega \upharpoonright k_n)\},$$

where $Q = \exists$ if n is odd and $Q = \forall$ if n is even.

Lastly, a Δ_n^0 class is a set $C \subseteq 2^{\mathbb{N}}$ that is both a Π_n^0 class and a Σ_n^0 class.

Every Π_n^0 class is also a Δ_{n+1}^0 , a Π_{n+1}^0 and a Σ_{n+1}^0 class, and the same holds for Σ_n^0 classes. A Σ_1^0 class is the Cantor space analogue of a computably enumerable set of natural numbers, while a Δ_1^0 class is the analogue of a computable set of natural numbers. In fact, the Σ_1^0 subsets of Cantor space are exactly the ones that are generated by computably enumerable sets of binary strings: that is, C is a Σ_1^0 class if and only if there is a computably enumerable prefix-free set $\mathcal{S} \subseteq 2^{<\mathbb{N}}$ such that $C = [\mathcal{S}]$.³ A Δ_1^0 class C , on the other hand, is one for which there is a finite prefix-free set $\mathcal{S} \subset 2^{<\mathbb{N}}$ such that $C = [\mathcal{S}]$. Thus, Σ_1^0 classes are the effectively open subsets of Cantor space, Π_1^0 classes are the effectively closed subsets, while Δ_1^0 classes are simply the clopen subsets. Similarly, Σ_2^0 classes are the effective analogue of F_σ sets (countable unions of closed sets), while Π_2^0 classes are the effective analogue of G_δ sets (countable intersections of open sets). These analogies of course propagate upwards throughout both hierarchies (the arithmetical hierarchy and the Borel hierarchy). A sequence $\{\mathcal{C}_n\}_{n \in \mathbb{N}}$ of Σ_n^0 classes is said to be a sequence of *uniformly* Σ_n^0 classes if $\bigcap_{n \in \mathbb{N}} \mathcal{C}_n$ is a Π_{n+1}^0 class, while a sequence $\{\mathcal{D}_n\}_{n \in \mathbb{N}}$ of Π_n^0 classes is said to be a sequence of *uniformly* Π_n^0 classes if $\bigcup_{n \in \mathbb{N}} \mathcal{D}_n$ is a Σ_{n+1}^0 class.

We will now see that a canonical statistical law, the Strong Law of Large Numbers, corresponds to an effectively specifiable property in the above sense.

Example 1.2.2 (The Strong Law of Large Numbers). *Fix the uniform measure λ and, given a sequence $\omega \in 2^{\mathbb{N}}$, let $\frac{\#0(\omega \upharpoonright k)}{k}$ denote the relative frequency of 0's in the first k digits of ω . The set of sequences that satisfy the Strong Law of Large Numbers (relative to λ) corresponds to the following subset of Cantor space, which is definable by a Π_3^0 formula*

³For a proof, see, for instance, [Downey and Hirschfeldt, 2010, Proposition 2.19.2, p. 74].

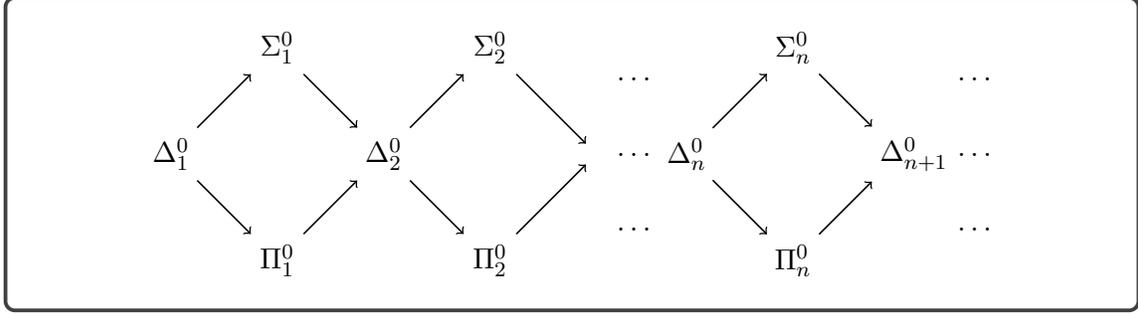


Figure 1.2: The arithmetical hierarchy of subsets of Cantor space.

(thereby being a Π_3^0 class):

$$\left\{ \omega \in 2^{\mathbb{N}} : (\forall n)(\exists m)(\forall k > m) \left| \frac{\#0(\omega \upharpoonright k)}{k} - \frac{1}{2} \right| \leq 2^{-n} \right\}$$

A sequence $\alpha \in 2^{\mathbb{N}}$ fails the Strong Law of Large Numbers (relative to λ) if there is some $n \in \mathbb{N}$ such that α belongs to the Π_2^0 class $\{ \omega \in 2^{\mathbb{N}} : (\forall m)(\exists k > m) \left| \frac{\#0(\omega \upharpoonright k)}{k} - \frac{1}{2} \right| > 2^{-n} \}$.

Arguably, the most prominent notion of algorithmic randomness in the literature is Martin-Löf randomness [1966]. Part of the rationale behind Martin-Löf’s definition comes from statistical hypothesis testing: effectively determining whether a probability-one law has been violated can be thought of as performing a *sequential statistical test* for randomness. Given a sequence $\omega \in 2^{\mathbb{N}}$, the conjecture (or null hypothesis) is that ω is an effectively typical outcome of the underlying measure: that it satisfies all relevant effective statistical laws. Then, ω is categorised as random if and only if it passes all effective sequential statistical tests for randomness.

Statistical hypothesis testing prescribes that a hypothesis be discarded if, upon supposing that said hypothesis is true, one observes a statistically significant outcome according to some pre-specified significance level. Martin-Löf randomness is defined in terms of tests whose significance levels are determined by a computable function $f : \mathbb{N} \rightarrow \mathbb{Q}$ with $\lim_{n \rightarrow \infty} f(n) = 0$ —without loss of generality, computable significance levels of the form 2^{-n} . A sequence ω is rejected at level 2^{-n} if and only if there is some $m \in \mathbb{N}$ such that we would reject the initial segment $\omega \upharpoonright m$ of ω at level 2^{-n} . On the other hand, a sequence ω is rejected *simpliciter* if, for every such significance level, there is an initial segment of ω that

we would discard at that level. A Martin-Löf random sequence is one which is *not* rejected at every significance level.

From now on, let μ denote an arbitrary computable measure. Martin-Löf randomness relative to μ is then defined as follows:

Definition 1.2.3 (Martin-Löf randomness).

- (a) Let $\{\mathcal{U}_n\}_{n \in \mathbb{N}}$ be a sequence of uniformly Σ_1^0 classes satisfying $\mu(\mathcal{U}_n) \leq 2^{-n}$ for all $n \in \mathbb{N}$. Such a sequence is called a sequential μ -Martin-Löf test.
- (b) A sequence $\omega \in 2^{\mathbb{N}}$ is μ -Martin-Löf random if and only if there is no sequential μ -Martin-Löf test $\{\mathcal{U}_n\}_{n \in \mathbb{N}}$ such that $\omega \in \bigcap_{n \in \mathbb{N}} \mathcal{U}_n$.

The collection of μ -Martin-Löf random sequences, which will be denoted by μ -MLR, is a set of μ -measure one. This is because the above definition ensures that there are only countably many sequential μ -Martin-Löf tests. Hence, the set of all sequences that fail at least one sequential μ -Martin-Löf test is a countable collection of μ -null sets and so, by countable additivity, a μ -null set itself. In fact, for analogous reasons, $\mu(\mu\text{-R}) = 1$ for every algorithmic randomness notion R , so we will not argue for it each time.

Note that, without loss of generality, sequential μ -Martin-Löf tests can be assumed to be nested. This is because, given any sequential μ -Martin-Löf test $\{\mathcal{U}_n\}_{n \in \mathbb{N}}$, one can construct a nested sequential μ -Martin-Löf test $\{\mathcal{V}_n\}_{n \in \mathbb{N}}$ such that $\bigcap_{n \in \mathbb{N}} \mathcal{V}_n = \bigcap_{n \in \mathbb{N}} \mathcal{U}_n$. This is done as follows: for every n , let $\mathcal{V}_n = \bigcup_{m > n} \mathcal{U}_m$. Since Σ_1^0 classes are closed under countable unions, $\{\mathcal{V}_n\}_{n \in \mathbb{N}}$ is a sequence of uniformly Σ_1^0 classes. Moreover, for each n , $\mu(\mathcal{V}_n) \leq \sum_{m > n} \mu(\mathcal{U}_m) \leq \sum_{m > n} 2^{-m} = 2^{-n}$. This establishes that $\{\mathcal{V}_n\}_{n \in \mathbb{N}}$ is a sequential μ -Martin-Löf test. Moreover, it is easy to see that $\mathcal{V}_0 \supseteq \mathcal{V}_1 \supseteq \mathcal{V}_2 \supseteq \dots$ and $\bigcap_{n \in \mathbb{N}} \mathcal{V}_n = \bigcap_{n \in \mathbb{N}} \mathcal{U}_n$.

An important feature of Martin-Löf randomness is the existence of *universal* Martin-Löf tests—where a sequential μ -Martin-Löf test $\{\mathcal{V}_n\}_{n \in \mathbb{N}}$ is universal when, for every sequential μ -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, to determine whether a sequence is μ -Martin-Löf random, it suffices to check whether it passes a *single* universal test.⁴

⁴The proof of the existence of universal sequential μ -Martin-Löf tests is due to Martin-Löf [1966] and it crucially relies on the fact that there is an effective enumeration of all sequential μ -Martin-Löf tests (see, for instance, [Downey and Hirschfeldt, 2010, pp. 233-234]).

Example 1.2.4. We now give an example (arguably, the best-known example) of a Martin-Löf random sequence. Let U be a universal prefix-free Turing machine⁵ and λ the uniform measure. Then, Chaitin’s halting probability Ω_U relative to U is the real number

$$\Omega_U = \lambda([\text{dom}(U)]) = \sum_{\sigma \in \text{dom}(U)} 2^{-|\sigma|}.$$

The binary expansion of Ω_U , which we also denote by Ω_U , is a λ -Martin-Löf random sequence (cf. [Chaitin, 1975] or [Downey and Hirschfeldt, 2010, Theorem 6.1.3]). Moreover, it is also left-computably enumerable.

A sequence $\{\mathcal{U}_n\}_{n \in \mathbb{N}}$ of uniformly Σ_1^0 classes such that $\sum_{n \in \mathbb{N}} \mu(\mathcal{U}_n) < \infty$ is called a μ -Solovay test. Martin-Löf randomness can also be characterised via Solovay tests:⁶

Theorem 1.2.5 (Solovay [1975]). *Let $\omega \in 2^{\mathbb{N}}$. The following are equivalent:*

- (1) ω is μ -Martin-Löf random;
- (2) for all μ -Solovay tests $\{\mathcal{U}_n\}_{n \in \mathbb{N}}$, there are at most finitely many n such that $\omega \in \mathcal{U}_n$.

By modifying the effectivity constraints imposed on sequential tests for randomness, one can obtain more or less logically strong notions of algorithmic randomness. As we will see, the other core algorithmic randomness notions that entail Martin-Löf randomness are defined in terms of more lenient tests, while the randomness notions that are entailed by Martin-Löf randomness are defined in terms of more demanding tests.

The first such notion that we consider here is Schnorr randomness, introduced by Schnorr [1971a,b], which is weaker⁷ than Martin-Löf randomness.⁸

⁵A Turing machine is prefix-free if its domain is a prefix-free set.

⁶The characterisation of Martin-Löf randomness (and, as we will see later, of other standard algorithmic randomness notions) in terms of Solovay tests is an effective analogue of the Borel-Cantelli Lemma from probability theory.

⁷Unless otherwise specified, we will use “weaker” and “stronger” in the non-strict sense. For most reasonable probability measures, the implications between algorithmic randomness notions are strict, but there are also measures (e.g., Dirac measures concentrated on a single sequence) that make the algorithmic randomness hierarchy collapse.

⁸See, for instance, [Schnorr, 1971a] or [Downey and Griffiths, 2004] for some examples of sequences (or reals) that are λ -Schnorr random but not λ -Martin-Löf random.

Definition 1.2.6 (Schnorr randomness).

- (a) Let $\{\mathcal{U}_n\}_{n \in \mathbb{N}}$ be a sequential μ -Martin-Löf test such that the measures $\mu(\mathcal{U}_n)$ are computable reals, uniformly in n . Then, $\{\mathcal{U}_n\}_{n \in \mathbb{N}}$ is called a sequential μ -Schnorr test.
- (b) A sequence $\omega \in 2^{\mathbb{N}}$ is μ -Schnorr random if and only if there is no sequential μ -Schnorr test $\{\mathcal{U}_n\}_{n \in \mathbb{N}}$ such that $\omega \in \bigcap_{n \in \mathbb{N}} \mathcal{U}_n$.

The collection of μ -Schnorr random sequences will be denoted by μ -SR.

Just as in the case of Martin-Löf randomness, the values 2^{-n} may be replaced by the values of any other computable function $f : \mathbb{N} \rightarrow \mathbb{Q}$ with $\lim_{n \rightarrow \infty} f(n) = 0$.

Schnorr randomness can also be characterised in terms of Solovay tests. The ones that give rise to Schnorr randomness are known as *total μ -Solovay tests*: namely, sequences $\{\mathcal{U}_n\}_{n \in \mathbb{N}}$ of uniformly Σ_1^0 classes such that $\sum_{n \in \mathbb{N}} \mu(\mathcal{U}_n)$ is not only finite, but also a computable real.

Theorem 1.2.7 (Downey and Griffiths [2004]). *Let $\omega \in 2^{\mathbb{N}}$. The following are equivalent:*

- (1) ω is μ -Schnorr random;
- (2) for all total μ -Solovay tests $\{\mathcal{U}_n\}_{n \in \mathbb{N}}$, there are at most finitely many n such that $\omega \in \mathcal{U}_n$.

Next, we consider another algorithmic randomness notion introduced by Schnorr [1971a,b], computable randomness, which is known to be weaker than Martin-Löf randomness but stronger than Schnorr randomness.⁹

Definition 1.2.8 (Computable randomness).

- (a) Let $\{\mathcal{U}_n\}_{n \in \mathbb{N}}$ be a sequential μ -Martin-Löf test for which there is a computable measure ν such that, for all $n \in \mathbb{N}$ and $\sigma \in 2^{<\mathbb{N}}$, $\mu(\mathcal{U}_n \cap [\sigma]) \leq 2^{-n} \nu([\sigma])$. Then, $\{\mathcal{U}_n\}_{n \in \mathbb{N}}$ is called a bounded sequential μ -Martin-Löf test.

⁹Computable randomness was defined by Schnorr in terms of betting strategies—that is, within the unpredictability paradigm that we will discuss in §1.3. The definition of computable randomness in terms of sequential tests given in Definition 1.2.8 is due to Merkle et al. [2006]. Another characterisation of computable randomness in terms of sequential tests is due to Downey et al. [2004]. For an example of a sequence that is λ -computably random but not λ -Martin-Löf random, see, for instance, [Schnorr, 1971a, Satz 7.2], [Schnorr, 1971b, Theorem 3.2], or [Wang, 1996, Theorem 3.2.1]. For an example of a sequence that is λ -Schnorr random but not λ -computably random, on the other hand, see, for instance, [Wang, 1996, Theorem 3.2.2].

- (b) A sequence $\omega \in 2^{\mathbb{N}}$ is μ -computably random if and only if there is no bounded sequential μ -Martin-Löf test $\{\mathcal{U}_n\}_{n \in \mathbb{N}}$ such that $\omega \in \bigcap_{n \in \mathbb{N}} \mathcal{U}_n$.

The collection of μ -computably random sequences will be denoted by μ -CR.

We conclude by defining an entire collection of algorithmic randomness notions: the weak n -randomness family.¹⁰

Definition 1.2.9 (Weak n -randomness). *Let $n \geq 1$. A sequence $\omega \in 2^{\mathbb{N}}$ is μ -weakly n -random if and only if it belongs to every Σ_n^0 class of μ -measure one.*

The collection of μ -weakly n -random sequences will be denoted by μ -WnR.

To see that this way of defining randomness is in line with the other definitions given above, note that belonging to every Σ_n^0 class of μ -measure one is equivalent to avoiding all Π_n^0 classes of μ -measure zero. Now, take a sequential μ -Martin-Löf test $\{\mathcal{U}_n\}_{n \in \mathbb{N}}$. Clearly, $\bigcap_{n \in \mathbb{N}} \mathcal{U}_n$ is a Π_2^0 class of μ -measure zero. More precisely, since the definition of a sequential μ -Martin-Löf test requires that there be a computable bound on the rate of convergence of the measures $\mu(\mathcal{U}_n)$, $\bigcap_{n \in \mathbb{N}} \mathcal{U}_n$ is a Π_2^0 class of *effective* μ -measure zero. Hence, being μ -Martin-Löf random amounts to avoiding all Π_2^0 classes of effective μ -measure zero.

For every $n \geq 1$, μ -weak $(n + 1)$ -randomness is stronger than μ -weak n -randomness. Moreover, μ -weak 2-randomness is stronger than μ -Martin-Löf randomness,¹¹ and μ -weak 1-randomness is weaker than μ -Schnorr randomness.¹² Among the notions in the weak n -randomness hierarchy, μ -weak 2-randomness and μ -weak 1-randomness are the ones that we shall mostly focus on in this dissertation, as they arguably amount to the logically strongest and the logically weakest core algorithmic randomness notions.

Both μ -weak 2-randomness and μ -weak 1-randomness can also be characterised in terms of sequential tests. Call a sequence $\{\mathcal{U}_n\}_{n \in \mathbb{N}}$ of uniformly Σ_1^0 classes with $\lim_{n \rightarrow \infty} \mu(\mathcal{U}_n) = 0$ a *generalised sequential μ -Martin-Löf test*. For any such test, the set $\bigcap_{n \in \mathbb{N}} \mathcal{U}_n$ is a Π_2^0 class of μ -measure zero, so the following proposition is immediate.

¹⁰The first systematic treatment of weak n -randomness is due to Kurtz [1981]. The first occurrence of these notions in print is in [Gaifman and Snir, 1982], although they already appear in [Solovay, 1975].

¹¹Recall that a sequence is Δ_2^0 if it is computable in the halting problem. As shown by Martin (cf. [Solovay, 1975]) and [Downey and Hirschfeldt, 2010, Corollary 7.2.9]), there are no Δ_2^0 λ -weakly 2-random sequences. However, as evidenced by Example 1.2.4, there are left-c.e. (and, so, Δ_2^0) λ -Martin-Löf random sequences.

¹²As shown by, for instance, Kurtz [1981] and Kautz [1991], there are λ -weakly 1-random sequences that do not satisfy the Strong Law of Large Numbers. However, all λ -Schnorr random sequences satisfy the Strong Law of Large Numbers (see, for example, Schnorr [1971a] and van Lambalgen [1987a]).

Proposition 1.2.10 (Folklore). *Let $\omega \in 2^{\mathbb{N}}$. The following are equivalent:*

- (1) ω is μ -weakly 2-random;
- (2) there is no generalised sequential μ -Martin-Löf test $\{\mathcal{U}_n\}_{n \in \mathbb{N}}$ such that $\omega \in \bigcap_{n \in \mathbb{N}} \mathcal{U}_n$.

The characterisation of μ -weak 1-randomness in terms of sequential tests, on the other hand, is due to Wang [1996]. A *sequential μ -weak 1-randomness test* is a sequential μ -Martin-Löf test $\{\mathcal{U}_n\}_{n \in \mathbb{N}}$ such that, for all $n \in \mathbb{N}$, $\mathcal{U}_n = [S_n]$ —where $\{S_n\}_{n \in \mathbb{N}}$ is a sequence of finite, prefix-free, uniformly computable subsets of $2^{<\mathbb{N}}$. This entails that both $\{\mathcal{U}_n\}_{n \in \mathbb{N}}$ and $\{\overline{\mathcal{U}}_n\}_{n \in \mathbb{N}}$ are sequences of uniformly Δ_1^0 classes. Then, the following holds:

Theorem 1.2.11 (Wang [1996]). *Let $\omega \in 2^{\mathbb{N}}$. The following are equivalent:*

- (1) ω is μ -weakly 1-random;
- (2) there is no sequential μ -weak 1-randomness test $\{\mathcal{U}_n\}_{n \in \mathbb{N}}$ such that $\omega \in \bigcap_{n \in \mathbb{N}} \mathcal{U}_n$.

Next, we will see that the algorithmic randomness notions defined here (see Figure 1.3) can also be characterised via randomness tests of a different kind.

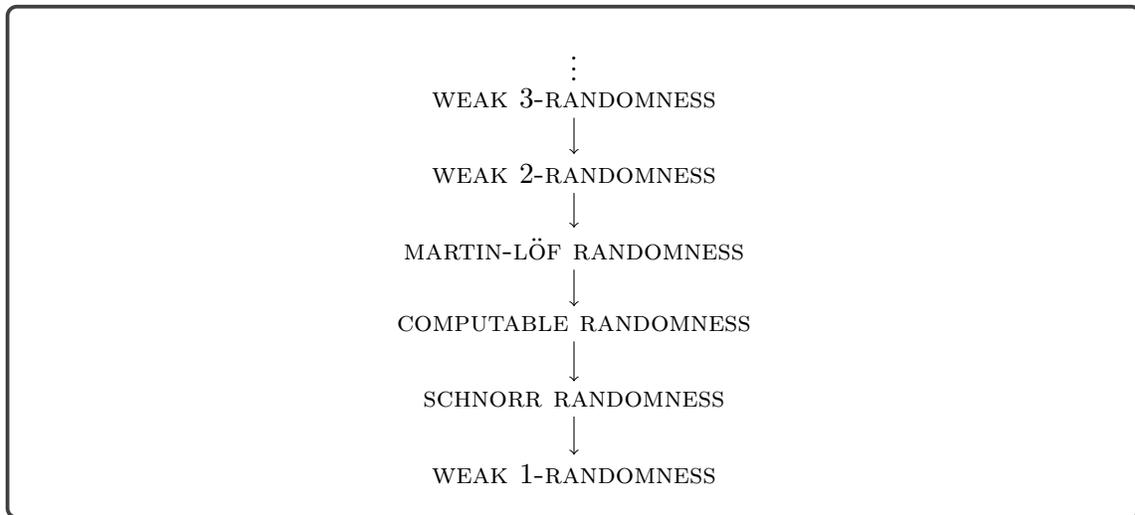


Figure 1.3: The algorithmic randomness notions introduced in §1.2.1, arranged according to logical strength. An arrow going from some randomness notion R to some other randomness notion R' indicates that, for *all* computable measures μ , μ - $R \subseteq \mu$ - R' . None of the displayed implications can be reversed.

1.2.2 Integral tests

Let $\overline{\mathbb{R}}$ denote $\mathbb{R} \cup \{+\infty\}$ (since we will not be dealing with $-\infty$, we shall write ∞ , rather than $+\infty$). Within the measure-theoretic typicality paradigm, algorithmic randomness can also be defined in terms of a different class of tests, called *integral tests* for randomness: roughly, classes of effectively approximable functions of the form $f : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ which meet certain additional measure-theoretic conditions. As we will see, integral tests will be particularly useful in Chapter 3, which focuses on Bayesian convergence to the truth, and Chapter 4, which is devoted to studying algorithmic randomness from a formal learning-theoretic perspective.

The relevant notion of effective approximability is captured in terms of *lower semi-computable* functions:

Definition 1.2.12 (Lower semi-computable function). *A function $f : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ is lower semi-computable if there is a sequence of uniformly computable rational-valued functions $f_k : 2^{<\mathbb{N}} \rightarrow \mathbb{Q}$ such that*

- (a) $f_{k+1}(\sigma) \geq f_k(\sigma)$ for all $k \geq 0$ and $\sigma \in 2^{<\mathbb{N}}$;
- (b) $f_k(\sigma\tau) \geq f_k(\sigma)$ for all $k \geq 0$ and all $\sigma, \tau \in 2^{<\mathbb{N}}$;
- (c) $f(\omega) = \sup\{f_k(\omega \upharpoonright n) : k, n \geq 0\}$ for all $\omega \in 2^{\mathbb{N}}$.

We will often make use of the following equivalent characterisation of lower semi-computable functions, so we rehearse its proof below.

Proposition 1.2.13 (Folklore). *A function $f : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ is lower semi-computable if and only if, for all $q \in \mathbb{Q}$, $f^{-1}((q, \infty]) = \{\omega \in 2^{\mathbb{N}} : f(\omega) > q\}$ is a Σ_1^0 class, uniformly in q .*

Proof. (\Rightarrow) Let $\{f_k\}_{k \in \mathbb{N}}$ be the sequence of uniformly computable functions witnessing the lower semi-computability of f . Fix $q \in \mathbb{Q}$. Then,

$$\begin{aligned} \{\omega \in 2^{\mathbb{N}} : f(\omega) > q\} &= \{\omega \in 2^{\mathbb{N}} : \sup\{f_k(\omega \upharpoonright n) : k, n \geq 0\} > q\} \\ &= \{\omega \in 2^{\mathbb{N}} : \exists n \exists k f_k(\omega \upharpoonright n) > q\} \end{aligned}$$

Hence, $\{\omega \in 2^{\mathbb{N}} : f(\omega) > q\}$ is a Σ_1^0 class, uniformly in q .

(\Leftarrow) Since f is lower semi-continuous, it is bounded below. To see this, for each $n > 0$, let

$\mathcal{U}_n = \{\omega \in 2^{\mathbb{N}} : f(\omega) > -n\}$. Then, $\mathcal{U}_n \subseteq \mathcal{U}_{n+1}$ and $\bigcup_{n>0} \mathcal{U}_n = 2^{\mathbb{N}}$. So, by the compactness of Cantor space,¹³ there is some $n_0 > 0$ such that $\mathcal{U}_{n_0} = 2^{\mathbb{N}}$. Hence, f is indeed bounded below. Let q_0, q_1, q_2, \dots be a computable enumeration of the rationals, where $q_0 = -n_0$. Now, for each q_m in \mathbb{Q} , write the open $\{\omega \in 2^{\mathbb{N}} : f(\omega) > q_m\}$ as a disjoint union of clopens $\bigcup_{i=0}^{N_m} [\tau_{q_m, i}]$, where N_m is either a natural number or ∞ (depending on whether the open is itself clopen or not), and where the sequence $\tau_{q_m, i}$ is computable. Then, we have that $N_0 = 0$ and $\tau_{q_0, 0} = \varepsilon$, corresponding to the fact that $\mathcal{U}_{n_0} = 2^{\mathbb{N}}$. This has the effect that, for each $\sigma \in 2^{<\mathbb{N}}$ and each $k \geq 0$, the following set is always non-empty and, in particular, it contains q_0 :

$$F_k(\sigma) = \{q_m : m \leq k, q_m \geq -n_0 \text{ and } \exists i \leq k \text{ with } \sigma \sqsupseteq \tau_{q_m, i}\}.$$

Finally, we define $f_k(\sigma) = \max(F_k(\sigma))$. Clearly, the f_k 's are uniformly computable rational-valued functions and they satisfy Conditions (a) and (b) from Definition 1.2.12. For Condition (c), take $\omega \in 2^{\mathbb{N}}$ and suppose first that $f(\omega) < \sup\{f_k(\omega \upharpoonright n) : k, n \geq 0\}$. Then, there are some $k, n \geq 0$ with $f_k(\omega \upharpoonright n) > f(\omega)$. Moreover, there must be some $m \leq k$ such that $f_k(\omega \upharpoonright n) = q_m \geq -n_0$ and some $i \leq k$ such that $\omega \upharpoonright n \sqsupseteq \tau_{q_m, i}$. Thus, $\omega \in [\tau_{q_m, i}]$, which means that $f(\omega) > q_m$. But this is a contradiction. Similarly, if $f(\omega) > \sup\{f_k(\omega \upharpoonright n) : k, n \geq 0\}$, then there is some rational $q_m \geq -n_0$ in the enumeration such that $f(\omega) > q_m > \sup\{f_k(\omega \upharpoonright n) : k, n \geq 0\}$. So, $\omega \in \{\alpha \in 2^{\mathbb{N}} : f(\alpha) > q_m\} = \bigcup_{i \in \mathbb{N}} [\tau_{q_m, i}]$. Let i be such that $\omega \in [\tau_{q_m, i}]$ and $j = \max\{m, i\}$. Then, $f_j(\tau_{q_m, i}) \geq q_m$, which implies that $\sup\{f_k(\omega \upharpoonright n) : k, n \geq 0\} \geq q_m$, a contradiction. Hence, $f(\omega) = \sup\{f_k(\omega \upharpoonright n) : k, n \geq 0\}$. \square

We will now present the integral-test characterisations of Martin-Löf randomness, Schnorr randomness, weak 2-randomness, and weak 1-randomness. Computable randomness has also been given an integral-test characterisation (see Rute [2013, 2016]), but it will not be needed in what follows. In general, the characterisations of algorithmic randomness notions via integral tests require said tests to be non-negative functions. In Cantor space, however, the non-negativity constraint may be dropped. This is because, as evinced by Definition 1.2.12, lower semi-computable functions are bounded below; hence, adding a computable

¹³Recall that a topological space is compact if each of its open covers has a finite sub-cover.

constant suffices to obtain a non-negative lower semi-computable function from one that takes negative values. The reason why the restriction to Cantor space is important is that, in more general settings, lower semi-computable functions are standardly defined in terms of the effective lower semi-continuity condition given in Proposition 1.2.13. However, to prove Proposition 1.2.13, we relied on the compactness of Cantor space to argue that any lower semi-continuous function on a compact domain is bounded below.

Once again, let μ denote an arbitrary computable measure. We begin with Martin-Löf randomness—the first algorithmic randomness notion to be given a characterisation in terms of integral tests:

Theorem 1.2.14 (Levin [1976]). *Let $\omega \in 2^{\mathbb{N}}$. The following are equivalent:*

- (1) ω is μ -Martin-Löf random;
- (2) $f(\omega) < \infty$ for all lower semi-computable functions $f : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ with finite expectation: i.e., such that $\int_{2^{\mathbb{N}}} f d\mu < \infty$;
- (3) $f(\omega) < \infty$ for all lower semi-computable functions $f : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ such that $\int_{2^{\mathbb{N}}} f d\mu \leq 1$.

Any lower semi-computable function $f : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ with finite expectation relative to μ (or with expectation at most 1 relative to μ) will be referred to as an *integral test for μ -Martin-Löf randomness*.

Schnorr randomness, on the other hand, has the following integral test characterisation:

Theorem 1.2.15 (Miyabe [2013a]). *Let $\omega \in 2^{\mathbb{N}}$. The following are equivalent:*

- (1) ω is μ -Schnorr random;
- (2) $f(\omega) < \infty$ for all lower semi-computable functions $f : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ with computable expectation: i.e., such that $\int_{2^{\mathbb{N}}} f d\mu$ is a computable real;
- (3) $f(\omega) < \infty$ for all lower semi-computable functions $f : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ such that $\int_{2^{\mathbb{N}}} f d\mu = 1$.

Any lower semi-computable function $f : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ with computable expectation relative to μ (or with expectation exactly 1 relative to μ) will be referred to as an *integral test for μ -Schnorr randomness*.

The characterisation of weak 2-randomness via integral tests is instead as follows:

Theorem 1.2.16 (Miyabe [2013a]). *Let $\omega \in 2^{\mathbb{N}}$. The following are equivalent:*

- (1) ω is μ -weakly 2-random;
- (2) $f(\omega) < \infty$ for all lower semi-computable functions $f : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ that are finite μ -almost everywhere: i.e., such that $\mu(\{\alpha \in 2^{\mathbb{N}} : f(\alpha) < \infty\}) = 1$.

Any lower semi-computable function that is finite μ -almost everywhere will be referred to as an *integral test for μ -weak 2-randomness*.

Now, a function $f : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ is said to be *extended computable* (*computable*, for short) if and only if, for all $q, p \in \mathbb{Q}$ with $q < p$, $f^{-1}((q, p))$ and $f^{-1}((q, \infty])$ are Σ_1^0 classes, uniformly in q, p (see Miyabe [2013b]). Weak 1-randomness can then be characterised as follows via computable integral tests:

Theorem 1.2.17 (Miyabe [2013b]). *Let $\omega \in 2^{\mathbb{N}}$. The following are equivalent:*

- (1) ω is μ -weakly 1-random;
- (2) $f(\omega) < \infty$ for all computable $f : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ that are finite μ -almost everywhere.

Any computable function $f : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ that is finite μ -almost everywhere will be referred to as an *integral test for μ -weak 1-randomness*.

This concludes our review of the measure-theoretic typicality paradigm. Next, we will discuss the unpredictability paradigm.

1.3 The unpredictability paradigm

According to the unpredictability paradigm, the essence of a random sequence, or data stream, is that past observations do not provide any information that can be exploited to make better-than-chance predictions about future outcomes. Accordingly, a sequence is defined as algorithmically random if it is impossible for a gambler to predict the bits of that sequence and devise an effective betting strategy that would allow them to gain infinite wealth by successively wagering on said bits.

The betting strategies employed to define randomness are called *dyadic martingales*.¹⁴

¹⁴Here, in generalising the concept of a dyadic martingale from the uniform measure to arbitrary computable measures, we follow Rute [2016].

Definition 1.3.1 (Dyadic martingale). *Given a measure μ , a dyadic μ -martingale is a partial function $d : \subseteq 2^{<\mathbb{N}} \rightarrow \mathbb{R}^{\geq 0}$ such that, for all $\sigma \in 2^{<\mathbb{N}}$,*

- (a) *if $d(\sigma)$ is undefined, then $\mu([\sigma]) = 0$ (impossibility condition);*
- (b) *$d(\sigma)\mu([\sigma]) = d(\sigma 0)\mu([\sigma 0]) + d(\sigma 1)\mu([\sigma 1])$ (fairness condition), where a term of the form $d(\tau)\mu([\tau])$ is taken to be equal to 0 if $\mu([\tau]) = 0$ even when $d(\tau)$ is undefined.*

A dyadic μ -martingale d is said to be normed if $d(\varepsilon) = 1$. It is said to succeed on a sequence $\omega \in 2^{\mathbb{N}}$ if and only if $\limsup_{n \rightarrow \infty} d(\omega \upharpoonright n) = \infty$.

To be precise, dyadic martingales formalise the capital functions associated with betting strategies. For each $\sigma \in 2^{<\mathbb{N}}$ on which d is defined, $d(\sigma)$ represents the capital accumulated after betting on the first $n = |\sigma|$ bits of a sequence whose initial segment of length n is σ . The impossibility condition and the convention that $d(\tau)\mu([\tau]) = 0$ if $\mu([\tau]) = 0$ and $d(\tau)$ is undefined ensure that the fairness condition is well-defined. In turn, as its name suggests, the fairness condition ensures that the game is fair: it requires that the gambler's expected winnings always equal their current capital. A dyadic martingale is successful on a sequence if the underlying betting strategy wins an unbounded amount of wealth when played against that sequence.

Dyadic martingales are but a special case of the general notion of a (discrete-time) martingale from probability theory. First, recall that a *random variable* is simply a measurable function. In particular, a function $f : 2^{\mathbb{N}} \rightarrow \mathbb{R}$ is a random variable if $\{\omega \in 2^{\mathbb{N}} : f(\omega) \geq r\} \in \mathcal{B}(2^{\mathbb{N}})$ for all $r \in \mathbb{R}$. Moreover, a function f is said to be *integrable* if the expectation $\mathbb{E}_{\mu}[|f|]$ of its absolute value relative to the underlying measure μ is finite. A martingale is then defined as follows: it is a countably infinite sequence of random variables, all of them integrable and such that, for each n , the conditional expectation of the $(n + 1)$ -st random variable given the previous n random variables is equal to the value of the n -th random variable. In other words:

Definition 1.3.2 (Martingale). *Given a measure μ , a μ -martingale is a sequence $\{M_n\}_{n \in \mathbb{N}}$ of random variables such that, for each $n \in \mathbb{N}$,*

- (i) $\mathbb{E}_{\mu}[|M_n|] < \infty$, and

$$(ii) \mathbb{E}_\mu[M_{n+1}|M_1, \dots, M_n] = M_n.$$

It is easy to see that dyadic martingales satisfy these conditions. A function of the form $d : \subseteq 2^{<\mathbb{N}} \rightarrow \mathbb{R}^{\geq 0}$ can in fact be thought of as a sequence of random variables X_0^d, X_1^d, \dots with $X_n^d(\omega) = d(\omega \upharpoonright n)$ whenever $d(\omega \upharpoonright n)$ is defined. If d is a dyadic μ -martingale, the impossibility condition ensures that each random variable X_n^d is defined μ -almost everywhere; the fairness condition, on the other hand, guarantees that $\mathbb{E}_\mu[X_{|\sigma|+1}^d | [\sigma]] = d(\sigma)$ whenever $d(\sigma)$ is defined.

A dyadic μ -martingale is *computable* if it is a total computable function and *almost everywhere computable* if it is a partial computable function [Rute, 2016]. Similarly, a dyadic μ -martingale is *left-c.e.* if it is a total left-c.e. function and *almost everywhere left-c.e.* if it is a partial left-c.e. function. Now, a measure μ is *strictly positive* if $\mu([\sigma]) > 0$ for all $\sigma \in 2^{<\mathbb{N}}$. For strictly positive measures, computable and almost everywhere computable dyadic martingales coincide. Moreover, every almost everywhere computable dyadic μ -martingale d can be turned into a left-c.e. dyadic μ -martingale d' by letting $d'(\sigma)$ equal $d(\sigma)$ whenever $d(\sigma)$ is defined and setting it to 0 otherwise (i.e., when $\mu([\sigma]) = 0$). Similarly, every almost everywhere left-c.e. dyadic μ -martingale can be turned into a left-c.e. dyadic μ -martingale.

The definition of a dyadic martingale can be relaxed as follows: given a measure μ , a function $d : \subseteq 2^{<\mathbb{N}} \rightarrow \mathbb{R}^{\geq 0}$ will be said to be a *dyadic μ -supermartingale* if, for all $\sigma \in 2^{<\mathbb{N}}$, d satisfies the impossibility condition from Definition 1.3.1, and $d(\sigma)\mu([\sigma]) \geq d(\sigma 0)\mu([\sigma 0]) + d(\sigma 1)\mu([\sigma 1])$. Dyadic supermartingales differ from dyadic martingales in that they can be wasteful: they are “allowed to discard part of [the] capital, such as by buying drinks or tipping the dealer” [Downey and Hirschfeldt, 2010, p. 235]. The notion of success, as well as the notions of computable, almost everywhere computable, left-c.e., and almost everywhere left-c.e. dyadic supermartingale are defined analogously to the case of dyadic martingales.

We are now ready to characterise randomness in terms of dyadic martingales. We will rehearse the characterisations of Martin-Löf randomness and computable randomness in this setting, and then we will define an additional randomness notion in terms of martingales: density randomness. All of these notions will play an important role in subsequent chapters. Schnorr randomness and weak 1-randomness have been given martingale-based

characterisations, as well (cf. [Schnorr, 1971a,b] and [Kurtz, 1981], respectively, or Downey and Hirschfeldt [2010]), but we will not make use of them in this dissertation.

We begin with Martin-Löf randomness, which can be characterised in terms of dyadic (super)martingales as follows:

Theorem 1.3.3 (Schnorr [1971a]). *Let $\omega \in 2^{\mathbb{N}}$. The following are equivalent:*

- (1) ω is μ -Martin-Löf random;
- (2) $\mu([\omega \upharpoonright n]) > 0$ for all $n \in \mathbb{N}$ and, for all (almost everywhere) left-c.e. dyadic μ -martingales d , $\limsup_{n \rightarrow \infty} d(\omega \upharpoonright n) < \infty$.
- (3) $\mu([\omega \upharpoonright n]) > 0$ for all $n \in \mathbb{N}$ and, for all (almost everywhere) left-c.e. dyadic μ -supermartingales d , $\limsup_{n \rightarrow \infty} d(\omega \upharpoonright n) < \infty$.

Computable randomness was first characterised in terms of (total) computable dyadic martingales by Schnorr [1971a,b] in the context of the uniform measure. For computable measures in general, the following holds:

Theorem 1.3.4 (Rute [2016]). *Let $\omega \in 2^{\mathbb{N}}$. The following are equivalent:*

- (1) ω is μ -computably random;
- (2) $\mu([\omega \upharpoonright n]) > 0$ for all $n \in \mathbb{N}$ and, for all almost everywhere computable dyadic μ -martingales d , $\limsup_{n \rightarrow \infty} d(\omega \upharpoonright n) < \infty$;
- (3) $\mu([\omega \upharpoonright n]) > 0$ for all $n \in \mathbb{N}$ and, for all almost everywhere computable dyadic μ -martingales d , $\lim_{n \rightarrow \infty} d(\omega \upharpoonright n)$ exists and is finite.

Lastly, we consider the following randomness notion, which results from a natural modification of the characterisation of μ -Martin-Löf randomness in terms of (almost everywhere) left-c.e. dyadic μ -martingales/ μ -supermartingales:

Definition 1.3.5 (Density randomness). *A sequence $\omega \in 2^{\mathbb{N}}$ is μ -density random if and only if $\mu([\omega \upharpoonright n]) > 0$ for all $n \in \mathbb{N}$ and, for all (almost everywhere) left-c.e. dyadic μ -martingales d , $\lim_{n \rightarrow \infty} d(\omega \upharpoonright n)$ exists and is finite. Equivalently, ω is μ -density random if and only if $\mu([\omega \upharpoonright n]) > 0$ for all $n \in \mathbb{N}$ and, for all (almost everywhere) left-c.e. dyadic μ -supermartingales d , $\lim_{n \rightarrow \infty} d(\omega \upharpoonright n)$ exists and is finite.*

The collection of μ -density random sequences will be denoted by μ -DR. Clearly, if a sequence is μ -density random, then it is μ -Martin-Löf random, for all μ . However, as we shall see, the converse does not hold.

Density randomness owes its name to the following fact. In [2016], in line with a current trend in algorithmic randomness whereby effective versions of classical almost-everywhere theorems in analysis are used to define new randomness notions, Miyabe et al.¹⁵ define a randomness concept which relies on an effectivisation of the Lebesgue Density Theorem. This randomness notion, which Miyabe et al. call density randomness, is defined as follows. Let $\mathcal{C} \in \mathcal{B}(2^{\mathbb{N}})$ and $\omega \in 2^{\mathbb{N}}$. Then, the *lower density* of ω in \mathcal{C} is given by

$$\underline{\rho}(\mathcal{C} | \omega) = \liminf_{n \rightarrow \infty} \frac{\lambda(\mathcal{C} \cap [\omega \upharpoonright n])}{\lambda([\omega \upharpoonright n])}.$$

A sequence $\omega \in 2^{\mathbb{N}}$ is a λ -dyadic positive density point if $\underline{\rho}(\mathcal{C} | \omega) > 0$ for all Π_1^0 classes \mathcal{C} such that $\omega \in \mathcal{C}$; it is a λ -dyadic density-one point if $\underline{\rho}(\mathcal{C} | \omega) = 1$ for all Π_1^0 classes \mathcal{C} such that $\omega \in \mathcal{C}$. Then, a sequence $\omega \in 2^{\mathbb{N}}$ is said to be λ -density random if and only if it is both λ -Martin-Löf random and a λ -dyadic density-one point.

This definition ensures that λ -density randomness entails λ -Martin-Löf randomness. As shown by Bienvenu et al. [2012b], if a sequence ω is λ -Martin-Löf random, then, for all Π_1^0 classes \mathcal{C} such that $\omega \in \mathcal{C}$, the *upper density* of ω in \mathcal{C} is one: i.e.,

$$\bar{\rho}(\mathcal{C} | \omega) = \limsup_{n \rightarrow \infty} \frac{\lambda(\mathcal{C} \cap [\omega \upharpoonright n])}{\lambda([\omega \upharpoonright n])} = 1.$$

However, a λ -Martin-Löf random sequence need not be a λ -dyadic density-one point. The above definition also entails that λ -density randomness is weaker than λ -weak 2-randomness (for all Π_1^0 classes \mathcal{C} and rationals $q \in [0, 1]$, the set $\{\omega \in \mathcal{C} : \underline{\rho}(\mathcal{C} | \omega) < q\}$ is a Π_2^0 class which, by the Lebesgue Density Theorem, has λ -measure zero).

Crucially, in the context of the uniform measure, this notion of density randomness has an equivalent characterisation in terms of dyadic (super)martingales. This result was proven by Andrews, Cai, Diamondstone, Lempp, and Miller, all members the UW-Madison Logic Group, and the proof was fully spelled out in print by Miyabe et al. [2016]:

¹⁵See also [Bienvenu et al., 2014].

Theorem 1.3.6 (Madison group). *Let $\omega \in 2^{\mathbb{N}}$. The following are equivalent:*

- (1) *ω is λ -density random (i.e., ω is λ -Martin-Löf random, as well as a λ -dyadic density-one point);*
- (2) *for all left-c.e. dyadic λ -martingales d , $\lim_{n \rightarrow \infty} d(\omega \upharpoonright n)$ exists and is finite.*
- (3) *for all left-c.e. dyadic λ -supermartingales d , $\lim_{n \rightarrow \infty} d(\omega \upharpoonright n)$ exists and is finite.*

Definition 1.3.5 is the obvious generalisation of Condition (2) and Condition (3) from Theorem 1.3.6 to arbitrary computable measures. Unless otherwise specified, when talking about density randomness we will always refer to the notion from Definition 1.3.5.

With this note, we end our discussion of the unpredictability paradigm, as well as our excursion into the fundamentals of the theory of algorithmic randomness. Next, we will see how this theory may be employed to better understand inductive learning and the learning performance of computationally limited learners. We will begin by considering the applications of algorithmic randomness in Bayesian learning and, in particular, in the study of Bayesian merging-of-opinions theorems.

Chapter 2

Algorithmic randomness and merging of opinions

Regularities are where you find them, and you can find them anywhere.

Goodman, *Fact, Fiction and Forecast*

[R]andomness... is going to be a concept which is relative to our body of knowledge, which will somehow reflect what we know and what we don't know.

Kyburg, *The Logical Foundations of Statistical Inference*

Bayesian learning encompasses a family of methods of statistical inference that crucially rely on prior probability distributions (priors), which are meant to encapsulate the experimenter's background knowledge and inductive assumptions before performing an experiment. This reliance on priors is often taken to be a cause for concern: when the available background knowledge does not suffice to reach inter-subjective agreement on prior probabilities, how can one be possibly guaranteed that the inferences drawn on the basis of one's own subjective prior provide any objective epistemic warrant? Perhaps most alarmingly, does the use of Bayesian methods—and, thus, of prior probability distributions—in the sciences threaten the objectivity of scientific knowledge?

According to *objective Bayesians*, such as Jaynes [1968] and Rosenkrantz [1981], this problem can be overcome by singling out the class of rationally permissible priors, the adoption of which ensures the objectivity of the conclusions derived from them. For instance, some objective Bayesians might contend that symmetry considerations play a crucial role in fixing the collection of rationally permissible priors, or that priors should be calibrated with known frequencies. On the other hand, *subjective Bayesians* such as Ramsey [1931], de Finetti [1937], Savage [1954], and Jeffrey [1977] maintain that the only requirement that rationality imposes on prior probability distributions is probabilistic coherence, and that there is no principled way of arguing for the superiority of a certain prior over another.

To rebuke accusations of excessive subjectivity, subjective Bayesians then often appeal to various asymptotic results from probability and measure theory that are meant to show that a Bayesian agent's initial beliefs or assumptions, in the form of a prior probability distribution, are essentially immaterial for the purpose of successful inquiry: the dynamics of Bayesian conditioning by themselves ensure that priors are eventually washed out by the shared evidence. Suppes, for instance, expresses this sentiment as follows:

It is of fundamental importance to any deep appreciation of the Bayesian viewpoint to realize the particular form of the prior distribution expressing beliefs held before the experiment is conducted is not a crucial matter. [...] The well-designed experiment is one that will swamp divergent prior distributions with the clarity and sharpness of its results, and thereby render insignificant the diversity of prior opinion. [Suppes, 1966, p. 204]

Similarly, Edwards, Lindman, and Savage argue that

Although your initial opinion about future behavior [...] may differ radically from your neighbor's, your opinions and his will ordinarily be so transformed by application of Bayes' theorem [...] as to become nearly indistinguishable. This approximate merging of initially divergent opinions is, we think, one reason why empirical research is called "objective". [Edwards et al., 1963, p. 197]

The theorems employed to argue that initial diversity of opinions is immaterial have

their roots in Savage’s work [1954], and they fall under the umbrella of “Bayesian merging-of-opinions” results. Roughly put, these results establish that, provided that their respective subjective priors are sufficiently compatible, two Bayesian agents¹ beginning the learning process with divergent beliefs are guaranteed to almost surely reach a consensus: as the number of observations increases, their beliefs (in the form of their posterior probability distributions) will become arbitrarily close with probability one.

In this chapter, we explore the phenomenon of merging of opinions in the context of more realistic, less-than-ideal agents. We do so by bringing into play the theory of computation: that is, by focusing on computationally limited Bayesian agents whose subjective priors are *computable* probability measures.² The key insight behind this work is that merging of opinions for computationally limited Bayesian agents—computable Bayesian agents, for short—can then be studied through the prism of algorithmic randomness. In particular, as we shall see, the algorithmic randomness concepts introduced in Chapter 1 can be employed to define refined notions of doxastic compatibility: given an algorithmic randomness notion R , the beliefs of two computable Bayesian agents will be said to be compatible relative to R if their respective computable priors agree on which data streams are R -random. More precisely, given two priors μ and ν , ν will be said to be compatible with μ relative to R if the collection of μ - R -random data streams is a subset of the collection of ν - R -random data streams.

The rationale for modelling doxastic compatibility in terms of algorithmic randomness is that the algorithmically random data streams, though maximally irregular when considered locally, bit by bit, are of necessity globally regular. As seen in Chapter 1, in spite of being unpredictable (observing a finite initial segment of a random data stream does not provide any useful information for predicting what the next observation is going to be), the algorithmically random sequences must nonetheless satisfy various effectively specifiable statistical laws (such as the Strong Law of Large Numbers and the Law of the Iterated Logarithm when the underlying measure is i.i.d.). From this perspective, different notions of algorithmic randomness may be seen as encoding different beliefs about the global uniformity of nature: each algorithmic randomness concept corresponds, from

¹As shown by Schervish and Seidenfeld [1990], these results are generalisable to the case where there are more than two Bayesian agents, provided that certain technical conditions are met.

²See Definition 1.1.1.

the viewpoint of the computable Bayesian agent with respect to whom that randomness notion is defined, to a precise class of effectively specifiable global regularities. So, when two computable Bayesian agents agree on what data streams are algorithmically random, they can be seen as making compatible inductive assumptions: they have compatible beliefs (or commitments) about which statistical properties the observational data is going to satisfy. In other words, they concur on the extent of nature's global uniformity, where the type of uniformity in question amounts to the satisfaction of certain effectively specifiable statistical laws.

To corroborate the legitimacy of using algorithmic randomness to define notions of doxastic compatibility, we will show that agreeing on which data streams are algorithmically random provably leads to asymptotic merging of opinions. In other words, the main results of this chapter establish that, when shared by computable Bayesian agents with differing subjective priors, the inductive assumptions pertaining to the global uniformity of nature encoded by algorithmic randomness notions guarantee the eventual (almost sure) attainment of inter-subjective agreement.

As explained in Chapter 1, the theory of algorithmic randomness is deeply rooted in measure theory, and the question of what equivalence relations are induced by randomness notions in the context of computable measures has already received some attention in the algorithmic randomness literature (see, in particular, [Bienvenu and Merkle, 2009]). We will review and make use of some of these results in what follows. Our most important contribution in this chapter consists in connecting algorithmic randomness to the study of notions of compatibility between subjective priors and in showing that the resulting notions of compatibility imply merging. This work therefore bridges the theory of algorithmic randomness and the formal epistemology (and statistics) literature on merging of opinions, Bayesian learning, and their philosophical ramifications.

The structure of this chapter is as follows. In §2.1.1, we review some canonical notions of agreement and disagreement between priors. In §2.1.2, we discuss the phenomenon of merging of opinions in the classical setting and, in particular, what is arguably the most prominent merging-of-opinions result: the Blackwell-Dubins Theorem [Blackwell and Dubins, 1962]. We then also consider the opposite phenomenon: polarisation of opinions. Our main results are in §2.2. In §2.2.1, we explore the relations between standard notions of

compatibility/incompatibility and the notions of agreement induced by algorithmic randomness. Then, we show that all core algorithmic randomness notions except for weak 1-randomness generate notions of compatibility that lead to merging. In §2.2.2, we conclude with some simple observations about how disagreement over which data streams are algorithmically random leads to asymptotic polarisation of opinions.

2.1 Merging of opinions

Recall that a *probability space* $(\Omega, \mathcal{E}, \mu)$ is a triple consisting of a set Ω , a σ -algebra \mathcal{E} on Ω , and a probability measure μ on \mathcal{E} (again, since the only measures that we will be dealing with in what follows are probability measures, we will simply call them measures from now on). The set Ω is the sample space: the collection of all possible basic outcomes of the experiment being modelled—or, more generally, the collection of all possible observational data associated with the inductive problem under consideration. The σ -algebra \mathcal{E} , on the other hand, corresponds to the collection of all events (involving observational data) that get assigned a probability: intuitively, all of the events that a Bayesian learner can entertain in the given situation. In this setting, hypotheses are thus elements of the algebra \mathcal{E} .

We are interested in situations where the same space comes equipped with two measures μ and ν . These measures may be given various interpretations. So far, we have been taking probability distributions to represent subjective priors; yet, measures admit a non-personalist interpretation, as well: they can be taken to encode objective distributions. For instance, a measure may be seen as representing the true chance distribution governing some stochastic process (e.g., a game of chance or Brownian motion). Merging-of-opinions theorems apply in this context, too. When one of the measures involved is an objective chance distribution while the other one is the subjective prior of a Bayesian agent, these results establish that, with increasing information, the agent’s beliefs will asymptotically align with the true chances, provided that said beliefs are sufficiently compatible with the truth. So, in what follows, while μ and ν will generally be assumed to be the subjective priors of two Bayesian agents, when appropriate, μ may also be viewed as representing the true distribution governing some process, and ν as the subjective prior of a Bayesian agent trying to approximate the true distribution.

Bodies of evidence can be naturally modelled in terms of σ -algebras. In particular, increasing bodies of evidence can be represented as *filtrations* on (Ω, \mathcal{E}) : namely, as sequences $\{\mathcal{E}_n\}_{n \in \mathbb{N}}$ of sub- σ -algebras³ of \mathcal{E} such that $\mathcal{E}_n \subseteq \mathcal{E}_{n+1}$ for all $n \in \mathbb{N}$ —where the latter condition ensures that, for each n , the information embodied by \mathcal{E}_{n+1} refines the information embodied by \mathcal{E}_n . Given a filtration $\{\mathcal{E}_n\}_{n \in \mathbb{N}}$ on (Ω, \mathcal{E}) , let \mathcal{E}^∞ denote the Borel σ -algebra $\sigma(\bigcup_{n \in \mathbb{N}} \mathcal{E}_n)$ generated by the union of the \mathcal{E}_n 's. If $\mathcal{E}^\infty = \mathcal{E}$, then we say that the filtration is *complete*: the cumulating evidence will eventually settle every event that a Bayesian agent can entertain.

Learning occurs by conditionalising on the (total) available evidence. Since, in this setting, the growing evidence is encapsulated by a filtration, we need to define the notion of conditional probability given a sub- σ -algebra. Fix a filtration $\{\mathcal{E}_n\}_{n \in \mathbb{N}}$ on (Ω, \mathcal{E}) , an event $A \in \mathcal{E}$, and a prior μ . The conditional probability $\mu(A \mid \mathcal{E}_n)$ of A given \mathcal{E}_n is an \mathcal{E}_n -measurable function⁴ (a random variable) $\mu(A \mid \mathcal{E}_n) : \Omega \rightarrow \mathbb{R}$ such that, for all $B \in \mathcal{E}_n$, $\mu(A \cap B) = \int_B \mu(A \mid \mathcal{E}_n) d\mu$. Such a function exists for any sub- σ -algebra⁵ and is unique up to sets of μ -measure zero.⁶

As explained in Chapter 1, in what follows we shall focus on the space $(2^{\mathbb{N}}, \mathcal{B}(2^{\mathbb{N}}))$ —where $2^{\mathbb{N}}$ is the set of countably infinite binary sequences and $\mathcal{B}(2^{\mathbb{N}})$ the Borel σ -algebra on $2^{\mathbb{N}}$ —and on measures μ, ν on $\mathcal{B}(2^{\mathbb{N}})$. We will think of sequences in $2^{\mathbb{N}}$ as data streams. In addition, we will restrict attention to the filtration $\{\mathcal{F}_n\}_{n \in \mathbb{N}}$, where, for each n , \mathcal{F}_n is the sub- σ -algebra of $\mathcal{B}(2^{\mathbb{N}})$ generated by the cylinders $[\sigma]$ associated to strings $\sigma \in 2^{<n}$ of length n . Each such algebra is thus generated by a finite partition of $2^{\mathbb{N}}$ (for instance, \mathcal{F}_1 is the algebra $\{\emptyset, [0], [1], 2^{\mathbb{N}}\}$ generated by the finite partition $\{[0], [1]\}$). Intuitively, \mathcal{F}_n represents all possible evidential situations that a Bayesian agent could find themselves in at the n -th stage of the learning process, after having made n observations (after having

³Given two σ -algebras \mathcal{F} and \mathcal{E} on the same set, \mathcal{F} is a sub- σ -algebra of \mathcal{E} if $\mathcal{F} \subseteq \mathcal{E}$.

⁴Recall that, given (Ω, \mathcal{E}) , a function $f : \Omega \rightarrow \mathbb{R}$ is \mathcal{E} -measurable if, for all $r \in \mathbb{R}$, $\{\omega \in \Omega : f(\omega) \geq r\} \in \mathcal{E}$.

⁵This is a consequence of the Radon-Nikodym Theorem (see, for instance, [Durrett, 2010, Theorem A.4.8., p. 470]).

⁶The standard definition of conditional probability given by Bayes' formula only applies to cases where the conditioning event has positive probability. Defining conditional probabilities in this more general setting allows to define conditionalisation with respect to probability zero events, as well. This definition also ensures that $\mu(A \mid \mathcal{E}_n)$ is a version of the conditional expectation $\mathbb{E}[\chi_A \mid \mathcal{E}_n]$ of the indicator function χ_A of A .

observed the first n digits of the true data stream). Since $\sigma(\bigcup_{n \in \mathbb{N}} \mathcal{F}_n) = \mathcal{B}(2^{\mathbb{N}})$, this filtration is complete. We are thus assuming that the evidence is both increasing and complete. Given that the \mathcal{F}_n 's are generated by finite partitions, learning in this setting essentially proceeds by standard Bayesian conditioning. We can in fact almost surely recover the familiar definition of conditional probabilities as follows: for any $\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})$, $n \in \mathbb{N}$, and μ -almost every $\omega \in 2^{\mathbb{N}}$,⁷

$$\mu(\mathcal{S} \mid [\omega \upharpoonright n]) = \frac{\mu(\mathcal{S} \cap [\omega \upharpoonright n])}{\mu([\omega \upharpoonright n])} = \frac{1}{\mu([\omega \upharpoonright n])} \int_{[\omega \upharpoonright n]} \mu(\mathcal{S} \mid \mathcal{F}_n) d\mu = \mu(\mathcal{S} \mid \mathcal{F}_n)(\omega),$$

where the second identity follows from the definition of $\mu(\mathcal{S} \mid \mathcal{F}_n)$, and the last identity from the fact that the value of $\mu(\mathcal{S} \mid \mathcal{F}_n)$ is constant within the partition cells generating \mathcal{F}_n —and so, in particular, within the cylinder $[\omega \upharpoonright n]$.

2.1.1 Classical notions of compatibility and incompatibility

We begin by reviewing some classical notions of compatibility and incompatibility between measures, which will serve as a springboard for our study of notions of agreement induced by algorithmic randomness.

Arguably, the most well-studied notion of compatibility between measures is *absolute continuity*, which is defined below.

Definition 2.1.1 (Absolute continuity). *Given measures μ and ν , μ is said to be absolutely continuous with respect to ν (in symbols, $\mu \ll \nu$) if, for any event $\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})$, $\mu(\mathcal{S}) > 0$ implies that $\nu(\mathcal{S}) > 0$. If both $\mu \ll \nu$ and $\nu \ll \mu$, then μ and ν are said to be mutually absolutely continuous.*

If μ and ν encode the subjective priors of two Bayesian agents, then μ being absolutely continuous with respect to ν intuitively means that μ is at least as dogmatic a prior as ν is. All of the events that are *a priori* “excluded” by the agent with prior ν (by virtue of having been assigned probability zero before any observations are made) are also “excluded” by the agent with prior μ .⁸ In other words, the agent with prior ν cannot be surprised by any

⁷In fact, this holds for every $\omega \in 2^{\mathbb{N}}$ such that $\mu([\omega \upharpoonright n]) > 0$ for all $n \in \mathbb{N}$.

⁸We write “excluded” within scare quotes because assigning probability zero to an event is not the same as considering that event impossible (similarly, assigning probability one to an event is not the same as

event to which the agent with prior μ assigns positive probability. It is however possible for μ to be strictly more dogmatic than ν : i.e., the agent with prior μ may assign probability zero to some events to which the agent with prior ν assigns positive probability. On the other hand, if μ represents the true distribution governing some stochastic process while ν is the subjective prior of a Bayesian agent, then μ being absolutely continuous with respect to ν means that the agent with prior ν assigns probability zero only to events that truly have probability zero.

Here are two examples to elucidate this type of compatibility.

Example 2.1.2. *If ν is a convex combination of μ_1 and μ_2 , then $\mu_1 \ll \nu$ and $\mu_2 \ll \nu$. Take the uniform measure λ and let $\nu = \frac{1}{4}\lambda + \frac{3}{4}\mu_{\frac{1}{3}}$, where $\mu_{\frac{1}{3}}$ is the Bernoulli measure given by $\mu_{\frac{1}{3}}([\sigma]) = \frac{1}{3}^k \cdot \frac{2}{3}^{n-k}$, with n the length of σ , k the number of 0's occurring in σ and $(n-k)$ the number of 1's occurring in σ . Measure ν is a convex combination of λ and $\mu_{\frac{1}{3}}$. Now, let $\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})$ such that $\lambda(\mathcal{S}) > 0$. Then, $\frac{1}{4}\lambda(\mathcal{S}) > 0$ and $\frac{3}{4}\mu_{\frac{1}{3}}(\mathcal{S}) \geq 0$, which together imply that $\nu(\mathcal{S}) > 0$. Hence, $\lambda \ll \nu$. An analogous argument shows that $\mu_{\frac{1}{3}} \ll \nu$.*

Example 2.1.3. *Given a measure ν , let $f : 2^{\mathbb{N}} \rightarrow \mathbb{R}$ be a unit-integrable random variable relative to ν : that is, let f be a measurable function with $\int_{2^{\mathbb{N}}} |f| d\nu = 1$. Then, define the measure μ as follows. For all $\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})$, let $\mu(\mathcal{S}) = \int_{\mathcal{S}} |f| d\nu$. Clearly, μ is a probability measure: since f is unit-integrable, $\mu(2^{\mathbb{N}}) = 1$ and $\mu(\emptyset) = 0$; moreover, for all countable collections $\{\mathcal{A}_i\}_{i \in \mathbb{N}}$ of pairwise disjoint sets in $\mathcal{B}(2^{\mathbb{N}})$, $\mu(\bigcup_{i \in \mathbb{N}} \mathcal{A}_i) = \int_{\bigcup_{i \in \mathbb{N}} \mathcal{A}_i} |f| d\nu = \sum_{i \in \mathbb{N}} \int_{\mathcal{A}_i} |f| d\nu = \sum_{i \in \mathbb{N}} \mu(\mathcal{A}_i)$. For any $\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})$ with $\nu(\mathcal{S}) = 0$, we have that $\mu(\mathcal{S}) = \int_{\mathcal{S}} |f| d\nu = 0$. So, $\mu \ll \nu$.*

We now consider a canonical notion of incompatibility between measures, as well as its dual notion, which yields a very minimal form of compatibility.

Definition 2.1.4 (Orthogonality). *Two measures μ and ν are said to be orthogonal (in symbols, $\mu \perp \nu$) if there is an event $\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})$ such that $\mu(\mathcal{S}) = 1$ but $\nu(\mathcal{S}) = 0$. If there is no such event, then μ and ν are non-orthogonal (in symbols, $\mu \not\perp \nu$).*

taking it to be certain). For a simple illustration of this difference, take the uniform measure λ on $\mathcal{B}(2^{\mathbb{N}})$. For every data stream $\omega \in 2^{\mathbb{N}}$, $\lambda(\{\omega\}) = 0$; however, one data stream has to correspond to the true sequence of outcomes of the experiment under consideration. So, assigning probability zero to an event amounts to treating it as extremely unlikely or negligible, but not as literally impossible.

Orthogonality is diametrically opposed to absolute continuity. If μ and ν are orthogonal, then absolute continuity fails in the most extreme way possible, as the event with (without loss of generality) ν -measure zero and positive μ -measure witnessing the failure of absolute continuity has in fact μ -measure one. So, orthogonality captures a radical type of disagreement. As a result, non-orthogonality embodies a very weak type of agreement.

Below is an example of two orthogonal measures, followed by an example of two non-orthogonal measures.

Example 2.1.5. *Take the uniform measure λ and the Bernoulli measure $\mu_{\frac{1}{3}}$ from Example 2.1.2. Given a sequence $\omega \in 2^{\mathbb{N}}$, let $\frac{\#0(\omega \upharpoonright n)}{n}$ denote the relative frequency of 0's in the first n digits of ω . By the Strong Law of Large Numbers, $\lambda\left(\left\{\omega \in 2^{\mathbb{N}} : \lim_{n \rightarrow \infty} \frac{\#0(\omega \upharpoonright n)}{n} = \frac{1}{2}\right\}\right) = 1$ and $\mu_{\frac{1}{3}}\left(\left\{\omega \in 2^{\mathbb{N}} : \lim_{n \rightarrow \infty} \frac{\#0(\omega \upharpoonright n)}{n} = \frac{1}{3}\right\}\right) = 1$. As a consequence, $\mu_{\frac{1}{3}}\left(\left\{\omega \in 2^{\mathbb{N}} : \lim_{n \rightarrow \infty} \frac{\#0(\omega \upharpoonright n)}{n} = \frac{1}{2}\right\}\right) = 0$, which shows that λ and $\mu_{\frac{1}{3}}$ are orthogonal.*

Example 2.1.6. *Let $\mu_{\frac{1}{3}}$ be the Bernoulli measure from the previous example, λ the uniform measure, and $\nu = \frac{1}{2}\mu_{\frac{1}{3}} + \frac{1}{2}\lambda$. Then, $\mu_{\frac{1}{3}} \ll \nu$ and $\lambda \ll \nu$. Let $\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})$. If $\mu_{\frac{1}{3}}(\mathcal{S}) = 0$, then $\frac{1}{2}\mu_{\frac{1}{3}}(\mathcal{S}) = 0$. Since $\frac{1}{2}\lambda(\mathcal{S}) \leq \frac{1}{2}$, we then have that $\nu(\mathcal{S}) \leq \frac{1}{2}$, too. On the other hand, if $\nu(\mathcal{S}) = 0$, then $\mu_{\frac{1}{3}}(\mathcal{S}) = 0$, as $\mu_{\frac{1}{3}} \ll \nu$. Hence, $\mu_{\frac{1}{3}}$ and ν are non-orthogonal. A similar argument establishes that λ and ν are non-orthogonal.*

If $\mu \ll \nu$, then μ and ν are non-orthogonal. The converse, however, does not hold: non-orthogonality is a much weaker form of compatibility than absolute continuity. The measures from Example 2.1.6 provide a counterexample: even though, as noted above, $\mu_{\frac{1}{3}} \ll \nu$, it is not the case that $\nu \ll \mu_{\frac{1}{3}}$. Since $\mu_{\frac{1}{3}}\left(\left\{\omega \in 2^{\mathbb{N}} : \lim_{n \rightarrow \infty} \frac{\#0(\omega \upharpoonright n)}{n} = \frac{1}{2}\right\}\right) = 0$ and $\lambda\left(\left\{\omega \in 2^{\mathbb{N}} : \lim_{n \rightarrow \infty} \frac{\#0(\omega \upharpoonright n)}{n} = \frac{1}{2}\right\}\right) = 1$, $\nu\left(\left\{\omega \in 2^{\mathbb{N}} : \lim_{n \rightarrow \infty} \frac{\#0(\omega \upharpoonright n)}{n} = \frac{1}{2}\right\}\right) = \frac{1}{2}$, which shows that $\nu \not\ll \mu_{\frac{1}{3}}$. Similarly, even though $\lambda \ll \nu$, $\nu \not\ll \lambda$. For an example of two measures that are non-orthogonal, and such that neither of them is absolutely continuous with respect to the other, consider the following. Let μ be such that, for all $\sigma \in 2^{<\mathbb{N}}$, $\mu([0\sigma]) = 2^{-(|\sigma|+1)}$, $\mu([10]) = 0$, and $\mu([11\sigma]) = 2^{-(|\sigma|+1)}$. In addition, let ν be such that, for all $\sigma \in 2^{<\mathbb{N}}$, $\nu([1\sigma]) = 2^{-(|\sigma|+1)}$, $\nu([01]) = 0$, and $\nu([00\sigma]) = 2^{-(|\sigma|+1)}$. Now, for any $\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})$ with $\mu(\mathcal{S}) = 1$, $\mu(\mathcal{S} \cap [11]) > 0$, since $\mu([11]) > 0$. But then $\nu(\mathcal{S} \cap [11]) = \nu(\mathcal{S} \mid [11])\nu([11]) = \mu(\mathcal{S} \mid [11])\frac{1}{2}\mu([11]) = \frac{1}{2}\mu(\mathcal{S} \cap [11]) > 0$, which, in turn,

implies that $\nu(\mathcal{S}) > 0$. On the other hand, for any $\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})$ with $\nu(\mathcal{S}) = 1$, $\nu(\mathcal{S} \cap [00]) > 0$. Hence, $\mu(\mathcal{S} \cap [00]) = \frac{1}{2}\nu(\mathcal{S} \cap [00]) > 0$ and, so, $\mu(\mathcal{S}) > 0$. Therefore, μ and ν are non-orthogonal. However, neither $\mu \ll \nu$ nor $\nu \ll \mu$, since $\nu([01]) = 0$ while $\mu([01]) = \frac{1}{4}$, and $\mu([10]) = 0$ while $\nu([10]) = \frac{1}{4}$.

The fact that absolute continuity and orthogonality are diametrically opposed notions is laid bare by the Lebesgue Decomposition Theorem, a classical result in measure theory.⁹ Given any two probability measures μ and ν , the Lebesgue Decomposition Theorem in fact implies that there is some $\alpha \in [0, 1]$ such that μ can be decomposed into two probability measures as follows: $\mu = \alpha\mu_a + (1 - \alpha)\mu_o$, where $\mu_a \ll \nu$ and $\mu_o \perp \nu$. Intuitively, μ_a is the part of μ that is compatible with ν , while μ_o is the part of μ that is incompatible with it. If $\alpha = 1$, then $\mu \ll \nu$, and if $\alpha = 0$, then $\mu \perp \nu$; if, on the other hand, $\alpha \in (0, 1)$, then the decomposition of μ into μ_a and μ_o can be shown to be unique. As we shall see, this result plays an important role in the context of merging-of-opinions theorems and their philosophical implications.

We conclude our review of compatibility notions by discussing a weaker form of absolute continuity: *local absolute continuity*.

Definition 2.1.7 (Local absolute continuity). *Given measures μ and ν , μ is said to be locally absolutely continuous with respect to ν (in symbols, $\mu \ll_{loc} \nu$) if, for every $n \in \mathbb{N}$ and every $\mathcal{S} \in \mathcal{F}_n$, $\mu(\mathcal{S}) > 0$ implies that $\nu(\mathcal{S}) > 0$.¹⁰ If both $\mu \ll_{loc} \nu$ and $\nu \ll_{loc} \mu$, then μ and ν are said to be mutually locally absolutely continuous.*

Given that the filtration $\{\mathcal{F}_n\}_{n \in \mathbb{N}}$ represents the possible evidence that the agents may observe, $\mu \ll_{loc} \nu$ means that ν agrees with μ about which evidence they expect to see. In other words, ν cannot be surprised by any piece of evidence to which μ assigns positive probability. Another way to think about local absolute continuity is that it amounts to absolute continuity restricted to finite-horizon events—that is, events that can be settled by a finite amount of evidence.

Recall that a measure μ is strictly positive if it assigns positive probability to every basic open set: i.e., if $\mu([\sigma]) > 0$ for all $\sigma \in 2^{<\mathbb{N}}$. Intuitively, strictly positive measures embody

⁹See, for instance, [Durrett, 2010, Theorem A.4.7, p. 469].

¹⁰Given the definition of the filtration $\{\mathcal{F}_n\}_{n \in \mathbb{N}}$, in our setting this condition is equivalent to the following: for any $\sigma \in 2^{<\mathbb{N}}$, $\mu([\sigma]) > 0$ implies that $\nu([\sigma]) > 0$.

a certain type of open-mindedness: they do not *a priori* rule out any finite sequence of observations. Clearly, any two strictly positive measures are mutually locally absolutely continuous. For a concrete example, take once again the uniform measure λ and the Bernoulli measure $\mu_{\frac{1}{3}}$, which are both strictly positive. While absolute continuity entails local absolute continuity, the reverse implication does not hold: as shown in Example 2.1.5, λ and $\mu_{\frac{1}{3}}$ are orthogonal and, so, neither of them is absolutely continuous with respect to the other.

The above also establishes that local absolute continuity does not imply non-orthogonality: Example 2.1.5 reveals that two measures can be locally absolutely continuous, and yet there can be an infinite-horizon event (i.e., an event that can only be settled by an infinite amount of evidence) on which these two measures maximally disagree. As a matter of fact, non-orthogonality and local absolute continuity are independent notions: neither of them implies the other. To see that non-orthogonality indeed fails to imply local absolute continuity, note that the example given earlier (after Example 2.1.6) of two non-orthogonal measures μ and ν such that $\mu \not\ll \nu$ and $\nu \not\ll \mu$ is also a case where $\mu \ll_{loc} \nu$ and $\nu \ll_{loc} \mu$.

The notions of compatibility that we discussed above and the logical relationships between them are summarised in Figure 2.1 below.

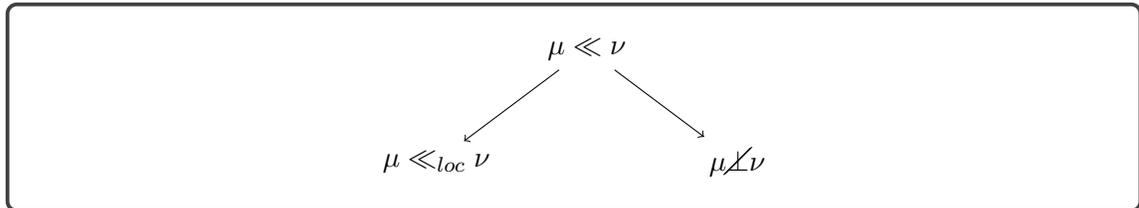


Figure 2.1: Logical dependencies between the notions of compatibility from §2.1.1. An arrow going from one notion to another indicates that, for any two measures μ and ν , compatibility in the sense of the first notion implies compatibility in the sense of the second notion.

2.1.2 Merging and polarisation in the classical framework

We are now ready to turn our attention to the phenomenon of merging of opinions—and to the other extreme: polarisation of opinions.

The most well-studied notion of merging of opinions in the literature was introduced

in a seminal article by Blackwell and Dubins [1962]:

Definition 2.1.8 (Merging). *Given measures μ and ν , ν is said to merge with μ (in symbols, $\nu \xrightarrow{M} \mu$) if, for every $\epsilon > 0$ and μ -almost every $\omega \in 2^{\mathbb{N}}$, there is some $N(\epsilon, \omega) \in \mathbb{N}$ such that, for all $n > N(\epsilon, \omega)$ and all $\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})$,*

$$\left| \nu(\mathcal{S} \mid \mathcal{F}_n)(\omega) - \mu(\mathcal{S} \mid \mathcal{F}_n)(\omega) \right| \leq \epsilon.$$

Having that $|\nu(\mathcal{S} \mid \mathcal{F}_n)(\omega) - \mu(\mathcal{S} \mid \mathcal{F}_n)(\omega)| \leq \epsilon$ for all $\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})$ is equivalent to having that $\sup_{\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})} |\nu(\mathcal{S} \mid \mathcal{F}_n)(\omega) - \mu(\mathcal{S} \mid \mathcal{F}_n)(\omega)| \leq \epsilon$. As a result, merging can be equivalently characterised as follows: $\nu \xrightarrow{M} \mu$ if, for μ -almost every $\omega \in 2^{\mathbb{N}}$,

$$\lim_{n \rightarrow \infty} \sup_{\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})} \left| \nu(\mathcal{S} \mid \mathcal{F}_n)(\omega) - \mu(\mathcal{S} \mid \mathcal{F}_n)(\omega) \right| = 0.$$

The distance $\sup_{\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})} |\nu(\mathcal{S}) - \mu(\mathcal{S})|$ between μ and ν is called the *total variation distance*. Essentially, it amounts to the largest possible difference between the probabilities that μ and ν can assign to the same event. Thus, $\sup_{\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})} |\nu(\mathcal{S} \mid \mathcal{F}_n)(\omega) - \mu(\mathcal{S} \mid \mathcal{F}_n)(\omega)|$ intuitively represents the maximum possible disagreement between μ and ν after having observed the outcomes of the first n experiments.

A crucial feature of this type of merging (and what makes it such a strong notion of consensus) is that it requires that the Bayesian agent with prior ν be eventually able to forecast correctly, or in agreement with measure μ , the probabilities of *any* event, including the probabilities of infinite-horizon events—more precisely, of events in the *tail σ -algebra* $\mathcal{G}_{\infty} = \bigcap_{n \in \mathbb{N}} \mathcal{G}_n$, where, for each $n \in \mathbb{N}$, $\mathcal{G}_n = \sigma(\bigcup_{i \geq n} \mathcal{F}_i)$.

Now, the following result, known as the Blackwell-Dubins Merging-of-Opinions Theorem (the Blackwell-Dubins Theorem, for short), is a foundational result in statistics and Bayesian epistemology. It establishes that absolute continuity is sufficient for merging: in other words, if ν is no more dogmatic than μ , then, with μ -probability one, ν will eventually agree with μ on the probability of all events, as the evidence accumulates.

Theorem 2.1.9 (Blackwell and Dubins [1962]). *Given measures μ and ν , if $\mu \ll \nu$, then $\nu \xrightarrow{M} \mu$.*

The Blackwell-Dubins Theorem also has a partial converse: merging entails absolute

continuity in the presence of local absolute continuity. In other words, under the assumption that ν shares the evidence with μ , merging of opinions in the sense of Blackwell and Dubins and absolute continuity are equivalent notions.

Theorem 2.1.10 (Kalai and Lehrer [1994]). *Given measures μ and ν such that $\mu \ll_{loc} \nu$, if $\nu \xrightarrow{M} \mu$, then $\mu \ll \nu$.*

As mentioned at the beginning of this chapter, the Blackwell-Dubins Theorem is philosophically significant because it is commonly taken to vindicate subjective Bayesianism by demonstrating that divergent initial opinions should not be a cause for concern: they do not threaten the objectivity of scientific inquiry, since objectivity can be recovered in a Peircean way in the form of inter-subjective agreement [Peirce, 1965]. Yet, merging of opinions is not guaranteed in all circumstances: as we have seen, it occurs when the agents' initial beliefs are sufficiently compatible. When the agent's priors are not similar enough, disagreement may persist, even as the evidence accumulates. For instance, it is easy to see that merging implies non-orthogonality, which means that if two measures are orthogonal, then they fail to merge. When combined with the Lebesgue Decomposition Theorem discussed in §2.1.1, this observation has an interesting consequence: for any two probability measures μ and ν , there is some $\alpha \in [0, 1]$ such that μ can be decomposed as $\mu = \alpha\mu_a + (1 - \alpha)\mu_o$, where μ_a is the part of μ with which ν asymptotically merges, while μ_o is the part of μ with which ν fails to merge.

The most radical failure of merging is polarisation of opinions, which occurs when disagreement, rather than being gradually eliminated by the shared evidence, becomes maximal as the available information increases.

Definition 2.1.11 (Polarisation). *Measures μ and ν are said to polarise (relative to μ) if there is a collection $\mathcal{P} \in \mathcal{B}(2^{\mathbb{N}})$ of data streams with $\mu(\mathcal{P}) > 0$ such that, for all $\omega \in \mathcal{P}$,*

$$\lim_{n \rightarrow \infty} \sup_{\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})} \left| \nu(\mathcal{S} \mid \mathcal{F}_n)(\omega) - \mu(\mathcal{S} \mid \mathcal{F}_n)(\omega) \right| = 1.$$

Maximal polarisation of opinions, which we denote by $\nu \parallel \mu$, occurs when $\mu(\mathcal{P}) = 1$.

Orthogonality and local absolute continuity together entail maximal polarisation of opinions: if two agents agree on what evidence is possible but their priors are orthogonal,

then, as they obtain more and more information, their beliefs become maximally divergent.

Observation 2.1.12. *Given measures μ and ν , if $\mu \perp \nu$ and $\mu \ll_{loc} \nu$, then $\nu \parallel \mu$.*

Proof. Suppose that there is some $\mathcal{C} \in \mathcal{B}(2^{\mathbb{N}})$ with $\mu(\mathcal{C}) = 0$ and $\nu(\mathcal{C}) = 1$, and that $\mu \ll_{loc} \nu$. Let \mathcal{U} be the set $\{\omega \in 2^{\mathbb{N}} : (\forall n) \mu([\omega \upharpoonright n]) > 0\}$. Clearly, $\mu(\mathcal{U}) = 1$. Moreover, for all $\omega \in \mathcal{U}$ and all $n \in \mathbb{N}$, $\nu([\omega \upharpoonright n]) > 0$, since $\mu \ll_{loc} \nu$. Hence, for all $\omega \in \mathcal{U}$ and all $n \in \mathbb{N}$, $\mu(\mathcal{C} \mid \mathcal{F}_n)(\omega) = 0$ and $\nu(\mathcal{C} \mid \mathcal{F}_n)(\omega) = 1$, from which it follows that $|\nu(\mathcal{C} \mid \mathcal{F}_n)(\omega) - \mu(\mathcal{C} \mid \mathcal{F}_n)(\omega)| = 1$. Therefore, $\sup_{\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})} |\nu(\mathcal{S} \mid \mathcal{F}_n)(\omega) - \mu(\mathcal{S} \mid \mathcal{F}_n)(\omega)| = 1$ for all $\omega \in \mathcal{U}$ and $n \in \mathbb{N}$, from which it follows that $\nu \parallel \mu$. \square

From the perspective of the Lebesgue Decomposition Theorem, this means that, for any two probability measures μ and ν , if $\mu \ll_{loc} \nu$, then there is some $\alpha \in [0, 1]$ such that μ can be decomposed as $\mu = \alpha\mu_a + (1-\alpha)\mu_o$, where μ_a is the part of μ with which ν asymptotically merges, while μ_o is the part of μ with which ν achieves maximal polarisation.¹¹

The results discussed in this section are summarised in Figure 2.2. In light of these observations, it appears paramount to provide a systematic analysis of the conditions under which merging occurs and those under which it fails. In particular, a clear understanding of which notions of compatibility lead to merging—and of how reasonable these notions are—would certainly help clarify the philosophical implications of the Blackwell-Dubins Theorem. Our results in the remainder of this chapter offer a step in this direction.

2.2 Algorithmic randomness

Classical merging-of-opinions results such as the Blackwell-Dubins Theorem are proven for arbitrary measures. In what follows, we will restrict attention to *computable* measures—and sometimes also make use of the more general concept of a *lower semi-computable* measure.¹² We will indicate it explicitly when the given measures are lower semi-computable; otherwise, from now on, all mentioned measures should be assumed to be computable. This

¹¹For a generalisation of the Blackwell-Dubins Theorem (which the authors dub the “Bayesian consensus-or-polarization law”) that combines merging and polarisation under the assumption of local absolute continuity, see Theorem 3 in [Nielsen and Stewart, 2019].

¹²See Definition 1.1.1 in Chapter 1.

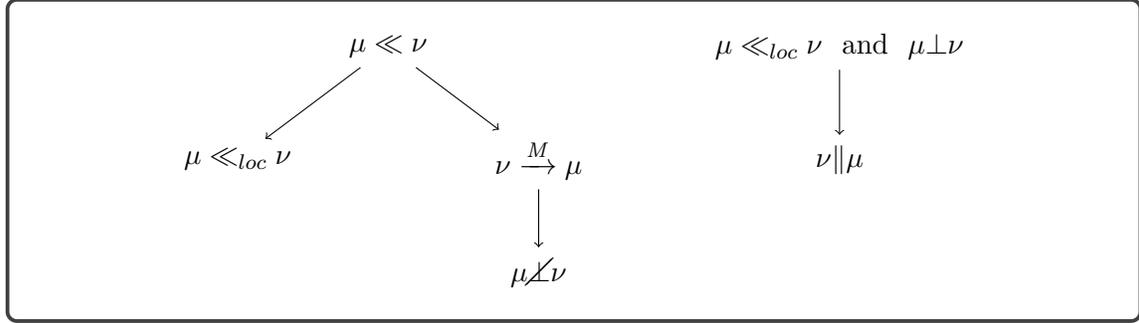


Figure 2.2: Logical dependencies between the notions of (in)compatibility discussed in §2.1.1 and the notions of merging/polarisation of opinions discussed in §2.1.2.

restriction is not motivated by normative considerations: we will not argue that computability is a rationality requirement on priors. Rather, it stems from the fact that our aim in this chapter is to elucidate the phenomenon of merging of opinions in the context of computationally limited Bayesian agents, which can be naturally identified with agents whose initial credences are given by computable (or, more generally, effective) priors. This perspective is in keeping with the overarching goal of this dissertation: namely, studying inductive learning and its connections with algorithmic randomness in the setting of less-than-ideal, computationally limited learners.

While focusing on effective measures of course means losing some generality, it also adds structure and allows to draw distinctions that were previously beyond reach (in this respect, our approach is consonant with that of Gaifman and Snir [1982]¹³). Notably, this computability-theoretic perspective allows to introduce many new more fine-grained notions of compatibility between priors, and to thereby represent the corresponding agents' inductive assumptions in a more detailed way.

From a methodological point of view, this is especially significant because, as mentioned earlier, the Blackwell-Dubins Theorem and the philosophical lesson standardly drawn from it crucially rely on absolute continuity: merging of opinions cannot be gotten for free, it follows when the agents' initial beliefs are sufficiently similar. Absolute continuity, however, is not without detractors (see, for instance, [Earman, 1992] and [Miller and Sanchirico,

¹³Our setting is, in a sense, less general, since we restrict attention to the Cantor space of infinite binary sequences, and to Δ_1^0 -definable (i.e., computable) and Σ_1^0 -definable (i.e., lower semi-computable) measures. See Chapter 3, §3.3.2 for a brief discussion of Gaifman and Snir's logical framework for modelling probabilistic learning.

1999]). Most recently, Nielsen and Stewart [2019] argued that there does not seem to be any plausible normative constraint on beliefs entailing that any two rationally permissible priors must be absolutely continuous, and that absolute continuity is also “of dubious descriptive value” [Nielsen and Stewart, 2019, p. 17].

The objection targeting the normative status of absolute continuity might very well be on point (and is in fact perfectly in line with the subjectivist outlook that merging-of-opinions theorems are meant to vindicate). The second worry voiced by Nielsen and Stewart, on the other hand, may be diffused by showing that there are many other forms of doxastic compatibility with a reasonable interpretation that either imply absolute continuity (thereby bolstering its descriptive value) or that, by themselves, guarantee the attainment of asymptotic consensus. Results of this type would provide a new way of looking at absolute continuity and show that, even if inter-subjective agreement cannot always be achieved, there are sufficiently many circumstances where a type of compatibility from which merging follows is realised to warrant holding out hope.

This chapter pursues this strategy to lend some additional credibility to the standard interpretation of merging-of-opinions theorems. In particular, we address the above worry by defining notions of compatibility induced by algorithmic randomness—a perspective that goes hand in hand with the computability-theoretic restrictions imposed on priors. Then, we show that agreeing on which data streams are algorithmically random indeed leads to merging of opinions between computable Bayesian agents—and that (maximally) disagreeing on which data streams are algorithmically random leads to (maximal) polarisation of opinions.

What motivates our focusing on algorithmic randomness? Though possibly surprising at first, as already hinted at by Skyrms, algorithmic randomness notions may be thought of as embodying beliefs in a special version of the principle of the uniformity of nature:

Without pursuing the matter in detail, I want to note a fact that is invariant over questions of fine tuning the analysis. It is that random sequences must have a limiting relative frequency. This is a rather spicy revelation in view of Reichenbach’s taking the existence of limiting relative frequencies as the principle of the uniformity of nature. The most chaotic and disordered alternative to uniformity that we can find *entails* uniformity-in-the-sense-of-Reichenbach!

[...] Randomness is indeed a kind of disorder, but it carries with it of necessity a kind of statistical order in the large. [Skyrms, 1984, p. 38]

In particular, as argued at the beginning of this chapter, algorithmic randomness notions may be taken to encode a specific type of inductive assumptions—or commitments (either explicit or implicit)—that result from the subjective prior with respect to which algorithmic randomness is defined. This is because algorithmic randomness notions embody all effective global regularities of a certain type that an agent expects to see in the data by virtue of having a certain prior. For instance, if their initial beliefs are captured by the Bernoulli measure with bias $\frac{2}{3}$ towards 1, the agent is (at least implicitly) making the inductive assumption that the limiting relative frequency of 0's in the true data stream is $\frac{1}{3}$. So, by believing that every sequence of n observations (or outcomes of the experiment under consideration) featuring k 0's has probability $\frac{1}{3}^k \frac{2}{3}^{(n-k)}$, the agent is also committed to believing in the relevant version of the Strong Law of Large Numbers. But algorithmic randomness captures inductive assumptions of exactly this type: i.e., commitments to believing that the data will display certain effective statistical regularities which stem from one's beliefs about events that can be settled with a finite number of observations.

2.2.1 Algorithmic randomness and merging

With this motivation in place, we can now turn to the properties of the notions of compatibility induced by algorithmic randomness. Recall that, given two computable priors μ and ν , as well as an algorithmic randomness notion R , we take ν to be compatible with μ with respect to R if $\mu\text{-}R \subseteq \nu\text{-}R$. Intuitively, this indicates that the agent with prior ν cannot be surprised by a data stream that the agent with prior μ considers typical.

Let us begin by reviewing some known results concerning the notions of compatibility yielded by Martin-Löf randomness and computable randomness. First, note that Martin-Löf randomness and computable randomness can also be characterised in terms of ratios of semi-measures and measures, respectively.

Theorem 2.2.1 (Folklore). *Let $\omega \in 2^{\mathbb{N}}$. The following are equivalent:*

- (1) ω is μ -Martin-Löf random;

- (2) $\mu([\omega \upharpoonright n]) > 0$ for all $n \in \mathbb{N}$, and $\limsup_{n \rightarrow \infty} \frac{\xi([\omega \upharpoonright n])}{\mu([\omega \upharpoonright n])} < \infty$ for all lower semi-computable semi-measures ξ .

Theorem 2.2.2 (Rute [2016]). *Let $\omega \in 2^{\mathbb{N}}$. The following are equivalent:*

- (1) ω is μ -computably random;
- (2) $\mu([\omega \upharpoonright n]) > 0$ for all $n \in \mathbb{N}$, and $\limsup_{n \rightarrow \infty} \frac{\xi([\omega \upharpoonright n])}{\mu([\omega \upharpoonright n])} < \infty$ for all computable measures ξ .

With these characterisations at hand, we can rehearse the proof of the following result, originally due to Muchnik et al. [1998], which establishes that, for any two computable measures, agreeing on which data streams are computably random entails agreeing on which data streams are Martin-Löf random.

Proposition 2.2.3 (Muchnik et al. [1998]). *If $\mu\text{-CR} \subseteq \nu\text{-CR}$, then $\mu\text{-MLR} \subseteq \nu\text{-MLR}$.*

Proof. Suppose that $\mu\text{-CR} \subseteq \nu\text{-CR}$ and that $\omega \in \mu\text{-MLR}$. Then, since $\mu\text{-MLR} \subseteq \mu\text{-CR}$, $\omega \in \mu\text{-CR}$, which entails that $\omega \in \mu\text{-MLR} \cap \nu\text{-CR}$. By Theorem 2.2.1, $\mu([\omega \upharpoonright n]) > 0$ for all $n \in \mathbb{N}$ and, by Theorem 2.2.2, $\nu([\omega \upharpoonright n]) > 0$ for all $n \in \mathbb{N}$. Now, suppose towards a contradiction that $\omega \notin \nu\text{-MLR}$. Then, by Theorem 2.2.1, there is a lower semi-computable semi-measure ξ such that $\limsup_{n \rightarrow \infty} \frac{\xi([\omega \upharpoonright n])}{\nu([\omega \upharpoonright n])} = \infty$. The fact that $\limsup_{n \rightarrow \infty} \frac{\xi([\omega \upharpoonright n])}{\nu([\omega \upharpoonright n])} = \infty$ entails that $\xi([\omega \upharpoonright n]) > 0$ for all $n \in \mathbb{N}$. Since ξ is a lower semi-computable semi-measure and $\omega \in \mu\text{-MLR}$, we also have that $\limsup_{n \rightarrow \infty} \frac{\xi([\omega \upharpoonright n])}{\mu([\omega \upharpoonright n])} < \infty$ by Theorem 2.2.1. But then,

$$\limsup_{n \rightarrow \infty} \frac{\mu([\omega \upharpoonright n])}{\nu([\omega \upharpoonright n])} = \limsup_{n \rightarrow \infty} \frac{\xi([\omega \upharpoonright n])}{\nu([\omega \upharpoonright n])} \bigg/ \frac{\xi([\omega \upharpoonright n])}{\mu([\omega \upharpoonright n])} = \infty,$$

which contradicts the fact that $\omega \in \nu\text{-CR}$. □

We now prove some new results pertaining to the type of compatibility induced by density randomness (cf. Definition 1.3.5). To this end, we first show that, just like Martin-Löf randomness and computable randomness, density randomness has a natural characterisation in terms of ratios of (semi-)measures, as well.

Theorem 2.2.4. *Let $\omega \in 2^{\mathbb{N}}$. The following are equivalent:*

- (1) ω is μ -density random;
- (2) $\mu([\omega \upharpoonright n]) > 0$ for all $n \in \mathbb{N}$, and $\lim_{n \rightarrow \infty} \frac{\xi([\omega \upharpoonright n])}{\mu([\omega \upharpoonright n])}$ exists and is finite for all lower semi-computable semi-measures ξ .

The proof of Theorem 2.2.4 relies on the following auxiliary lemma:

Lemma 2.2.5. *Fix a (computable) measure μ .*

- (i) *If d is a normed dyadic μ -martingale, then $\xi([\sigma]) = d(\sigma)\mu([\sigma])$ defines a measure. If d is left-c.e. or almost-everywhere left-c.e., then ξ is lower semi-computable, uniformly from d .*
- (ii) *If ξ is a semi-measure, then*

$$d(\sigma) \begin{cases} = \frac{\xi([\sigma])}{\mu([\sigma])} & \text{if } \mu([\sigma]) > 0, \\ \text{undefined} & \text{if } \mu([\sigma]) = 0 \end{cases}$$

is a dyadic μ -supermartingale. If ξ is a lower semi-computable semi-measure, then d is an almost-everywhere left-c.e dyadic μ -supermartingale. If ξ is a lower semi-computable semi-measure and μ is strictly positive, then d is a left-c.e dyadic μ -supermartingale.

Proof. (i) First, note that ξ is well-defined: by Definition 1.3.1, if $d(\sigma)$ is undefined, then $\mu([\sigma]) = 0$ and $d(\sigma)\mu([\sigma]) = 0$. Now, since d is normed, $d(\varepsilon) = 1$. Hence, $\xi([\varepsilon]) = d(\varepsilon)\mu([\varepsilon]) = 1$. Moreover, for all $\sigma \in 2^{<\mathbb{N}}$,

$$\begin{aligned} \xi([\sigma]) &= d(\sigma)\mu([\sigma]) \\ &= d(\sigma 0)\mu([\sigma 0]) + d(\sigma 1)\mu([\sigma 1]) \\ &= \xi([\sigma 0]) + \xi([\sigma 1]), \end{aligned}$$

where the second identity follows from the fairness condition in Definition 1.3.1. Given that d and μ are both non-negative, so is ξ . So, all that we have left to show is that $\xi([\sigma]) \leq 1$ for all $\sigma \in 2^{<\mathbb{N}}$. This follows from a simple argument by induction. We already know that $\xi([\varepsilon]) = 1$. Now, suppose that $\xi([\sigma]) \leq 1$. Then, $\xi([\sigma 0]) = \xi([\sigma]) - \xi([\sigma 1]) \leq \xi([\sigma]) \leq 1$.

The reasoning is analogous in the case of $\xi([\sigma 1])$. Hence, ξ is a measure.

Next, suppose that d is left-c.e. (and, thus, total). Let $h : 2^{<\mathbb{N}} \times \mathbb{N} \rightarrow \mathbb{Q}$ be a total computable function such that, for all $(\sigma, n) \in 2^{<\mathbb{N}} \times \mathbb{N}$, the sequence $\{h(\sigma, n)\}_{n \in \mathbb{N}}$ is non-decreasing and $\lim_{n \rightarrow \infty} h(\sigma, n) = d(\sigma)$. Without loss of generality, h can be assumed to be non-negative. Since μ is computable, it is also lower semi-computable. Therefore, for each $\sigma \in 2^{<\mathbb{N}}$, $\mu([\sigma])$ is a left-c.e. real, uniformly in σ . For each $\sigma \in 2^{<\mathbb{N}}$, let $\{q_{\sigma, n}\}_{n \in \mathbb{N}}$ be a computable non-decreasing sequence of rationals with $\lim_{n \rightarrow \infty} q_{\sigma, n} = \mu([\sigma])$. Without loss of generality, the $q_{\sigma, n}$'s can be assumed to be non-negative. For each $\sigma \in 2^{<\mathbb{N}}$, $\{h(\sigma, n) \cdot q_{\sigma, n}\}_{n \in \mathbb{N}}$ is thus a computable non-decreasing sequence of rational numbers such that $\lim_{n \rightarrow \infty} h(\sigma, n) \cdot q_{\sigma, n} = d(\sigma)\mu([\sigma]) = \xi([\sigma])$. Hence, $\xi([\sigma])$ is a left-c.e. real, uniformly in σ , which means that ξ is lower semi-computable. If, on the other hand, d is merely almost everywhere left-c.e., then it is a partial left-c.e. function. Let $h : \subseteq 2^{<\mathbb{N}} \times \mathbb{N} \rightarrow \mathbb{Q}$ be a non-negative partial computable function such that (1) for all $\sigma \in 2^{<\mathbb{N}}$, $d(\sigma)$ is defined if and only if $h(\sigma, n)$ is defined for all $n \in \mathbb{N}$, and (2) for all $\sigma \in 2^{<\mathbb{N}}$ such that $d(\sigma)$ is defined, $\{h(\sigma, n)\}_{n \in \mathbb{N}}$ is a non-decreasing sequence with $\lim_{n \rightarrow \infty} h(\sigma, n) = d(\sigma)$. As before, for each σ , let $\{q_{\sigma, n}\}_{n \in \mathbb{N}}$ be a computable non-decreasing sequence of non-negative rationals with $\lim_{n \rightarrow \infty} q_{\sigma, n} = \mu([\sigma])$. Now, for each σ and n , let

$$q'_{\sigma, n} = \begin{cases} h(\sigma, n) \cdot q_{\sigma, n} & \text{if } q_{\sigma, n} > 0; \\ 0 & \text{otherwise.} \end{cases}$$

Then, $\{q'_{\sigma, n}\}_{n \in \mathbb{N}}$ is a computable non-decreasing sequence of rational numbers such that $\lim_{n \rightarrow \infty} q'_{\sigma, n} = d(\sigma)\mu([\sigma]) = \xi([\sigma])$, which establishes that ξ is lower semi-computable.

(ii) If $d(\sigma)$ is undefined, then $\mu([\sigma]) = 0$ by definition. Hence, the impossibility condition is satisfied. If $\mu([\sigma]) = 0$, then $\mu([\sigma 0]) = \mu([\sigma 1]) = 0$. Hence, $d(\sigma)\mu([\sigma]) = 0 = d(\sigma 0)\mu([\sigma 0]) + d(\sigma 1)\mu([\sigma 1])$ (again, recall that we follow the convention that $d(\tau)\mu([\tau]) = 0$ if $\mu([\tau]) = 0$ even when $d(\tau)$ is undefined). If, on the other hand, $\mu([\sigma]) > 0$, then we have two cases to consider. First, suppose that $\mu([\sigma 0]) > 0$ and $\mu([\sigma 1]) > 0$. Then,

$$\begin{aligned} d(\sigma)\mu([\sigma]) &= \frac{\xi([\sigma])}{\mu([\sigma])}\mu([\sigma]) \\ &\geq \xi([\sigma 0]) + \xi([\sigma 1]) \end{aligned}$$

$$\begin{aligned}
&= \frac{\xi([\sigma 0])}{\mu([\sigma 0])} \mu([\sigma 0]) + \frac{\xi([\sigma 1])}{\mu([\sigma 1])} \mu([\sigma 1]) \\
&= d(\sigma 0) \mu([\sigma 0]) + d(\sigma 1) \mu([\sigma 1]),
\end{aligned}$$

where the inequality holds because ξ is by assumption a semi-measure. Second, suppose that either $\mu([\sigma 0]) = 0$ or $\mu([\sigma 1]) = 0$. Without loss of generality, assume that $\mu([\sigma 0]) = 0$. Then,

$$\begin{aligned}
d(\sigma) \mu([\sigma]) &= \frac{\xi([\sigma])}{\mu([\sigma])} \mu([\sigma]) \\
&\geq \xi([\sigma 0]) + \xi([\sigma 1]) \\
&\geq 0 + \xi([\sigma 1]) \\
&= d(\sigma 0) \mu([\sigma 0]) + d(\sigma 1) \mu([\sigma 1]).
\end{aligned}$$

Hence, the version of the fairness condition for dyadic supermartingales is satisfied in all cases.

Now, suppose that ξ is lower semi-computable. Define the function $h : \subseteq 2^{<\mathbb{N}} \times \mathbb{N} \rightarrow \mathbb{Q}$ as follows. If $\mu([\sigma]) = 0$, let $h(\sigma, n)$ be undefined for all $n \in \mathbb{N}$. If $\mu([\sigma]) > 0$, on the other hand, we do the following. Let $\{q_{\sigma, n}\}_{n \in \mathbb{N}}$ be a computable non-decreasing sequence of (without loss of generality) non-negative rationals witnessing the fact that $\xi([\sigma])$ is a left-c.e. real, uniformly in σ . Since μ is a computable measure, it is also upper semi-computable, which means that $\mu([\sigma])$ is a right-c.e. real, uniformly in σ . Let $\{q'_{\sigma, n}\}_{n \in \mathbb{N}}$ be a computable non-increasing sequence of positive rationals witnessing the fact that $\mu([\sigma])$ is right-c.e.: that is, $\lim_{n \rightarrow \infty} q'_{\sigma, n} = \mu([\sigma])$. Hence, $\{\frac{1}{q'_{\sigma, n}}\}_{n \in \mathbb{N}}$ is a computable non-decreasing sequence of positive rationals that converges to $\frac{1}{\mu([\sigma])}$. Define $h(\sigma, n)$ as $\frac{q_{\sigma, n}}{q'_{\sigma, n}}$ for all n . Then, h is a partial computable function, and the sequence $\{h(\sigma, n)\}_{n \in \mathbb{N}}$ is non-decreasing and converges to $\frac{\xi([\sigma])}{\mu([\sigma])} = d(\sigma)$ for all $\sigma \in 2^{<\mathbb{N}}$ such that $d(\sigma)$ is defined (i.e., all $\sigma \in 2^{<\mathbb{N}}$ such that $\mu([\sigma]) > 0$). Hence, d is almost everywhere left-c.e. (and it is left-c.e. if μ is strictly positive). \square

We are now ready to prove Theorem 2.2.4, which offers an alternative characterisation of the notion of density randomness.

Proof of Theorem 2.2.4. For the (1)-to-(2) direction, suppose that ω is μ -density random. Then, ω is also μ -Martin-Löf random. Hence, by Theorem 2.2.1, $\mu([\omega \upharpoonright n]) > 0$ for all $n \in \mathbb{N}$. Now, let ξ be a lower semi-computable semi-measure. By Lemma 2.2.5(ii), $\frac{\xi}{\mu}$ is an almost everywhere left-c.e. dyadic μ -supermartingale. By Theorem 1.3.6, $\lim_{n \rightarrow \infty} d(\omega \upharpoonright n)$ exists and is finite for all almost everywhere left-c.e. dyadic μ -supermartingales d . Hence, $\lim_{n \rightarrow \infty} \frac{\xi([\omega \upharpoonright n])}{\mu([\omega \upharpoonright n])}$ exists and is finite.

For the (2)-to-(1) direction, suppose that $\omega \notin \mu$ -DR. Then, by Theorem 1.3.6, there is a left-c.e. dyadic μ -martingale d that fails to converge to a finite value along ω . Without loss of generality, we can assume d to be normed. If $\mu([\omega \upharpoonright n]) = 0$ for some $n \in \mathbb{N}$, then we are done. So, suppose that $\mu([\omega \upharpoonright n]) > 0$ for all $n \in \mathbb{N}$. For each $\sigma \in 2^{<\mathbb{N}}$, let $\xi([\sigma]) = d(\sigma)\mu([\sigma])$. Then, by Lemma 2.2.5(i), ξ is a lower semi-computable measure. But then the sequence $\left\{ \frac{\xi([\omega \upharpoonright n])}{\mu([\omega \upharpoonright n])} \right\}_{n \in \mathbb{N}}$ either does not have a limit or $\lim_{n \rightarrow \infty} \frac{\xi([\omega \upharpoonright n])}{\mu([\omega \upharpoonright n])} = \infty$. \square

The above characterisation can then be put to use to show that agreeing on which data streams are Martin-Löf random entails agreeing on which data streams are density random.¹⁴

Proposition 2.2.6. *If μ -MLR $\subseteq \nu$ -MLR, then μ -DR $\subseteq \nu$ -DR.*

Proof. Suppose that μ -MLR $\subseteq \nu$ -MLR and $\omega \in \mu$ -DR. Then, $\omega \in \mu$ -MLR, which, in turn, entails that $\omega \in \mu$ -DR $\cap \nu$ -MLR. Suppose that $\omega \notin \nu$ -DR. Since $\omega \in \nu$ -MLR, Theorem 2.2.1 entails that $\nu([\omega \upharpoonright n]) > 0$ for all n . Thus, by Theorem 2.2.4, there must be a lower semi-computable semi-measure ξ such that either $\lim_{n \rightarrow \infty} \frac{\xi([\omega \upharpoonright n])}{\nu([\omega \upharpoonright n])} = \infty$ or the sequence

$\left\{ \frac{\xi([\omega \upharpoonright n])}{\nu([\omega \upharpoonright n])} \right\}_{n \in \mathbb{N}}$ does not have a limit.

Let us consider the second case first. The fact that $\left\{ \frac{\xi([\omega \upharpoonright n])}{\nu([\omega \upharpoonright n])} \right\}_{n \in \mathbb{N}}$ does not have a limit entails that $\xi([\omega \upharpoonright n]) > 0$ for all n , and that there are $a, b \in \mathbb{R}$ with $0 < a < b$ such that the number of upcrossings of the sequence $\left\{ \frac{\xi([\omega \upharpoonright n])}{\nu([\omega \upharpoonright n])} \right\}_{n \in \mathbb{N}}$ across the interval $[a, b]$ is infinite.

In addition, since $\omega \in \mu$ -DR, Theorem 2.2.4 entails that $\mu([\omega \upharpoonright n]) > 0$ for all n , and $\lim_{n \rightarrow \infty} \frac{\xi([\omega \upharpoonright n])}{\mu([\omega \upharpoonright n])}$ exists and is finite. Call this limit ℓ . We have two sub-cases to examine.

¹⁴Note that the implications that hold among the notions of compatibility induced by algorithmic randomness discussed so far are a mirror image of the implications that hold between the underlying algorithmic randomness concepts. We have that μ -CR $\subseteq \nu$ -CR entails that μ -MLR $\subseteq \nu$ -MLR, which, in turn, entails that μ -DR $\subseteq \nu$ -DR. On the other hand, as seen in Chapter 1, density randomness entails Martin-Löf randomness, which entails computable randomness.

First, suppose that $\ell = 0$. Then, $\lim_{n \rightarrow \infty} 1 \left/ \frac{\xi([\omega \upharpoonright n])}{\mu([\omega \upharpoonright n])} \right. = \infty$. For each n ,

$$\frac{\mu([\omega \upharpoonright n])}{\nu([\omega \upharpoonright n])} = \frac{\xi([\omega \upharpoonright n])}{\nu([\omega \upharpoonright n])} \left/ \frac{\xi([\omega \upharpoonright n])}{\mu([\omega \upharpoonright n])} \right.$$

(all three ratios are well-defined and positive because $\mu([\omega \upharpoonright n]) > 0$, $\nu([\omega \upharpoonright n]) > 0$, and $\xi([\omega \upharpoonright n]) > 0$ for all n). Since there are infinitely many n with $\frac{\xi([\omega \upharpoonright n])}{\nu([\omega \upharpoonright n])} \geq b > 0$, we have that

$$\limsup_{n \rightarrow \infty} \frac{\mu([\omega \upharpoonright n])}{\nu([\omega \upharpoonright n])} = \limsup_{n \rightarrow \infty} \frac{\xi([\omega \upharpoonright n])}{\nu([\omega \upharpoonright n])} \left/ \frac{\xi([\omega \upharpoonright n])}{\mu([\omega \upharpoonright n])} \right. \geq \lim_{n \rightarrow \infty} \left(b \left/ \frac{\xi([\omega \upharpoonright n])}{\mu([\omega \upharpoonright n])} \right. \right) = \infty.$$

This, however, contradicts the fact that $\omega \in \nu$ -MLR. So, suppose instead that $\ell > 0$. Let $a' = \frac{2}{3}a + \frac{1}{3}b$ and $b' = \frac{1}{3}a + \frac{2}{3}b$. Then, $0 < a' < b'$, and the number of upcrossings of the sequence $\left\{ \frac{\xi([\omega \upharpoonright n])}{\nu([\omega \upharpoonright n])} \right\}_{n \in \mathbb{N}}$ across the interval $[a', b']$ is infinite. Moreover, for each of the infinitely many n such that $\frac{\xi([\omega \upharpoonright n])}{\nu([\omega \upharpoonright n])} \leq a < a'$, $a' - \frac{\xi([\omega \upharpoonright n])}{\nu([\omega \upharpoonright n])} > \frac{1}{3}(b - a)$, and for each of the infinitely many n such that $\frac{\xi([\omega \upharpoonright n])}{\nu([\omega \upharpoonright n])} \geq b > b'$, $\frac{\xi([\omega \upharpoonright n])}{\nu([\omega \upharpoonright n])} - b' > \frac{1}{3}(b - a)$. But then we have that

$$\limsup_{n \rightarrow \infty} \frac{\nu([\omega \upharpoonright n])}{\mu([\omega \upharpoonright n])} = \limsup_{n \rightarrow \infty} \frac{\xi([\omega \upharpoonright n])}{\mu([\omega \upharpoonright n])} \left/ \frac{\xi([\omega \upharpoonright n])}{\nu([\omega \upharpoonright n])} \right. \geq \frac{\ell}{a'}, \text{ and}$$

$$\liminf_{n \rightarrow \infty} \frac{\nu([\omega \upharpoonright n])}{\mu([\omega \upharpoonright n])} = \liminf_{n \rightarrow \infty} \frac{\xi([\omega \upharpoonright n])}{\mu([\omega \upharpoonright n])} \left/ \frac{\xi([\omega \upharpoonright n])}{\nu([\omega \upharpoonright n])} \right. \leq \frac{\ell}{b'}.$$

Since, ℓ, a' and b' are all positive and $a' < b'$, $\frac{\ell}{a'} > \frac{\ell}{b'}$. Hence, the sequence $\left\{ \frac{\nu([\omega \upharpoonright n])}{\mu([\omega \upharpoonright n])} \right\}_{n \in \mathbb{N}}$ fails to converge, which, by Theorem 2.2.4, contradicts the assumption that $\omega \in \mu$ -DR.

Let us now consider the first case: that is, suppose that $\lim_{n \rightarrow \infty} \frac{\xi([\omega \upharpoonright n])}{\nu([\omega \upharpoonright n])} = \infty$. This implies that $\xi([\omega \upharpoonright n]) > 0$ for all n . Since $\lim_{n \rightarrow \infty} \frac{\xi([\omega \upharpoonright n])}{\mu([\omega \upharpoonright n])} < \infty$ by Theorem 2.2.4, we also have that

$$\limsup_{n \rightarrow \infty} \frac{\mu([\omega \upharpoonright n])}{\nu([\omega \upharpoonright n])} = \limsup_{n \rightarrow \infty} \frac{\xi([\omega \upharpoonright n])}{\nu([\omega \upharpoonright n])} \left/ \frac{\xi([\omega \upharpoonright n])}{\mu([\omega \upharpoonright n])} \right. = \infty,$$

which contradicts the fact that $\omega \in \nu$ -MLR. Hence, $\omega \in \nu$ -DR. \square

Next, we will see that agreeing on which data streams are density random entails absolute continuity. *A fortiori*, by Proposition 2.2.6 and Proposition 2.2.3, agreeing on which data streams are Martin-Löf random and agreeing on which data streams are computably random entail absolute continuity, as well.¹⁵

Proposition 2.2.7. *If μ -DR \subseteq ν -DR, then $\mu \ll \nu$.*

Proof. Suppose there is some $\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})$ with $\nu(\mathcal{S}) = 0$, but $\mu(\mathcal{S}) > 0$. Then, there is some $q \in \mathbb{Q}$ with $\mu(\mathcal{S}) > q > 0$. Since ν is regular,¹⁶ $\nu(\mathcal{S}) = \inf\{\nu(\mathcal{U}) : \mathcal{S} \subseteq \mathcal{U} \text{ and } \mathcal{U} \in \mathcal{B}(2^{\mathbb{N}}) \text{ is an open set}\}$. Hence, for all $n \in \mathbb{N}$, there is an open set \mathcal{U}_n with $\mathcal{S} \subseteq \mathcal{U}_n$ such that $\nu(\mathcal{U}_n) < 2^{-n}$ and $\mu(\mathcal{U}_n) > q$. Every \mathcal{U}_n is of the form $\bigcup_{i \in \mathbb{N}} [\sigma_{n,i}]$ —where, without loss of generality, the sets $[\sigma_{n,i}]$ can be taken to be pairwise disjoint. For each \mathcal{U}_n , there is some K_n such that (1) $\mu(\bigcup_{i \leq K_n} [\sigma_{n,i}]) \geq q$, while (2) $\nu(\bigcup_{i \leq K_n} [\sigma_{n,i}]) < 2^{-n}$. Let $V_n = \{\sigma_{n,0}, \dots, \sigma_{n,K_n}\}$ and let $\bigcup_{i \leq K_n} [\sigma_{n,i}]$ be denoted as $[V_n]$. For each $m \in \mathbb{N}$, let $\mathcal{V}_m = \bigcup_{n > m} [V_n]$. Then, we have that $\nu(\mathcal{V}_m) \leq \sum_{n > m} \nu([V_n]) \leq \sum_{n > m} 2^{-n} \leq 2^{-m}$ and, since $\mu([V_n]) \geq q$ for all n , $\mu(\mathcal{V}_m) \geq q$, as well. Note that the sets V_n can be chosen in such a way that $\{\mathcal{V}_m\}_{m \in \mathbb{N}}$ is a sequence of uniformly Σ_1^0 classes. Given n , simply enumerate the strings in $2^{< \mathbb{N}}$ (for instance, in the length-lexicographic order) until conditions (1) and (2) are met. By the above, we are guaranteed that a finite, prefix-free collection of cylinders satisfying these conditions will eventually be found, effectively. Hence, $\{\mathcal{V}_m\}_{m \in \mathbb{N}}$ is a ν -Martin-Löf test. This entails that $\bigcap_{m \in \mathbb{N}} \mathcal{V}_m \cap \nu\text{-MLR} = \emptyset$. Given that $\nu\text{-DR} \subseteq \nu\text{-MLR}$, $\bigcap_{m \in \mathbb{N}} \mathcal{V}_m \cap \nu\text{-DR} = \emptyset$. However, since $\mu(\mathcal{V}_m) \geq q$ for all m and the sequence $\{\mathcal{V}_m\}_{m \in \mathbb{N}}$ is nested, $\mu(\bigcap_{m \in \mathbb{N}} \mathcal{V}_m) > 0$. Due to the fact that $\mu\text{-DR}$ has μ -measure one, we therefore have that $\bigcap_{m \in \mathbb{N}} \mathcal{V}_m \cap \mu\text{-DR} \neq \emptyset$. Hence, $\mu\text{-DR} \not\subseteq \nu\text{-DR}$. \square

As mentioned above, Proposition 2.2.7 allows to conclude that agreeing on which data streams are computably random, Martin-Löf random, and density random all entail absolute continuity. This is epistemologically significant because, by the Blackwell-Dubins Theorem, we then have that these three different forms of compatibility induced by algorithmic randomness all ensure asymptotic merging of opinions. In other words, the

¹⁵The fact that $\mu\text{-MLR} \subseteq \nu\text{-MLR}$ entails $\mu \ll \nu$ was proven by Bienvenu and Merkle [2009]. The proof of Proposition 2.2.7 is analogous to the proof of this fact.

¹⁶Regularity follows from the fact that ν is a Borel probability measure and Cantor space is a locally compact Hausdorff space with a countable base.

inductive assumptions encoded by these core algorithmic randomness notions, the commitments to the global uniformity of nature that they each represent, when shared, guarantee the attainment of inter-subjective agreement between different Bayesian agents.

Corollary 2.2.8. *If $\mu\text{-DR} \subseteq \nu\text{-DR}$, then $\nu \xrightarrow{\text{M}} \mu$. A fortiori, if $\mu\text{-MLR} \subseteq \nu\text{-MLR}$, then $\nu \xrightarrow{\text{M}} \mu$, and if $\mu\text{-CR} \subseteq \nu\text{-CR}$, then $\nu \xrightarrow{\text{M}} \mu$.*

While it does not entail the other notions of compatibility induced by randomness, agreeing on which data streams are Schnorr random entails absolute continuity, too:¹⁷

Proposition 2.2.9 (Bienvenu and Merkle [2009]). *If $\mu\text{-SR} \subseteq \nu\text{-SR}$, then $\mu \ll \nu$.*

Thus, the inductive assumptions encapsulated by Schnorr randomness lead to almost sure inter-subjective agreement, as well.

Corollary 2.2.10. *If $\mu\text{-SR} \subseteq \nu\text{-SR}$, then $\nu \xrightarrow{\text{M}} \mu$.*

None of the above implications can be reversed: that is, for each of them, it is possible to find two computable measures for which the converse implication fails (see Appendix 2.A). This evinces that approaching the question of when inter-subjective agreement is attainable from the perspective of algorithmic randomness affords a richer, more fine-grained analysis of the types of commitments and inductive assumptions that a computable Bayesian agent can make. Taken together, the above results attest to the utility of applying algorithmic randomness, as a tool for modelling various types of agreement about the global uniformity of nature, in the study of Bayesian learning—and, specifically, of the phenomenon of merging of opinions.

Next, we turn our attention to the weak n -randomness hierarchy. First of all, in the effective setting, it is natural to consider the following versions of absolute continuity, which only apply to Π_n^0 and Σ_n^0 classes,¹⁸ respectively.

Definition 2.2.11 (Π_n^0 -absolute continuity and Σ_n^0 -absolute continuity). *A measure μ is said to be Π_n^0 -absolutely continuous with respect to measure ν (in symbols, $\mu \ll_{\Pi_n^0} \nu$) if, for any Π_n^0 class $\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})$, $\mu(\mathcal{S}) > 0$ entails that $\nu(\mathcal{S}) > 0$. Similarly, μ is said to be*

¹⁷In fact, by inspecting the proof of Proposition 2.2.7, it can be seen that the measures $\nu(\mathcal{V}_m)$ are computable reals, uniformly in m , which means that $\{\mathcal{V}_m\}_{m \in \mathbb{N}}$ is also a ν -Schnorr test.

¹⁸See Definition 1.2.1 in Chapter 1.

Σ_n^0 -absolutely continuous with respect to ν (in symbols, $\mu \ll_{\Sigma_n^0} \nu$) if, for any Σ_n^0 class $\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})$, $\mu(\mathcal{S}) > 0$ entails that $\nu(\mathcal{S}) > 0$.

Without loss of generality, we can focus on Π_n^0 -absolute continuity, given that Π_n^0 -absolute continuity is equivalent to Σ_{n+1}^0 -absolute continuity:

Observation 2.2.12. *For all $n \geq 1$, $\mu \ll_{\Pi_n^0} \nu$ if and only if $\mu \ll_{\Sigma_{n+1}^0} \nu$.*

Proof. For the left-to-right direction, suppose that $\mu \ll_{\Pi_n^0} \nu$ and let \mathcal{S} be a Σ_{n+1}^0 class with $\nu(\mathcal{S}) = 0$. Then, $\mathcal{S} = \bigcup_{m \in \mathbb{N}} \mathcal{S}_m$, where $\{\mathcal{S}_m\}_{m \in \mathbb{N}}$ is a sequence of uniformly Π_n^0 classes. Since $\nu(\mathcal{S}) = 0$, $\nu(\mathcal{S}_m) = 0$ for all $m \in \mathbb{N}$. Hence, given that $\mu \ll_{\Pi_n^0} \nu$, $\mu(\mathcal{S}_m) = 0$ for all $m \in \mathbb{N}$, which, in turn, entails that $\mu(\mathcal{S}) = 0$.

Since every Π_1^0 class is also a Σ_{n+1}^0 class, the right-to-left direction is immediate. \square

For each n , the notion of compatibility yielded by weak n -randomness coincides with Π_n^0 -absolute continuity.

Observation 2.2.13. *For all $n \geq 1$, $\mu \ll_{\Pi_n^0} \nu$ if and only if $\mu\text{-WnR} \subseteq \nu\text{-WnR}$.*

Proof. For the left-to-right direction, suppose that $\omega \in \mu\text{-WnR}$ and let \mathcal{C} be a Σ_n^0 class of ν -measure one. Then, $\bar{\mathcal{C}}$ is a Π_n^0 class of ν -measure zero. Since $\mu \ll_{\Pi_n^0} \nu$, it follows that $\mu(\bar{\mathcal{C}}) = 0$ and $\mu(\mathcal{C}) = 1$. Then, the fact that $\omega \in \mu\text{-WnR}$ entails that $\omega \in \mathcal{C}$. But since \mathcal{C} was an arbitrary Σ_n^0 class of ν -measure one, we can conclude that $\omega \in \nu\text{-WnR}$.

For the right-to-left direction, suppose that there is a Π_n^0 class \mathcal{A} such that $\nu(\mathcal{A}) = 0$, but $\mu(\mathcal{A}) > 0$. Then, $\bar{\mathcal{A}}$ is Σ_n^0 class of ν -measure one, which entails that $\nu\text{-WnR} \subseteq \bar{\mathcal{A}}$. However, $\mu(\bar{\mathcal{A}}) < 1$, so $\mu\text{-WnR} \not\subseteq \bar{\mathcal{A}}$, because the collection of μ -weakly n -random sequences has μ -measure one. Hence, $\mu\text{-WnR} \not\subseteq \nu\text{-WnR}$. \square

Recall that a Π_2^0 class is the effective analogue of a G_δ subset of Cantor space: namely, a countable intersection of open sets. The proposition below is the effective version (in the context of computable measures) of the well-known equivalence of absolute continuity and absolute continuity restricted to G_δ sets.

Proposition 2.2.14 (Folklore). *Given measures μ and ν , $\mu \ll \nu$ if and only if $\mu \ll_{\Pi_2^0} \nu$.*

Proof. We only go through the non-trivial direction. Suppose there is $\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})$ with $\nu(\mathcal{S}) = 0$, but $\mu(\mathcal{S}) > q > 0$, for some rational q . Since $\nu(\mathcal{S}) = \inf\{\nu(\mathcal{U}) : \mathcal{S} \subseteq \mathcal{U} \text{ and } \mathcal{U} \in$

$\mathcal{B}(2^{\mathbb{N}})$ is an open set}, for all n , there is an open set \mathcal{U}_n with $\mathcal{S} \subseteq \mathcal{U}_n$ such that $\nu(\mathcal{U}_n) < 2^{-n}$ and $\mu(\mathcal{U}_n) > q$. Every \mathcal{U}_n is of the form $\bigcup_{i \in \mathbb{N}} [\sigma_{n,i}]$, where the cylinders $[\sigma_{n,i}]$ are pairwise disjoint. For each \mathcal{U}_n , we can effectively find some K_n such that $\mu(\bigcup_{i \leq K_n} [\sigma_{n,i}]) \geq q - 2^{-n}$, while $\nu(\bigcup_{i \leq K_n} [\sigma_{n,i}]) < 2^{-n}$. We will denote $\bigcup_{i \leq K_n} [\sigma_{n,i}]$ as \mathcal{V}_n . Let $\mathcal{C} = \bigcap_{m \in \mathbb{N}} \bigcup_{n \geq m} \mathcal{V}_n$. Then, \mathcal{C} is a Π_2^0 class. By the First Borel-Cantelli Lemma,¹⁹ $\nu(\mathcal{C}) = 0$. However, by the Reverse Fatou Lemma,²⁰ $\mu(\mathcal{S}) \geq \limsup_{n \rightarrow \infty} \mu(\mathcal{V}_n) = q$. \square

Proposition 2.2.14 establishes that, past the second level, the Π_n^0 -absolute continuity hierarchy collapses: that is, for all $n \geq 2$, Π_n^0 -absolute continuity is not a weaker form of absolute continuity, it actually coincides with it. Hence, Observation 2.2.13 and Proposition 2.2.14 together entail that, for every $n \geq 2$, having compatible beliefs about which data streams are weakly n -random is the same as absolute continuity.

Corollary 2.2.15. *For all $n \geq 2$, $\mu \ll \nu$ if and only if μ -WnR \subseteq ν -WnR.*

An immediate consequence of this equivalence is that, for all $n \geq 2$, the type of compatibility yielded by weak n -randomness guarantees merging of opinions—a fact that lends further credibility to the use of algorithmic randomness to define notions of doxastic compatibility.

Corollary 2.2.16. *For all $n \geq 2$, if μ -WnR \subseteq ν -WnR, then $\nu \xrightarrow{M} \mu$.*

In light of the above, it is then natural to ask whether Π_1^0 -absolute continuity coincides with absolute continuity, as well—and if not (that is, if Proposition 2.2.14 turns out to be the strongest possible result), whether the notion of compatibility yielded by weak 1-randomness nonetheless entails merging. The first question, first raised by Gaifman and Snir [1982], was given a negative answer by Bienvenu and Merkle [2009]. The notion of compatibility induced by weak 1-randomness is the only one, among the compatibility

¹⁹The First Borel-Cantelli Lemma is the following result (see, for instance, [Durrett, 2010, §2.3]):

Theorem (First Borel-Cantelli Lemma). *Given a measure μ , let $\{\mathcal{S}_n\}_{n \in \mathbb{N}}$ be a sequence of events in $\mathcal{B}(2^{\mathbb{N}})$ such that $\sum_{n \in \mathbb{N}} \mu(\mathcal{S}_n) < \infty$. Then, $\mu(\bigcap_{m \in \mathbb{N}} \bigcup_{n > m} \mathcal{S}_n) = 0$.*

²⁰The Reverse Fatou Lemma is the following result:

Theorem (Reverse Fatou Lemma for sets). *Given a measure μ , let $\{\mathcal{S}_n\}_{n \in \mathbb{N}}$ be a sequence of events in $\mathcal{B}(2^{\mathbb{N}})$. Then, $\mu(\bigcap_{m \in \mathbb{N}} \bigcup_{n > m} \mathcal{S}_n) \geq \limsup_{n \rightarrow \infty} \mu(\mathcal{S}_n)$.*

concepts generated by the weak n -randomness family, that is strictly weaker than absolute continuity.

Proposition 2.2.17 (Bienvenu and Merkle [2009]). *There exist computable measures μ and ν such that μ -W1R \subseteq ν -W1R, but $\mu \not\ll \nu$.*

The two measures used in the proof of Proposition 2.2.17 are strictly positive. This allows us to provide a negative answer to the second question above:

Proposition 2.2.18. *There exist computable measures μ and ν with μ -W1R \subseteq ν -W1R, and yet $\nu \not\stackrel{M}{\rightarrow} \mu$.*

Proof. Let μ and ν be the measures from the proof of Proposition 2.2.17 (see Proposition 56 in [Bienvenu and Merkle, 2009]). Since μ and ν are both strictly positive, we have that $\mu \ll_{loc} \nu$. Hence, by Theorem 2.1.10, if ν were to merge with μ , we would have that $\mu \ll \nu$. But, by Proposition 2.2.17, we know that $\mu \not\ll \nu$. Hence, $\nu \not\stackrel{M}{\rightarrow} \mu$. \square

Thus, while agreement on randomness generally entails merging, weak 1-randomness is the exception: having compatible inductive assumptions about the global regularities encoded by weak 1-randomness does not suffice to attain inter-subjective agreement in the sense of Blackwell and Dubins. This finding is in itself interesting because, as discussed in Chapter 1, weak 1-randomness is a bit of an outlier within the algorithmic randomness hierarchy. In particular, it does not entail several basic statistical laws and global regularities, such as the Strong Law of Large Numbers, the Law of the Iterated Logarithm, and Borel normality. Hence, it is perhaps not so surprising that agreeing on which data streams are weakly 1-random does not ensure merging. In fact, this failure can be taken to corroborate our explanation for why it is reasonable to use algorithmic randomness to define notions of compatibility, in that it shows that making compatible inductive assumptions about the global uniformity of nature—about sufficiently many statistical laws—is necessary for merging.

Now, even though agreement on weak 1-randomness does not entail absolute continuity, it does entail local absolute continuity (this simply follows from the fact that cylinders, being Δ_1^0 classes, are also Π_1^0 classes). The converse, however, does not hold, as shown by the following example.

Example 2.2.19. *Take the uniform measure λ and the Bernoulli measure $\mu_{\frac{1}{3}}$. Then, $\lambda \ll_{loc} \mu_{\frac{1}{3}}$. Let $\mathcal{C} = \{\omega \in 2^{\mathbb{N}} : (\exists m)(\forall n \geq m) \frac{\#0(\omega|n)}{n} > \frac{7}{20}\}$. Clearly, $\mu_{\frac{1}{3}}(\mathcal{C}) = 0$, while $\lambda(\mathcal{C}) = 1$. Since \mathcal{C} is a Σ_2^0 class, we then have that $\lambda \not\ll_{\Sigma_2^0} \mu_{\frac{1}{3}}$ and, consequently, that λ -W1R $\not\subseteq \mu_{\frac{1}{3}}$ -W1R (by Observation 2.2.12 and Observation 2.2.13).*

A more general reason for why local absolute continuity does not entail agreement on weak 1-randomness is that the latter entails non-orthogonality (see Proposition 40 and Corollary 41 in [Bienvenu and Merkle, 2009]), while, as seen earlier, local absolute continuity does not. In [2009], Bienvenu and Merkle also show that non-orthogonality does not entail agreement on weak 1-randomness (see Proposition 54).

Since the type of compatibility induced by weak 1-randomness does not entail merging in the sense of Blackwell and Dubins, an immediate question is whether there is some weaker type of merging that agreement on weak 1-randomness might nonetheless guarantee. A natural candidate is the following notion, first investigated by Kalai and Lehrer [1994]:

Definition 2.2.20 (Weak merging). *Given measures μ and ν , ν is said to weakly merge with μ (in symbols, $\nu \xrightarrow{\text{WM}} \mu$) if, for any $\ell \in \mathbb{N}$, any $\epsilon > 0$, and μ -almost every $\omega \in 2^{\mathbb{N}}$, there is some $N(\ell, \epsilon, \omega) \in \mathbb{N}$ such that, for all $n > N(\ell, \epsilon, \omega)$ and all $\mathcal{S} \in \mathcal{F}_{n+\ell}$,*

$$\left| \nu(\mathcal{S} \mid \mathcal{F}_n)(\omega) - \mu(\mathcal{S} \mid \mathcal{F}_n)(\omega) \right| \leq \epsilon.$$

As shown by Kalai and Lehrer [1994] (see, also, [Lehrer and Smorodinsky, 1996]), weak merging can be equivalently defined by letting $\ell = 1$. Hence, the above definition can be rewritten as follows: $\nu \xrightarrow{\text{WM}} \mu$ if, for μ -almost every $\omega \in 2^{\mathbb{N}}$,

$$\lim_{n \rightarrow \infty} \sup_{\mathcal{S} \in \mathcal{F}_{n+1}} \left| \nu(\mathcal{S} \mid \mathcal{F}_n)(\omega) - \mu(\mathcal{S} \mid \mathcal{F}_n)(\omega) \right| = 0.$$

The reason why weak merging is a natural candidate is that it only requires merging of opinions on finite-horizon events (as opposed to requiring merging on all events, as in the case of the notion of merging considered by Blackwell and Dubins). Likewise, the type of compatibility induced by weak 1-randomness targets statistical laws that correspond to Σ_1^0 classes: namely, laws whose satisfaction can be verified with a finite number of observations. Therefore, while the fact that weak 1-randomness does not entail some crucial statistical

regularities prevents it from yielding a type of compatibility that ensures merging, the inductive assumptions captured by Σ_1^0 classes of measure one might nonetheless suffice for weak merging, which is less demanding. We leave it as an open question whether this is indeed the case.

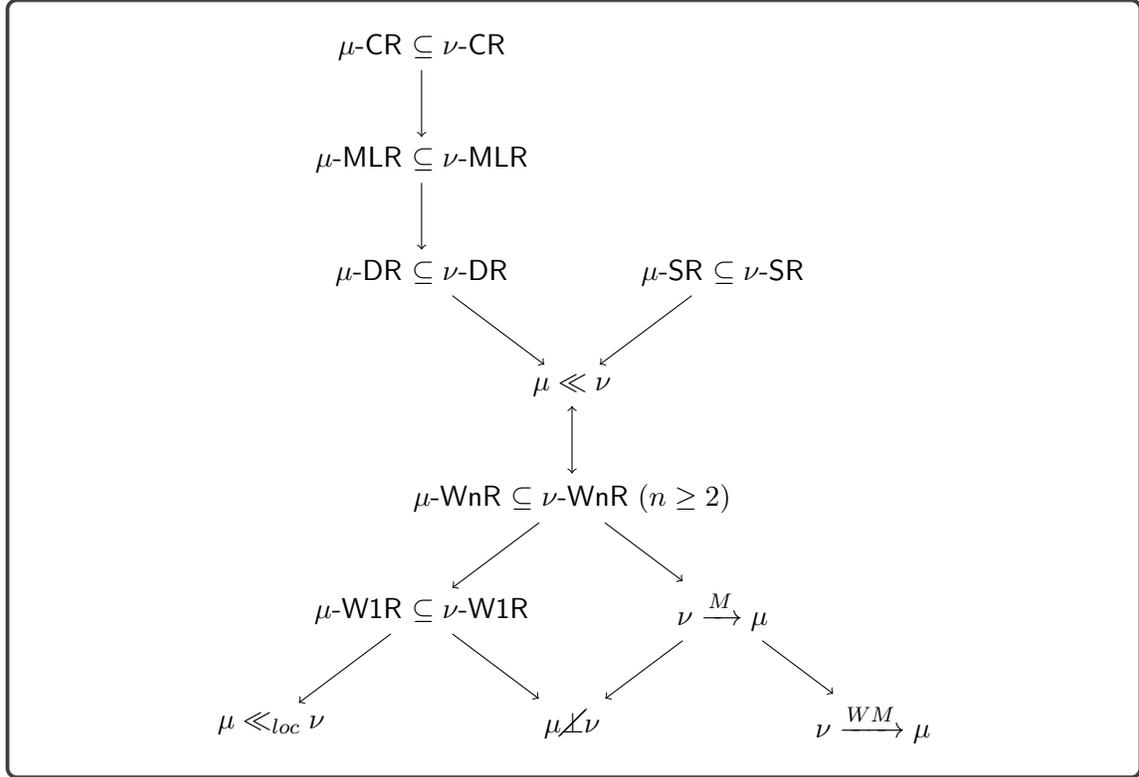


Figure 2.3: Logical dependencies between the notions of compatibility induced by algorithmic randomness, the classical notions of compatibility we discussed, merging in the sense of Blackwell and Dubins, and weak merging.

2.2.2 Algorithmic randomness and polarisation

Algorithmic randomness can be used to define not only notions of compatibility between priors, but also notions of incompatibility or disagreement. Just as orthogonality corresponds to the most radical failure of absolute continuity, for any algorithmic randomness notion R , two measures μ and ν are radically R -incompatible when $\mu-R \cap \nu-R = \emptyset$.

Surprisingly, the logical dependencies between the compatibility notions yielded by algorithmic randomness are very different from the logical relations that hold among the

underlying randomness concepts (compare Figure 2.3 and Figure 1.3). The logical dependencies between the types of incompatibility induced by algorithmic randomness, on the other hand, are a mirror image of the algorithmic randomness hierarchy.

Observation 2.2.21. *Given measures μ and ν , the following hold:*

- (i) *if $\mu\text{-W1R} \cap \nu\text{-W1R} = \emptyset$, then $\mu\text{-SR} \cap \nu\text{-SR} = \emptyset$;*
- (ii) *if $\mu\text{-SR} \cap \nu\text{-SR} = \emptyset$, then $\mu\text{-CR} \cap \nu\text{-CR} = \emptyset$;*
- (iii) *if $\mu\text{-CR} \cap \nu\text{-CR} = \emptyset$, then $\mu\text{-MLR} \cap \nu\text{-MLR} = \emptyset$;*
- (iv) *if $\mu\text{-MLR} \cap \nu\text{-MLR} = \emptyset$, then $\mu\text{-DR} \cap \nu\text{-DR} = \emptyset$;*
- (v) *if $\mu\text{-MLR} \cap \nu\text{-MLR} = \emptyset$, then $\mu\text{-W2R} \cap \nu\text{-W2R} = \emptyset$;*
- (vi) *if $\mu\text{-WnR} \cap \nu\text{-WnR} = \emptyset$, then $\mu\text{-Wn+1R} \cap \nu\text{-Wn+1R} = \emptyset$ for all $n \geq 1$.*

Proof. All of the above cases are proved in the same way: they rely on the logical dependencies between the algorithmic randomness notions involved. As an example, consider case (i). Suppose that $\mu\text{-W1R} \cap \nu\text{-W1R} = \emptyset$. Since $\mu\text{-SR} \subseteq \mu\text{-W1R}$ and $\nu\text{-SR} \subseteq \nu\text{-W1R}$, we can immediately conclude that $\mu\text{-SR} \cap \nu\text{-SR} = \emptyset$. \square

It is also easy to see that, for any algorithmic randomness notion R and measures μ, ν , if $\mu\text{-R} \cap \nu\text{-R} = \emptyset$, then $\mu \perp \nu$. This follows from the fact that $\mu(\mu\text{-R}) = \nu(\nu\text{-R}) = 1$, which, together with $\mu\text{-R} \cap \nu\text{-R} = \emptyset$, entails that $\nu(\mu\text{-R}) = 0$ and $\mu(\nu\text{-R}) = 0$. Hence, if $\mu\text{-R} \cap \nu\text{-R} = \emptyset$, then $\nu \not\stackrel{M}{\rightarrow} \mu$ and $\mu \not\stackrel{M}{\rightarrow} \nu$. Moreover, if $\mu\text{-R} \cap \nu\text{-R} = \emptyset$ and $\mu \ll_{loc} \nu$, then $\nu \parallel \mu$: radical disagreement over which data streams are algorithmically random and local absolute continuity entail polarisation of opinions.

2.3 Conclusion

We conclude the chapter by offering a synopsis of our results, and by discussing some open questions and possible avenues for future research.

The results presented here evince that the phenomenon of merging of opinions can be fruitfully studied in a computability-theoretic setting by appealing to the theory of

algorithmic randomness. Algorithmic randomness can in fact be employed to define notions of agreement and disagreement between priors which reflect a Bayesian agent's initial beliefs about the statistical regularities that they expect to see in the observational data. Our main contribution consisted in showing that the resulting notions of agreement and disagreement provably lead to merging and polarisation of opinions, respectively.

We focused on the notion of merging of opinions introduced by Blackwell and Dubins [1962]. More precisely, we showed that, apart from weak 1-randomness, all core algorithmic randomness notions give rise to forms of compatibility which secure merging in the sense of Blackwell and Dubins (by virtue of entailing absolute continuity). The notion of compatibility yielded by weak 1-randomness, on the other hand, is too weak to entail merging in the sense of Blackwell and Dubins. Weaker notions of merging have been studied by Kalai and Lehrer [1994], Lockett [1971], Diaconis and Freedman [1986, 1990], and D'Aristotile et al. [1988]. Their results prompt the question of whether agreement with respect to weak 1-randomness might be sufficient to achieve these other weaker types of merging.

We saw that, past level two, the notions of compatibility yielded by weak n -randomness coincide with absolute continuity, and that there is a hierarchy of compatibility notions induced by algorithmic randomness that are strictly stronger than absolute continuity. This raises the question of whether there is a stronger variant of absolute continuity that these strong forms of compatibility coincide with or are entailed by. Hoyrup and Rojas [2011] offer a first step in this direction. They define the following variant of absolute continuity, which they call *effective absolute continuity*: μ is effectively absolutely continuous with respect to ν if there is a computable function $\varphi : \mathbb{N} \rightarrow \mathbb{N}$ such that, for all $\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})$ and all $n \in \mathbb{N}$, $\nu(\mathcal{S}) < 2^{-\varphi(n)}$ entails that $\mu(\mathcal{S}) < 2^{-n}$ (this is indeed a form of absolute continuity because the standard version of absolute continuity can be defined in the same way, but dropping the computability requirement on φ). Then, they note that effective absolute continuity entails compatibility relative to Martin-Löf randomness, but not vice-versa. An immediate question prompted by this observation is where exactly effective absolute continuity fits within the hierarchy of compatibility notions induced by algorithmic randomness. More generally, effective absolute continuity and possible variants thereof are worth investigating further to better understand the relationship between (forms of) absolute continuity and the types of agreement generated by algorithmic randomness.

The above considerations also beg the question of whether there are any forms of merging stronger than the one introduced by Blackwell and Dubins that the new notions of compatibility yielded by algorithmic randomness entail. In other words, are there any types of merging for which absolute continuity is not enough, but which are nonetheless guaranteed to hold if the two agents agree on the relevant algorithmic randomness notions? In particular, are there any *effective* types of merging that are entailed by concurring on which data streams are algorithmically random?

While absolute continuity is arguably the most prominent type of compatibility in the literature on Bayesian merging of opinions, there are several alternatives to it, often studied in the context of weaker forms of merging (see, for instance, [Kalai and Lehrer, 1994] and [Lehrer and Smorodinsky, 1996]). Another immediate question is thus how these forms of compatibility relate to the randomness-based notions of agreement investigated here.

In addition to the type of compatibility displayed by the agents' priors, one might wonder whether there are any other aspects of a learning situation that can have an impact on whether merging will occur. For instance, both notions of merging of opinions considered in this chapter rely on the fact that the agents update their respective beliefs via Bayesian conditioning (more precisely, as explained in §2.1, using conditional probabilities defined with respect to the sub- σ -algebras in the filtration $\{\mathcal{F}_n\}_{n \in \mathbb{N}}$). The use of Bayesian conditioning, while standard, in itself reflects an important assumption about how learning proceeds: namely, that making an observation amounts to learning that an event has occurred for certain (in C. I. Lewis' words, "[i]f anything is to be probable then something must be certain" [Lewis, 1946, p. 186]). However, not all learning situations involve certain evidence. For instance, being told that some event \mathcal{S} has occurred by a source one does not deem fully reliable may increase one's belief in \mathcal{S} without convincing them that \mathcal{S} has happened for certain. The canonical alternative to Bayesian conditioning for dealing with uncertain evidence is Jeffrey's *probability kinematics* (or *Jeffrey conditioning*) [1957; 1965; 1968]. So, a natural question is whether merging-of-opinions results still hold when the evidential inputs are uncertain and the agents update their beliefs via Jeffrey conditioning.

This issue is addressed in [Huttegger, 2015a]. Huttegger considers two possible ways of precisifying the idea that, at each step of the learning process, the agents are faced with the same uncertain evidence: *hard Jeffrey shifts*, which are independent of the agents'

priors, and *soft Jeffrey shifts*, which instead rely on the agents' priors. Then, he proves that merging of opinions is still guaranteed to hold if the agents update their beliefs via Jeffrey conditioning with hard Jeffrey shifts, while updating via Jeffrey conditioning with soft Jeffrey shifts does not, in general, guarantee merging. This finding and the results in this chapter then prompt the following question: what happens to merging-of-opinions theorems if, in addition to using Jeffrey conditioning to update their beliefs, the agents begin the learning process with credences that are compatible relative to some algorithmic randomness notion? Does agreement on algorithmic randomness facilitate merging in this setting?

To conclude, recall that the Lebesgue Decomposition Theorem discussed earlier entails that, for any two probability measures μ and ν , there is some $\alpha \in [0, 1]$ such that μ can be decomposed as follows: $\mu = \alpha\mu_a + (1 - \alpha)\mu_o$, where $\mu_a \ll \nu$ and $\mu_o \perp \nu$. In view of our results, a natural question is whether there is a randomness analogue of the Lebesgue Decomposition Theorem. In other words, given some algorithmic randomness notion R , is it the case that, for any two computable measures μ and ν , there is some $\alpha \in [0, 1]$ such that μ can be decomposed as $\mu = \alpha\mu_a + (1 - \alpha)\mu_o$, where ν is R -compatible with μ_a (i.e., $\mu_a\text{-}R \subseteq \nu\text{-}R$) but R -incompatible with μ_o (i.e., $\mu_o\text{-}R \cap \nu\text{-}R = \emptyset$)?

2.A Appendix: converse implications

Bienvu and Merkle [2009] proved that (i) agreement on Martin-Löf randomness does not entail agreement on computable randomness, (ii) agreement on computable randomness does not entail agreement on Schnorr randomness, and (iii) absolute continuity entails neither agreement on Martin-Löf randomness nor agreement on Schnorr randomness. In this appendix we complete the picture by showing that none of the implications involving density randomness can be reversed either.

All of the counterexamples discussed below rely on the use of strictly positive measures. This is due to the fact that there is an exact correspondence between normed total dyadic martingales and strictly positive measures which also carries over to the computable setting:

Observation 2.A.1. *For every strictly positive measure μ , the normed total dyadic μ -martingales are exactly the functions of the form $\frac{\xi}{\mu}$, where ξ is a measure. For every*

computable strictly positive measure μ , the computable normed dyadic μ -martingales are exactly the functions of the form $\frac{\xi}{\mu}$, where ξ is a computable measure.

Proof. Let μ be a strictly positive measure and $d : 2^{<\mathbb{N}} \rightarrow \mathbb{R}^{\geq 0}$ a normed dyadic μ -martingale. For each $\sigma \in 2^{\mathbb{N}}$, let $\xi([\sigma]) = d(\sigma)\mu([\sigma])$. Then, by an argument analogous to the one provided in the proof of Lemma 2.2.5, ξ is a measure. Moreover, if μ and d are computable, then so is ξ .

For the other direction, take $\frac{\xi}{\mu}$, where μ and ξ are measures and μ is strictly positive. For all σ , define $d(\sigma)$ as $\frac{\xi([\sigma])}{\mu([\sigma])}$. Then, d is normed and non-negative. Moreover, $d(\sigma)\mu([\sigma]) = \xi([\sigma]) = \xi([\sigma 0]) + \xi([\sigma 1]) = d(\sigma 0)\mu([\sigma 0]) + d(\sigma 1)\mu([\sigma 1])$. So, d is a normed μ -martingale. If μ and ξ are computable, then so is d . \square

Now, given two computable strictly positive measures μ, ν and $k \in \mathbb{R}^{>0}$, let $\mathcal{S}_{\mu/\nu}^k = \left\{ \omega \in 2^{\mathbb{N}} : \sup_{n \in \mathbb{N}} \frac{\mu([\omega \upharpoonright n])}{\nu([\omega \upharpoonright n])} \geq k \right\}$. Then, define $\mathcal{S}_{\mu/\nu}$ as $\bigcap_{k \in \mathbb{N}} \mathcal{S}_{\mu/\nu}^k$. The set $\mathcal{S}_{\mu/\nu}$ consists of all ω with $\limsup_{n \in \mathbb{N}} \frac{\mu([\omega \upharpoonright n])}{\nu([\omega \upharpoonright n])} = \infty$. Since ν is strictly positive, we have that $\nu(\mathcal{S}_{\mu/\nu}) = 0$ by Ville's Theorem²¹ and Observation 2.A.1. Moreover, by the characterisation of computable randomness in terms of dyadic martingales and Observation 2.A.1, $\mathcal{S}_{\mu/\nu} \cap \nu\text{-CR} = \emptyset$. *A fortiori*, $\mathcal{S}_{\mu/\nu} \cap \nu\text{-MLR} = \emptyset$ and $\mathcal{S}_{\mu/\nu} \cap \nu\text{-DR} = \emptyset$.

Proposition 2.A.2. *Let μ and ν be computable strictly positive measures. We then have that $\mu\text{-DR} = \nu\text{-DR}$ if and only if $\mathcal{S}_{\mu/\nu} \cap \mu\text{-DR} = \mathcal{S}_{\nu/\mu} \cap \nu\text{-DR} = \emptyset$.*

Proof. The left-to-right direction follows from the fact that $\mathcal{S}_{\nu/\mu} \cap \mu\text{-DR} = \emptyset$ and the fact that $\mathcal{S}_{\mu/\nu} \cap \nu\text{-DR} = \emptyset$. For the right-to-left direction, suppose $\mathcal{S}_{\mu/\nu} \cap \mu\text{-DR} = \mathcal{S}_{\nu/\mu} \cap \nu\text{-DR} = \emptyset$, but that, without loss of generality, there is some $\omega \in \mu\text{-DR}$ such that $\omega \notin \nu\text{-DR}$. Then, there is a lower semi-computable semi-measure ξ such that either $\lim_{n \rightarrow \infty} \frac{\xi([\omega \upharpoonright n])}{\nu([\omega \upharpoonright n])}$ is infinite or the sequence $\left\{ \frac{\xi([\omega \upharpoonright n])}{\nu([\omega \upharpoonright n])} \right\}_{n \in \mathbb{N}}$ fails to have a limit. If $\lim_{n \rightarrow \infty} \frac{\xi([\omega \upharpoonright n])}{\nu([\omega \upharpoonright n])} = \infty$, it then follows that

$$\limsup_{n \rightarrow \infty} \frac{\mu([\omega \upharpoonright n])}{\nu([\omega \upharpoonright n])} = \limsup_{n \rightarrow \infty} \frac{\xi([\omega \upharpoonright n])}{\nu([\omega \upharpoonright n])} \bigg/ \frac{\xi([\omega \upharpoonright n])}{\mu([\omega \upharpoonright n])} = \infty.$$

Hence, $\omega \in \mathcal{S}_{\mu/\nu}$. Since $\mathcal{S}_{\mu/\nu} \cap \mu\text{-DR} = \emptyset$, $\omega \notin \mu\text{-DR}$, which is a contradiction. If, on the other hand, the sequence $\left\{ \frac{\xi([\omega \upharpoonright n])}{\nu([\omega \upharpoonright n])} \right\}_{n \in \mathbb{N}}$ does not have a limit, then the sequence $\left\{ \frac{\nu([\omega \upharpoonright n])}{\mu([\omega \upharpoonright n])} \right\}_{n \in \mathbb{N}}$

²¹ Ville's Theorem [Ville, 1939] establishes that the success set of a dyadic (super)martingale is a measure-zero set.

does not have a limit either, since $\frac{\nu([\omega]_n)}{\mu([\omega]_n)} = \frac{\xi([\omega]_n)}{\mu([\omega]_n)} / \frac{\xi([\omega]_n)}{\nu([\omega]_n)}$. Hence, $\omega \notin \mu\text{-DR}$, which is again a contradiction. \square

Note that Proposition 2.A.2 continues to hold if we replace $\mathcal{S}_{\mu/\nu}$ with $\mathcal{C}_{\mu/\nu} = \bigcap_{k \in \mathbb{N}} \mathcal{C}_{\mu/\nu}^k$, where $\mathcal{C}_{\mu/\nu}^k = \left\{ \omega \in 2^{\mathbb{N}} : \sup_{n \in \mathbb{N}} \frac{\mu([\omega]_{3^n})}{\nu([\omega]_{3^n})} \geq k \right\}$. We will appeal to Proposition 2.A.2 and to this latter observation in proving that none of the implications involving density randomness established in §2.2 can be reversed. We will also make use of the notion of a Δ_2^0 sequence, where $\omega \in 2^{\mathbb{N}}$ is said to be Δ_2^0 if there is a computable function $f : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$ such that, for all $n \in \mathbb{N}$, (i) $\omega(n) = \lim_{k \rightarrow \infty} f(n, k)$, and (ii) $f(n, k)$ is eventually constant (i.e., there are only finitely many distinct values among $f(n, 0), f(n, 1), f(n, 2), \dots$). If ω is Δ_2^0 , then it is computable in the halting problem. We are now ready to prove the following:

Proposition 2.A.3. (a) *There are computable measures μ and ν with $\mu\text{-DR} \subseteq \nu\text{-DR}$, but $\mu\text{-MLR} \not\subseteq \nu\text{-MLR}$.*

(b) *There are computable measures μ and ν with $\mu\text{-DR} \subseteq \nu\text{-DR}$, but $\mu\text{-SR} \not\subseteq \nu\text{-SR}$.*

(c) *There are computable measures μ and ν with $\mu \ll \nu$, but $\mu\text{-DR} \not\subseteq \nu\text{-DR}$.*

Proof. (a) Take the uniform measure λ . By Theorem 3.1 in [Day and Miller, 2015], there is a Δ_2^0 sequence $\omega \in \lambda\text{-MLR}$ that is a λ -dyadic positive density point, but not a λ -dyadic density-one point. Hence, $\omega \in \lambda\text{-MLR} \setminus \lambda\text{-DR}$. Since $\omega \in \lambda\text{-MLR}$, it follows that $\omega \in \lambda\text{-SR}$. Therefore, by Proposition 49 in [Bienvenu and Merkle, 2009], there is a strictly positive measure ν such that (i) $\omega \notin \nu\text{-SR}$, (ii) $\mathcal{C}_{\nu/\lambda} = \emptyset$, and (iii) $\mathcal{C}_{\lambda/\nu} = \{\omega\}$. By (i), we have that $\omega \notin \nu\text{-MLR}$, because $\nu\text{-MLR} \subseteq \nu\text{-SR}$. By (ii), it clearly follows that $\mathcal{C}_{\nu/\lambda} \cap \nu\text{-DR} = \emptyset$. By (iii) and the fact that $\omega \notin \lambda\text{-DR}$, on the other hand, we have that $\mathcal{C}_{\lambda/\nu} \cap \lambda\text{-DR} = \emptyset$. Hence, by Proposition 2.A.2, $\lambda\text{-DR} = \nu\text{-DR}$. Thus, $\lambda\text{-DR} \subseteq \nu\text{-DR}$, while $\lambda\text{-MLR} \not\subseteq \nu\text{-MLR}$.

(b) The proof of part (a) also establishes this claim.

(c) Let $\omega \in 2^{\mathbb{N}}$ be a Δ_2^0 sequence such that $\omega \in \lambda\text{-DR}$. The existence of such a sequence follows, e.g., from the fact that λ -density randomness is entailed by a randomness notion called λ -Demuth randomness, together with the fact that there are λ -Demuth random sequences (see, for instance, [Nies, 2009, Theorem 3.6.25] and Demuth [1982]). Then, $\omega \in \lambda\text{-SR}$. Once again, by Proposition 49 in [Bienvenu and Merkle, 2009], there is a strictly positive measure ν such that (i) $\omega \notin \nu\text{-SR}$, (ii) $\mathcal{C}_{\nu/\lambda} = \emptyset$, and (iii) $\mathcal{C}_{\lambda/\nu} = \{\omega\}$. By

(ii), $\nu(\mathcal{C}_{\nu/\lambda}) = 0$. Clearly, $\lambda(\{\omega\}) = 0$. Hence, (iii) entails that $\lambda(\mathcal{C}_{\lambda/\nu}) = 0$. Thus, by Proposition 47 in [Bienvenu and Merkle, 2009], λ and ν are mutually absolutely continuous (that is, $\lambda \ll \nu$ and $\nu \ll \lambda$). However, (i) implies that $\omega \notin \nu\text{-DR}$, since $\nu\text{-DR} \subseteq \nu\text{-SR}$. Hence, $\lambda \ll \nu$ while $\lambda\text{-DR} \not\subseteq \nu\text{-DR}$. \square

Chapter 3

Algorithmic randomness and convergence to the truth

But human opinion universally tends in the long run to a definite form, which is the truth. Let any human being have enough information and exert enough thought upon any question, and the result will be that he will arrive at a certain definite conclusion, which is the same that any other mind will reach under sufficiently favorable circumstances.

Peirce, *Review of Fraser's "The Works of George Berkeley"*

Together with merging-of-opinions theorems, convergence-to-the-truth results are a staple of Bayesian epistemology: their use in philosophy (especially in debates concerning the tenability of subjective Bayesianism) dates back to the work of Savage [1954]. In a nutshell, convergence-to-the-truth results show that, in a wide array of learning scenarios, Bayesian agents expect their future credences to almost surely converge to the truth as the evidence accumulates.

Bayesian convergence to the truth is epitomised by Lévy's Upward Martingale Convergence Theorem [1937] (Lévy's Upward Theorem, for short), which establishes that, given some quantity that a Bayesian agent is trying to estimate, the probability of observing a data stream that will lead the agent's successive estimates to asymptotically align with the truth is one. In other words, a Bayesian agent performing successive experiments to

estimate some quantity expects that almost every sequence of observations will bring about inductive success: eventual convergence to the truth.

While in and of itself significant, convergence to the truth with probability one remains a somewhat elusive notion. In its classical form, Lévy's Upward Theorem does not specify which data streams belong to the probability-one set of sequences on which convergence to the truth occurs. It does not reveal how the composition of this set changes depending on the particular quantity that the Bayesian agent is trying to estimate, nor does it indicate whether the data streams that ensure eventual convergence to the truth share any property that might explain their conduciveness to successful learning, that sets them apart from the data streams along which learning fails. Thus, a natural question raised by Lévy's classical result is whether the kinds of data streams that are conducive to learning for Bayesian agents are uniformly characterisable in an informative way.

The results presented here provide an answer to this question. Just as in Chapter 2, the driving idea behind this chapter is to approach the phenomenon of convergence to the truth from the perspective of computability theory and, in particular, the theory of algorithmic randomness. The classical version of Lévy's Upward Theorem is in fact very general: it does not impose any restrictions on the kind of prior probability distributions that a Bayesian agent can begin the learning process with, and the only constraint imposed on the quantities to be estimated is that they be integrable random variables. This high degree of generality is precisely the reason why Lévy's Upward Theorem, in its classical form, does not afford a precise characterisation of the collection of truth-conducive data streams. Here, on the other hand, we restrict attention to Bayesian agents whose subjective priors are computable probability measures and whose goal is estimating quantities that can be effectively approximated. As argued in Chapter 2, these are natural restrictions to impose when studying the inductive performance of more realistic, computationally limited learners. Additionally, as we shall see, they allow to provide a more fine-grained analysis of the phenomenon of Bayesian convergence to the truth.

More specifically, this chapter is devoted to proving several effective versions of Lévy's Upward Theorem, each depending on a different class of effective integrable random variables. The main upshot of these results is that, in this effective setting, a certain type of convergence to the truth holds if and only if the observed data stream is algorithmically

random. We thus have a positive answer to our initial question: for computable Bayesian agents whose goal is estimating the values of effectively approximable random variables, the truth-conducive data streams can indeed be characterised in a uniform way—they are exactly the algorithmically random ones. More precisely, different natural effectivity constraints imposed on random variables result in different algorithmic randomness notions corresponding to the collection of data streams on which Lévy’s Upward Theorem holds. The weaker the effectivity constraints, the stronger the randomness notion; thus, as the random variables corresponding to the inductive problems faced by the Bayesian agent become harder to compute, the collection of data streams on which successful learning occurs shrinks in a fully specifiable way.

While both this chapter and Chapter 2 appeal to computability theory and algorithmic randomness to clarify issues of epistemological interest, the results established here are of an importantly different nature. In Chapter 2, algorithmic randomness was used to define notions of compatibility between computable priors, which were in turn shown to imply asymptotic merging of opinions. In this chapter, on the other hand, we show that, in the effective setting, algorithmic randomness naturally emerges as the property that characterises the truth-conducive data streams—the data streams along which an agent’s credences eventually converge to the correct hypotheses. Thus, the provided effectivisations of Lévy’s Upward Theorem bridge the theory of algorithmic randomness and the literature on Bayesian convergence to the truth, and they attest, from a different perspective, to the utility of algorithmic randomness in the study of probabilistic inference.¹

These effectivisations also yield characterisations of some standard algorithmic randomness notions. In this sense, this work continues (and is indebted to) a recent line of research in computability theory which aims at characterising algorithmic randomness concepts in terms of effective versions of classical almost-everywhere theorems from analysis and probability, such as the Lebesgue Differentiation Theorem, the Lebesgue Density Theorem, or Birkhoff’s Ergodic Theorem.² Our work is also importantly related to that of Gaifman

¹In the conclusion of this chapter, we will see how the results in Chapter 2 and the ones presented here may be combined.

²For some papers exploring the connections between algorithmic randomness and computable analysis, see, for instance, [V’yugin, 1998], [Bienvenu et al., 2012a], [Franklin et al., 2012], [Rute, 2012], [Miyabe, 2013b], [Bienvenu et al., 2014], [Franklin and Towsner, 2014], [Freer et al., 2014], [Pathak et al., 2014], [Brattka et al., 2016], [Miyabe et al., 2016], and [Rute, 2018].

and Snir [1982], who first drew the connection between randomness and Bayesian learning in our sense. A discussion of Gaifman and Snir’s framework and of how our results differ from theirs can be found in §3.3.2.

The structure of the remainder of this chapter is as follows. In §3.1, we present the classical version of Lévy’s Upward Theorem and discuss its epistemic significance. The main results are presented in §3.2. We consider in turn various natural classes of effective random variables and show that the corresponding effectivisations of Lévy’s Upward Theorem each yield a characterisation of some standard algorithmic randomness notion. These results therefore establish that, in the Bayesian setting, there is a robust correspondence between algorithmic randomness and truth-conduciveness. In §3.3, we conclude by offering a discussion of the results and their philosophical consequences.

3.1 Convergence to the truth

Just as in Chapter 2, a learning scenario is given by a probability space $(\Omega, \mathcal{E}, \mu)$, with Ω the sample space, \mathcal{E} a σ -algebra on Ω , and μ a probability measure on \mathcal{E} . Unless otherwise specified, we take μ to represent the subjective prior of a Bayesian agent: namely, the agent’s initial degrees of belief, or credences, about the events in \mathcal{E} , which encapsulate their background knowledge and assumptions before any observation has been made.

In keeping with much of the Bayesian epistemology literature on the topic (see, for instance, [Earman, 1992], [Belot, 2013], and [Huttegger, 2015b, 2017]), we shall discuss Bayesian convergence to the truth in the context of the Cantor space of infinite binary sequences. Once again, we will think of such sequences as data streams, environments, possible worlds, or possible states of the world. Bayesian agents will be identified with their respective subjective priors—in this case, with (probability) measures over the Borel σ -algebra $\mathcal{B}(2^{\mathbb{N}})$ on $2^{\mathbb{N}}$. The inductive problems faced by Bayesian agents, on the other hand, will be modelled as random variables—more precisely, given a measure μ over $\mathcal{B}(2^{\mathbb{N}})$, as μ -measurable functions of the form $f : \subseteq 2^{\mathbb{N}} \rightarrow \mathbb{R}$ that are defined on a μ -measure-one set of sequences. Intuitively, the values of these random variables represent quantities that a Bayesian agent is interested in estimating by performing repeated experiments, and they crucially depend on which data stream is observed as a result of such experiments.

3.1.1 The classical version of Lévy’s Upward Theorem

Bayesian convergence-to-the-truth theorems follow from a fundamental result in probability theory: Doob’s Martingale Convergence Theorem [1953]. Recall that, in the general setting,³ a (discrete-time) martingale is an infinite sequence of random variables, where, for each n , the conditional expectation of the $(n + 1)$ -st random variable given the previous n random variables equals the value of the n -th random variable. Loosely put, Doob’s Martingale Convergence Theorem states that, as long as certain technical conditions are met, the limit of a martingale exists and is finite with probability one. A martingale can be seen as representing the evolution of a gambler’s capital through an infinite sequence of fair gambles, where fairness is ensured by the fact that, at each stage, the gambler’s expected capital at the following stage is exactly the same as their current capital. So, Doob’s Martingale Convergence Theorem implies that a gambler will almost surely fail to win an unbounded amount of capital.

From the perspective of Bayesian epistemology, there is an especially salient class of martingales (also known as *Lévy martingales*) that goes back to Lévy [1937]: namely, the conditional expectations of integrable random variables. Recall that, given a measure μ on $\mathcal{B}(2^{\mathbb{N}})$ and a random variable $f : \subseteq 2^{\mathbb{N}} \rightarrow \mathbb{R}$ that is defined μ -almost everywhere, $\mathbb{E}_{\mu}[f]$ denotes the (unconditional) expectation of f with respect to μ (that is, the average value of f weighted by μ). A random variable f is integrable, or L^1 , if $\mathbb{E}_{\mu}[|f|] < \infty$.

As in Chapter 2, we restrict attention to conditional expectations defined with respect to the filtration $\{\mathcal{F}_n\}_{n \in \mathbb{N}}$ —where, for each n , \mathcal{F}_n is the sub- σ -algebra of $\mathcal{B}(2^{\mathbb{N}})$ yielded by the cylinders $[\sigma]$ generated by strings $\sigma \in 2^{<\mathbb{N}}$ of length n . We do so because, in our setting, this collection of algebras has an especially natural epistemic interpretation: each \mathcal{F}_n intuitively captures the possible information that the Bayesian agent may obtain at the n -th stage of the learning process.

The conditional expectation $\mathbb{E}_{\mu}[f \mid \mathcal{F}_n] : 2^{\mathbb{N}} \rightarrow \mathbb{R}$ of a random variable f given \mathcal{F}_n is itself a random variable that, on input $\omega \in 2^{\mathbb{N}}$, returns the best estimate of f ’s value conditional on the first n digits $\omega \upharpoonright n$ of ω from the perspective of the Bayesian agent with prior μ . More suggestively, when ω is the true state of the world, $\mathbb{E}_{\mu}[f \mid \mathcal{F}_n](\omega)$ encodes

³As opposed to the special case of algorithmic randomness, where the focus is on dyadic martingales (see Chapter 1, §1.3).

the agent's beliefs regarding the true value of f (namely, $f(\omega)$) after having observed the outcomes $\omega \upharpoonright n = \omega(0)\dots\omega(n-1)$ of the first n experiments.

We use throughout the following version of the conditional expectation (since it is only almost surely unique): for all $\omega \in 2^{\mathbb{N}}$,

$$\mathbb{E}_{\mu}[f \mid \mathcal{F}_n](\omega) = \begin{cases} \frac{1}{\mu([\omega \upharpoonright n])} \int_{[\omega \upharpoonright n]} f d\mu & \text{if } \mu([\omega \upharpoonright n]) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Then, $\{\mathbb{E}_{\mu}[f \mid \mathcal{F}_n]\}_{n \in \mathbb{N}}$ is a martingale, since, by the tower property of conditional expectations,

$$\mathbb{E}_{\mu}[\mathbb{E}_{\mu}[f \mid \mathcal{F}_{n+1}] \mid \mathcal{F}_n] = \mathbb{E}_{\mu}[f \mid \mathcal{F}_n].$$

Given two integrable functions f and g , $\|f - g\|_1$ denotes $\int_{2^{\mathbb{N}}} |f - g| d\mu$. A sequence of functions $\{f_n\}_{n \in \mathbb{N}}$ converges to an integrable function f in the L^1 -norm (which we abbreviate as $f_n \rightarrow f$ in L^1) if $\lim_{n \rightarrow \infty} \|f_n - f\|_1 = 0$. Now, Lévy's Upward Martingale Convergence Theorem (also known as Lévy's 0-1 Law) is the following result:⁴

Theorem 3.1.1 (Lévy's Upward Theorem, Lévy [1937]). *Let $f : \subseteq 2^{\mathbb{N}} \rightarrow \mathbb{R}$ be an integrable random variable that is defined μ -almost everywhere. Then,*

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mu}[f \mid \mathcal{F}_n] = f$$

μ -almost everywhere. Moreover, $\mathbb{E}_{\mu}[f \mid \mathcal{F}_n] \rightarrow f$ in L^1 .

Let us call the set of sequences $\{\omega \in 2^{\mathbb{N}} : \lim_{n \rightarrow \infty} \mathbb{E}_{\mu}[f \mid \mathcal{F}_n](\omega) = f(\omega)\}$ along which convergence occurs the *success set* of agent μ with respect to f , and its complement the *failure set*. As mentioned above, we can think of f as a quantity that the agent is trying to estimate—for instance, f could record the value of some unknown parameter which may vary between possible worlds (or different possible states of the world). Then, Lévy's Upward Theorem says that if this quantity can be modelled as an integrable random variable, then, from the agent's viewpoint, their failure set is negligible. In other words, the agent believes that they will eventually converge to the truth about the value of f with

⁴See, for instance, [Williams, 1991, §14.2].

probability one: $\mu(\{\omega \in 2^{\mathbb{N}} : \lim_{n \rightarrow \infty} \mathbb{E}_{\mu}[f \mid \mathcal{F}_n](\omega) = f(\omega)\}) = 1$.

A special case that is often the focus of discussions on convergence to the truth in the philosophy literature is when the integrable random variable to be estimated is the indicator function $\chi_{\mathcal{A}}$ of some measurable subset \mathcal{A} of Cantor space.⁵ This corresponds to the case where the inductive problem that the Bayesian agent is facing is a binary decision problem: does the true world, corresponding to the observed data stream, belong to \mathcal{A} ? Or, put differently, does the true world possess the property encoded by \mathcal{A} ? In this setting, the quantity that the agent is trying to estimate is thus the truth value of \mathcal{A} , and learning almost surely proceeds by Bayesian conditioning, since, whenever $\mu([\omega \upharpoonright n]) > 0$,

$$\mathbb{E}_{\mu}[\chi_{\mathcal{A}} \mid \mathcal{F}_n](\omega) = \frac{1}{\mu([\omega \upharpoonright n])} \int_{[\omega \upharpoonright n]} \chi_{\mathcal{A}} d\mu = \frac{\mu(\mathcal{A} \cap [\omega \upharpoonright n])}{\mu([\omega \upharpoonright n])} = \mu(\mathcal{A} \mid [\omega \upharpoonright n]).$$

So, Lévy's Upward Theorem entails that $\lim_{n \rightarrow \infty} \mu(\mathcal{A} \mid [\omega \upharpoonright n]) = \chi_{\mathcal{A}}(\omega)$ for μ -almost every $\omega \in 2^{\mathbb{N}}$. Hence, a Bayesian agent with prior μ expects their beliefs, given by the above sequence of posterior probabilities, to converge almost surely to the truth about whether \mathcal{A} is the case with increasing information. In other words, the agent is essentially certain that they will be able to eventually settle the question of membership in \mathcal{A} .

3.1.2 The epistemic significance of Lévy's Upward Theorem

Given a cursory reading of the above discussion, it might be tempting to interpret Lévy's Upward Theorem as providing a vindication of the subjectivist (or personalist) brand of Bayesianism,⁶ according to which there are no rationality constraints that can be reasonably imposed on an agent's initial credences besides having to comply with the laws of the probability calculus. After all, does the theorem not establish that prior probability distributions do not matter in the long run? Does it not ensure that subjective priors are eventually swamped by the cumulating evidence, so that different agents starting off the learning process with divergent beliefs are nonetheless all guaranteed to converge to the truth in the limit, provided that they make the same observations and correctly update their credences on the basis of the new evidence?

⁵That is, $\chi_{\mathcal{A}}$ is the function given by $\chi_{\mathcal{A}}(\omega) = 1$ if $\omega \in \mathcal{A}$ and $\chi_{\mathcal{A}}(\omega) = 0$ if $\omega \notin \mathcal{A}$.

⁶A variety of Bayesianism championed, for instance, by Ramsey [1931], de Finetti [1937], Savage [1954], and Jeffrey [1977].

As already evinced by the above description of the result, the main problem with this reading is that the almost-sure convergence to the truth achieved via Lévy’s Upward Theorem is always relative to the agent’s prior: that is, to their initial beliefs. The precise sense in which an agent with prior μ converges to the truth almost surely about a given hypothesis is that the μ -measure of their success set with respect to that hypothesis is one. This means that the agent, from their own perspective, is basically *ex ante* (that is, before performing any experiments) certain that, with increasing information, their beliefs will eventually converge to the truth. Yet, there is no objective, external guarantee that this will indeed be the case.

Thus, as noted by many authors,⁷ Lévy’s Upward Theorem does not establish the universal reliability of Bayesian learning methods from an objective, third-person standpoint. Its epistemic significance stems from within, for it establishes that a certain kind of scepticism about induction is impossible: if an agent is independently committed to probabilistic coherence (synchronic and diachronic), then, by Lévy’s Upward Theorem, that agent cannot be a sceptic about the possibility of learning from experience. The agent’s independent commitment to the Bayesian framework as a way of modelling their uncertainty implies that, *by their own light*, their recourse to inductive reasoning is justified. As observed by Skyrms, from the perspective of the Bayesian agent, it is “inappropriate for you to ask the standard question, “Why should I believe that the real situation is not in that set of measure zero?” The measure in question *is* your degree of belief. You *do* believe that the real situation is not in that set, with degree of belief one” [Skyrms, 1984, p. 62].

The epistemic significance of Lévy’s Upward Theorem may also be appraised through the prism of inter-subjective agreement, since it can also be viewed as a merging-of-opinions result (though one where the type of merging involved, as we will now see, is rather weak). Recall that, given two measures μ and ν on $\mathcal{B}(2^{\mathbb{N}})$, μ is absolutely continuous with respect to ν ($\mu \ll \nu$) if, for every event $\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})$, $\nu(\mathcal{S}) = 0$ implies that $\mu(\mathcal{S}) = 0$.⁸ Take two measures μ and ν with $\mu \ll \nu$, and let $f : \subseteq 2^{\mathbb{N}} \rightarrow \mathbb{R}$ be an almost-everywhere defined integrable random variable relative to the probability spaces $(2^{\mathbb{N}}, \mathcal{B}(2^{\mathbb{N}}), \mu)$ and $(2^{\mathbb{N}}, \mathcal{B}(2^{\mathbb{N}}), \nu)$. Let $\mathbb{E}_{\mu}[f \mid \mathcal{F}_n]$ denote the conditional expectation of f with respect to

⁷See, for instance, [Glymour, 1980], [Earman, 1992], [Kelly, 1996], and [Belot, 2013].

⁸See Definition 2.1.1 in Chapter 2 and the ensuing discussion.

μ and $\mathbb{E}_\nu[f \mid \mathcal{F}_n]$ its conditional expectation with respect to ν . Then, Lévy's Upward Theorem entails that $\lim_{n \rightarrow \infty} \mathbb{E}_\mu[f \mid \mathcal{F}_n] = f$ μ -almost everywhere and $\lim_{n \rightarrow \infty} \mathbb{E}_\nu[f \mid \mathcal{F}_n] = f$ ν -almost everywhere. Moreover, since μ is absolutely continuous with respect to ν ,

$$\lim_{n \rightarrow \infty} \mathbb{E}_\nu[f \mid \mathcal{F}_n] = f \text{ } \mu\text{-almost everywhere.} \quad (\star)$$

This means that the collection of data streams along which both $\lim_{n \rightarrow \infty} \mathbb{E}_\mu[f \mid \mathcal{F}_n] = f$ and $\lim_{n \rightarrow \infty} \mathbb{E}_\nu[f \mid \mathcal{F}_n] = f$ hold has μ -measure one. Therefore, on this μ -measure-one set, we have that ν merges with μ in the following sense: for each data streams ω in this set and for every $\epsilon > 0$, there is some N such that, for all $n > N$, $|\mathbb{E}_\mu[f \mid \mathcal{F}_n](\omega) - \mathbb{E}_\nu[f \mid \mathcal{F}_n](\omega)| < \epsilon$.

If μ and ν are both subjective priors, this type of merging means that, after performing N experiments, the two agents' respective estimates of the value of f will differ by less than ϵ ; hence, in the limit, their estimates—and, thus, their beliefs—will coincide. As already remarked by Earman [1992], it should however be pointed out that, compared to other merging-of-opinions results (such as the seminal Blackwell-Dubins Theorem [1962]), Lévy's Upward Theorem yields a rather weak form of merging. This is because the type of convergence attained is not at all uniform: the value of N in fact depends on ν, μ, f, ω , as well as ϵ .

As seen in Chapter 2, merging-of-opinions theorems can also be taken to establish a certain type of convergence to the truth. This is because, besides representing subjective priors, probability measures can also represent objective chance distributions. One may thus view one of the two measures involved in merging-of-opinions results as representing the true distribution governing some phenomenon, while the other one is the subjective prior of a Bayesian agent. From this perspective, these theorems establish that, as long as the agent's subjective prior is sufficiently compatible with the truth, their beliefs will almost surely align with the true measure—where the achievable almost-sure convergence is with respect to the true measure. Viewing Lévy's Upward Theorem as a merging-of-opinions result where one of the two measures is a chance distribution offers a limited vindication of the intuition with which we opened this discussion: namely, that Lévy's Upward Theorem may be seen as establishing that priors, and where they come from, should not be a cause for concern. This is because, even though, as argued above, the type of merging yielded by

Lévy’s Upward Theorem is weak, it nonetheless implies that priors will be asymptotically swamped by the evidence, as long as they are absolutely continuous with respect to the objectively true distribution.

We conclude our discussion of the philosophical import of Lévy’s Upward Theorem by noting that Condition (\star) above is significant in itself. In cases where μ can be reasonably interpreted as an objective chance distribution, Condition (\star) tells us that convergence to the truth, in the original sense of Lévy’s Upward Theorem, occurs with objective probability one. Hence, in this setting, Lévy’s Upward Theorem does provide an external guarantee that convergence to the truth is bound to almost surely occur.

3.2 Effectivising Lévy’s Upward Theorem

As noted earlier, Lévy’s Upward Theorem is very general: it applies to any probability measure and any integrable random variable. However, while establishing that the collection of data streams along which convergence to the truth occurs is “measure-theoretically large” (that is, convergence to the truth happens with probability one), the theorem does not offer any indication as to what this success set might look like. Yet, it is in natural to wonder whether it is possible to pinpoint any specific data streams that guarantee convergence to the truth, and what a data stream must be like in order to belong to the success set of a Bayesian learner. In particular, are there any properties that *all* the data streams that guarantee convergence to the truth share? Is it possible to single out some property of data streams that is responsible for leading to convergence to the truth?

In what follows, we will address these questions from the perspective of computability theory, obtaining in this way various effective versions of Lévy’s Upward Theorem. The computability-theoretic constraints imposed throughout the chapter essentially amount to restricting attention to computationally limited learners. The classical versions of the theorems that play a central role in Bayesian epistemology, including Lévy’s Upward Theorem, apply to ideal agents that do not suffer from any limitations: the agents’ subjective priors can be arbitrarily complex, and so can the inductive problems that they have to solve. As already discussed in Chapter 2, computability theory offers a natural framework for modelling less-than-ideal agents, whose computational power does not exceed that of a Turing

machine: a computationally limited, or computable, Bayesian agent is one whose subjective prior is a computable measure (see Definition 1.1.1). In addition, in the computability-theoretic setting, one can provide a more fine-grained analysis of the inductive problems faced by the learners. In particular, as we go along, we will impose various effectivity constraints on the random variables that the Bayesian agents are trying to estimate. The aim of these constraints is to specify the extent to which the quantities to be estimated are effectively approximable. The more difficult to compute a random variable is, the more complex the corresponding inductive problem. This classification, as we shall see, allows to clarify how the learning performance of a computable Bayesian agent varies as a function of the complexity of the inductive problem that they have to solve.

Besides computability, some of the results in this chapter also rely on a second property of priors: strict positivity.⁹ Strict positivity is often assumed in the Bayesian epistemology literature, since strictly positive priors have a natural epistemic interpretation: they intuitively capture the beliefs of agents who are maximally open-minded with respect to the evidence, in that they do not *a priori* exclude any finite sequence of observations. From this perspective, if a result holds for all computable strictly positive measures, then it means that it applies to all computationally limited Bayesian agents that are evidentially open-minded.¹⁰

We do not take strict positivity to be a rationality requirement on priors, and it should be noted that, while intuitive, strict positivity is sometimes criticised in the literature (see, for instance, [Belot, 2013]). The type of open-mindedness encoded by strict positivity is in fact compatible with various kinds of closed-mindedness. For instance, consider the uniform measure λ , which is strictly positive: λ assigns probability zero to many events, including to the collection of data streams that are eventually zero; hence, although open-minded with respect to all finite sequences of observations, the uniform measure is closed-minded with respect to the possibility of observing only finitely many ones (and this is just one possible example among many). This is however a feature more than a bug, as no prior can be open-minded with respect to every event in $\mathcal{B}(2^{\mathbb{N}})$: in order for this to be the

⁹Recall that a measure is strictly positive if it assigns positive probability to all basic open sets (in our setting, to all cylinders).

¹⁰Most of the characterisation results proven below that rely on strict positivity can in fact be generalised to arbitrary computable measures. For these more general results, see [Huttegger et al., 2021].

case, a prior would have to assign positive probability to every singleton $\{\omega\}$, which is not possible. As noted by Huttegger, “[t]he open-mindedness of a prior with respect to a set is not a maximally open state of mind that doesn’t rule out any possibilities; it rather represents a state of mind that is committed to some possibilities at the expense of others. Open-mindedness with respect to one set implies closed-mindedness with respect to others” [Huttegger, 2015b, p. 593], and not all closed-mindedness is unreasonable.

The main findings from this chapter may be summarised as follows. In this effective setting, it is indeed possible to single out a property of data streams that guarantees that a Bayesian agent’s beliefs will eventually align with the truth: this property is being algorithmically random. In particular, in what follows, we will consider in succession various natural effectivity constraints on random variables, and we will show that, for computable (strictly positive) measures, the resulting effectivisations of Lévy’s Upward Theorem provide characterisations of several distinct algorithmic randomness notions.^{11,12} In other words, in each case, the algorithmically random data streams will be shown to correspond to the collection of sequences on which the relevant effectivisation of Lévy’s Upward Theorem holds.

3.2.1 L^1 -computable functions

We begin by considering L^1 -computable functions, first introduced by Pour-El and Richards [1989], which play an important role in computable analysis. According to Weihrauch’s definition [2000] of computability for functions on the real numbers (and related sets), all computable functions are continuous. Yet, some discontinuous functions are so simple that it seems they should count as computable in some sense. For an example of a very simple discontinuous function, take the function $f : 2^{\mathbb{N}} \rightarrow \mathbb{R}$ given by $f(\omega) = 1$ if ω is the constant-one sequence $\bar{1}$, and $f(\omega) = 0$ otherwise—that is, f is the indicator function of the singleton $\{\bar{1}\}$. The concept of L^1 -computability is meant to remedy this problem.

Just as the computable reals are the ones that can be computably approximated via

¹¹More precisely, the forward direction of all of our characterisation results will be shown to hold for all computable measures. In two instances, the backward direction will be shown to hold provided that the underlying measure is strictly positive, in addition to being computable, and, in two other instances, the backward direction will be given in the context of the uniform measure.

¹²Again, see [Huttegger et al., 2021] for the most recent versions of these results.

computable sequences of rational numbers,¹³ the L^1 -computable functions are those that can be computably approximated via computable sequences of appropriately simple functions:

Definition 3.2.1 (Rational-valued step function). *A rational-valued step function is a function $f : 2^{\mathbb{N}} \rightarrow \mathbb{R}$ of the form $f = \sum_{i=1}^k q_i \chi_{[\tau_i]}$, where $q_1, \dots, q_k \in \mathbb{Q}$ and $\tau_1, \dots, \tau_k \in 2^{<\mathbb{N}}$. Without loss of generality, the strings τ_1, \dots, τ_k can be assumed to be pairwise disjoint.*

Then, a function is L^1 -computable if it can be approximated at a computable rate via a computable sequence of rational-valued step functions in the following sense:

Definition 3.2.2 (L^1 -computable function). *A function $f : \subseteq 2^{\mathbb{N}} \rightarrow \mathbb{R}$ is L^1 -computable, relative to a computable measure μ , if there is a computable sequence $\{f_n\}_{n \in \mathbb{N}}$ of rational-valued step functions such that, for all $n \in \mathbb{N}$,*

$$\|f_n - f\|_1 = \int_{2^{\mathbb{N}}} |f_n - f| d\mu \leq 2^{-n}.$$

The sequence $\{f_n\}_{n \in \mathbb{N}}$ is said to be a witness to the L^1 -computability of f .

A computable sequence $\{f_n\}_{n \in \mathbb{N}}$ of rational-valued step functions is said to be a *fast computable L^1 -Cauchy sequence* relative to μ if, for all $n, m \in \mathbb{N}$ with $m \geq n$, $\|f_n - f_m\|_1 \leq 2^{-n}$. The following proposition offers an alternative characterisation of L^1 -computability which will later turn out to be useful (from now on, we will omit reference to the underlying computable measure μ when no ambiguity arises):¹⁴

Proposition 3.2.3. (i) *Suppose that $\{f_n\}_{n \in \mathbb{N}}$ is a fast computable L^1 -Cauchy sequence such that $f_n \rightarrow f$ in L^1 . Then $\{f_{n+1}\}_{n \in \mathbb{N}}$ is a witness to the L^1 -computability of f .*

(ii) *Suppose that f is L^1 -computable with witness $\{f_n\}_{n \in \mathbb{N}}$. Then $\{f_{n+1}\}_{n \in \mathbb{N}}$ is a fast computable L^1 -Cauchy sequence with $f_{n+1} \rightarrow f$ in L^1 .*

Recall the indicator function $f : 2^{\mathbb{N}} \rightarrow \mathbb{R}$ of the singleton set $\{\bar{1}\}$ mentioned above. Fix the uniform measure λ . We will see that, in spite of being discontinuous (and, so, uncomputable, given Weihrauch's definition), this function is L^1 -computable. For each

¹³See Chapter 1, §1.1.

¹⁴For a proof of Proposition 3.2.3, see [Huttegger et al., 2021].

$n \in \mathbb{N}$ and $\omega \in 2^{\mathbb{N}}$, let $f_n(\omega) = 1$ if $\omega(k) = 1$ for all $0 \leq k < n$, and $f_n(\omega) = 0$ otherwise. In other words, f_n is the indicator function of the cylinder $[1^n]$ generated by n consecutive 1's. Then, the f_n 's form a computable sequence of rational-valued step functions. Moreover, for each n ,

$$\int_{2^{\mathbb{N}}} |f_n - f| d\lambda = \int_{2^{\mathbb{N}}} f_n d\lambda - \int_{2^{\mathbb{N}}} f d\lambda = 2^{-n},$$

so the f_n 's are indeed a witness to the L^1 -computability of f .

Now, what happens to Lévy's Upward Theorem if, rather than considering arbitrary integrable random variables, one focuses instead on L^1 -computable random variables? Is there a property of data streams that characterises exactly those sequences that guarantee convergence to the truth in the sense of Lévy's Upward Theorem when the quantity to be estimated is L^1 -computable? The result below establishes that there is indeed such a property, and that this property corresponds to a central algorithmic randomness notion: Schnorr randomness. Before we can state the result, we need some notation: given an L^1 -computable function $f : \subseteq 2^{\mathbb{N}} \rightarrow \mathbb{R}$ with witness $\{f_n\}_{n \in \mathbb{N}}$, define $\hat{f} : \subseteq 2^{\mathbb{N}} \rightarrow \mathbb{R}$ as $\hat{f}(\omega) = \lim_{n \rightarrow \infty} f_n(\omega)$ if such limit exists and is finite; otherwise, let $\hat{f}(\omega)$ be undefined. We then have the following:^{15,16}

Theorem 3.2.4. *Let μ be a computable strictly positive measure and $\omega \in 2^{\mathbb{N}}$. Then, the following are equivalent:*

- (1) ω is μ -Schnorr random;
- (2) for all L^1 -computable functions $f : \subseteq 2^{\mathbb{N}} \rightarrow \mathbb{R}$ with witness $\{f_n\}_{n \in \mathbb{N}}$, $\hat{f}(\omega)$ is defined, and

$$\lim_{k \rightarrow \infty} \mathbb{E}_{\mu}[f \mid \mathcal{F}_k](\omega) = \hat{f}(\omega).$$

As a matter of fact, the forward direction of Theorem 3.2.4 holds for all computable

¹⁵For a proof of (the most recent version of) this result, see [Huttegger et al., 2021].

¹⁶Theorem 3.2.4 is the Cantor space analogue of the characterisation of Schnorr randomness by Pathak et al. [2014] in terms of the Lebesgue Differentiation Theorem and the uniform measure in Euclidean space. Theorem 3.2.4 is also related to, yet distinct from, some similar characterisations of Schnorr randomness due to Rute [2012]. In particular, Rute independently proved the forward direction of Theorem 3.2.4. He, however, did not prove an exact analogue of our backward direction. In particular, as opposed to Rute's results, the backward direction of the proof of Theorem 3.2.4 establishes that if a sequence ω fails to be μ -Schnorr random, then there is an L^1 -computable function for which the limit of the conditional expectations does not exist along ω (and, so, convergence to the truth fails to occur on ω). See [Huttegger et al., 2021] for a more detailed discussion of the relation between Theorem 3.2.4 and Rute's results.

measures, not just the strictly positive ones. This is because, for any computable measure μ , ω being μ -Schnorr random entails that $\mu([\omega \upharpoonright k]) > 0$ for all k : if there were some k such that $\mu([\omega \upharpoonright k]) = 0$, then the sequence $\{\mathcal{U}_n\}_{n \in \mathbb{N}}$, where $\mathcal{U}_n = [\omega \upharpoonright k]$ for all n , would be a sequential μ -Schnorr test that ω would fail. As a result, ω being μ -Schnorr random by itself ensures that $\mathbb{E}_\mu[f \mid \mathcal{F}_k](\omega) = \frac{1}{\mu([\omega \upharpoonright k])} \int_{[\omega \upharpoonright k]} f d\mu$ for all k , which, in turn, allows to do away with the strict positivity assumption.

Theorem 3.2.4 has a natural epistemic interpretation: the Schnorr random data streams are exactly the sequences of observations along which a computable (open-minded) Bayesian agent's beliefs—in the form of their best estimates of the true value of an L^1 -computable random variable—asymptotically align with the truth. So, the collection of Schnorr random sequences coincides with the collection of data streams where a specific type of inductive success is attained.

Note that this result does not establish that, given a fixed L^1 -computable random variable, it is impossible for a data stream that is not μ -Schnorr random to be truth-conducive. Take once again the L^1 -computable function f (relative to λ) that outputs 1 when given as input the constant-one sequence and 0 otherwise. Let $\bar{0}$ denote the constant-zero sequence. For all n and k , $f_n(\bar{0}) = 0$ and $\mathbb{E}_\lambda[f \mid \mathcal{F}_k](\bar{0}) = 0$. We therefore have that $\lim_{k \rightarrow \infty} \mathbb{E}_\lambda[f \mid \mathcal{F}_k](\bar{0}) = \lim_{n \rightarrow \infty} f_n(\bar{0}) = \hat{f}(\bar{0}) = f(\bar{0}) = 0$: the sequence $\bar{0}$ is truth-conducive. Yet, $\bar{0}$ not only fails to be λ -Schnorr random, but it is also a computable data stream—and, thus, it is highly non-random. What Theorem 3.2.4 does establish, however, is that if the observed data stream is μ -Schnorr random, then convergence to the truth is guaranteed for all L^1 -computable inductive problems; conversely, if the observed data stream ω fails to be μ -Schnorr random, then there is at least one L^1 -computable inductive problem such that the Bayesian agent's successive estimates do not align with the truth along ω . Hence, μ -Schnorr randomness characterises the collection of data streams along which inductive success in the sense of Lévy's Upward Theorem is attained across all L^1 -computable problems.

This result bridges the theory of algorithmic randomness and Bayesian learning: in particular, it shows that algorithmic randomness, when taken to be a property of data streams, allows to offer a more informative analysis of the phenomenon of Bayesian convergence to the truth for computationally limited learners. In addition, Theorem 3.2.4 immediately raises the question of whether the correspondence between the data streams

along which convergence to the truth occurs and the algorithmically random data streams survives when one imposes other natural effectivity constraints on the random variables to be estimated. In other words, is the correspondence between algorithmic randomness and truth-conduciveness robust?

The effectivity constraints imposed on random variables intuitively track the complexity of the corresponding inductive problems (L^1 -computability marking one possible level of complexity). Theorem 3.2.4 thus begs the question of whether, using the theory of algorithmic randomness, it is possible to further gauge the “size” and structure of the success set of a computable Bayesian agent in terms of the complexity of the inductive problem that they are facing.

These are the questions to which the remainder of this chapter is devoted. We will consider various natural classes of effective random variables and see how they yield different effectivisations of Lévy’s Upward Theorem. Most importantly, each time the success set of a Bayesian agent across problems in that class will be shown to correspond to a standard algorithmic randomness notion. As we shall see, the results presented in what follows also bring to the fore in a perspicuous way the fact that succeeding becomes more and more difficult as the complexity of the inductive problem faced by the computable Bayesian agent increases. More precisely, the harder to compute a random variable is, the logically stronger (and, thus, more restrictive) the algorithmic randomness notion characterising the success set for functions in that class.

3.2.2 Integral tests for randomness

We begin by studying effectivisations of Lévy’s Upward Theorem where the effective random variables under consideration are integral tests for randomness (as defined in §1.2.2 in the context of the measure-theoretic typicality paradigm).

Lower semi-computable random variables with computable expectation

Recall that an integral test for Schnorr randomness is a lower semi-computable random variable with computable expectation (cf. Theorem 1.2.15). Many inductive problems can be naturally modelled as random variables in this class. Suppose, for instance, that a Bayesian agent with prior λ (the uniform measure) were repeatedly tossing a coin and

wanted to test the following hypothesis: is the coin being tossed going to land heads more often than tails at least once? This hypothesis corresponds to the set $\mathcal{U} = \{\omega \in 2^{\mathbb{N}} : (\exists n) \frac{\#1(\omega \upharpoonright n)}{n} > \frac{1}{2}\}$, where 1 represents a heads outcome and 0 a tails outcome. From a Bayesian perspective, this inductive problem amounts to estimating the value of the indicator function $\chi_{\mathcal{U}} : 2^{\mathbb{N}} \rightarrow \{0, 1\}$ of \mathcal{U} . Clearly, $\chi_{\mathcal{U}}$ is lower semi-computable, since it is the indicator function of a Σ_1^0 class. Moreover, given that $\lambda(\mathcal{U}) = 1$, $\chi_{\mathcal{U}}$ has a computable expectation.

We now show that if one replaces L^1 -computable functions with lower semi-computable functions with computable expectation, then the resulting effectivisation of Lévy's Upward Theorem once again yields Schnorr randomness.

Theorem 3.2.5. *Let μ be a computable strictly positive measure and $\omega \in 2^{\mathbb{N}}$. Then, the following are equivalent:*

- (1) ω is μ -Schnorr random;
- (2) for all lower semi-computable functions $f : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ with computable expectation,

$$\lim_{k \rightarrow \infty} \mathbb{E}_{\mu}[f \mid \mathcal{F}_k](\omega) = f(\omega) < \infty.$$

As we shall see, just as in the case of Theorem 3.2.4, the forward direction of this result holds for all computable measures, not just the strictly positive ones.

To prove Theorem 3.2.5, we will appeal to a result due to Miyabe [2013b] (Theorem 3.2.6 below), which reveals that there is a tight correspondence between integral tests for Schnorr randomness and L^1 -computable functions. First, we need the following definition: a function $h : \subseteq 2^{\mathbb{N}} \rightarrow \mathbb{R}$ is said to be the *difference between two integral tests for Schnorr randomness* if there are two integral tests for Schnorr randomness g and ℓ such that $h(\omega) = g(\omega) - \ell(\omega)$ whenever $g(\omega) < \infty$ and $\ell(\omega) < \infty$, and such that $h(\omega)$ is undefined otherwise. Miyabe then proves the following, for all computable measures μ :

Theorem 3.2.6 (Miyabe [2013b]). (i) *Let h be the difference between two integral tests for μ -Schnorr randomness g and ℓ . Then, there is an L^1 -computable function f with witness $\{f_n\}_{n \in \mathbb{N}}$ such that $\hat{f}(\omega) = h(\omega)$ for all μ -Schnorr random sequences ω .¹⁷*

¹⁷Recall that $\hat{f}(\omega)$ equals $\lim_{n \rightarrow \infty} f_n(\omega)$ whenever this limit exists and is finite, and it is undefined otherwise.

- (ii) Let f be an L^1 -computable function with witness $\{f_n\}_{n \in \mathbb{N}}$. Then, there is a function h that is the difference between two integral tests for μ -Schnorr randomness such that $h(\omega) = \hat{f}(\omega)$ for all μ -Schnorr random sequences ω .

We additionally make use of the following simple lemma, which also holds for all computable measures μ (as opposed to just the strictly positive ones):

Lemma 3.2.7. *Let $f : \subseteq 2^{\mathbb{N}} \rightarrow \mathbb{R}$ be an L^1 -computable function with witness $\{f_n\}_{n \in \mathbb{N}}$. Then, $\|\hat{f} - f\|_1 = 0$.*

Proof. By Theorem 3.2.4, $\lim_{n \rightarrow \infty} f_n(\omega)$ exists and is finite for all μ -Schnorr random $\omega \in 2^{\mathbb{N}}$. Therefore, $\{f_n\}_{n \in \mathbb{N}}$ converges pointwise—and, as a consequence, \hat{f} is defined— μ -almost everywhere. By Proposition 3.2.3, $\{f_{n+1}\}_{n \in \mathbb{N}}$ is a fast computable L^1 -Cauchy sequence with $f_{n+1} \rightarrow f$ in L^1 . Moreover, we clearly have that $\lim_{n \rightarrow \infty} f_{n+1}(\omega) = \lim_{n \rightarrow \infty} f_n(\omega) = \hat{f}(\omega)$ for all μ -Schnorr random $\omega \in 2^{\mathbb{N}}$. Since $\sum_{n \in \mathbb{N}} \|f_{n+2} - f_{n+1}\|_1 < \infty$, it follows that $f_{n+1} \rightarrow \hat{f}$ in L^1 . But then, since $f_{n+1} \rightarrow f$ in L^1 , as well, we have that $\|\hat{f} - f\|_1 = 0$. \square

We are now ready to prove Theorem 3.2.5, which establishes the following, for computable strictly positive measures μ : (i) if a data stream is μ -Schnorr random, then the successive estimates of the Bayesian agent with prior μ converge to the truth for all lower semi-computable random variables with computable expectation; (ii) if a data stream is not μ -Schnorr random, then there is at least one lower semi-computable random variable with computable expectation such that the agent's successive estimates fail to converge to a limit—and so, *a fortiori*, fail to converge to the truth. Notably, since ω being μ -Schnorr random guarantees that, for all n , $\mu([\omega \upharpoonright n]) > 0$ even in the absence of strict positivity—which, as mentioned earlier, implies that the forward direction of Theorem 3.2.4 holds for all computable measures—the proof of (i) works for all computable measures, as well.

Proof of Theorem 3.2.5. (1) \Rightarrow (2) Suppose that $\omega \in 2^{\mathbb{N}}$ is μ -Schnorr random, and let f be an integral test for μ -Schnorr randomness. The function $f - 0$ (where 0 here denotes the constant-zero function) is clearly the difference between two integral tests for μ -Schnorr randomness. Hence, by Theorem 3.2.6, there is an L^1 -computable function ξ with witness $\{\xi_m\}_{m \in \mathbb{N}}$ such that $\hat{\xi}$ and f agree on all μ -Schnorr random sequences. So, in particular, $\hat{\xi}(\omega) = f(\omega)$. Moreover, for each n , $\mathbb{E}_\mu[\xi \mid \mathcal{F}_n](\omega) = \mathbb{E}_\mu[\hat{\xi} \mid \mathcal{F}_n](\omega) = \mathbb{E}_\mu[f \mid \mathcal{F}_n](\omega)$,

since $\xi, \hat{\xi}$, and f are μ -almost everywhere equal: ξ and $\hat{\xi}$ are μ -almost everywhere equal because, by Lemma 3.2.7, $\|\hat{\xi} - \xi\|_1 = 0$, while $\hat{\xi}$ and f are μ -almost everywhere equal because the collection of μ -Schnorr random sequences has μ -measure one. By Theorem 3.2.4, $\lim_{n \rightarrow \infty} \mathbb{E}_\mu[\xi \mid \mathcal{F}_n](\omega) = \lim_{m \rightarrow \infty} \xi_m(\omega) = \hat{\xi}(\omega)$. Hence, $\lim_{n \rightarrow \infty} \mathbb{E}_\mu[f \mid \mathcal{F}_n](\omega) = f(\omega)$.

(2) \Rightarrow (1) Suppose that ω is not μ -Schnorr random. We will show that there is a lower semi-computable function f with computable expectation such that $\{\mathbb{E}_\mu[f \mid \mathcal{F}_i](\omega)\}_{i \in \mathbb{N}}$ does not have a limit.¹⁸ Since ω is not μ -Schnorr random, there is a sequential test $\{\mathcal{U}_n\}_{n \in \mathbb{N}}$ for μ -Schnorr randomness such that $\omega \in \bigcap_{n \in \mathbb{N}} \mathcal{U}_n$. We will proceed as follows:

- firstly, we will define another sequential test $\{\mathcal{V}_n\}_{n \in \mathbb{N}}$ for μ -Schnorr randomness such that $\bigcap_{n \in \mathbb{N}} \mathcal{U}_n \subseteq \bigcap_{n \in \mathbb{N}} \mathcal{V}_n$;
- secondly, we will show that the function $g : 2^{\mathbb{N}} \rightarrow \mathbb{R}$ given by $g(\alpha) = 0$ whenever $\alpha \in \bigcap_{n \in \mathbb{N}} \mathcal{V}_n$ and $g(\alpha) = \sum_{n \in \mathbb{N}} (-1)^n \chi_{\mathcal{V}_n}$ otherwise—where $\chi_{\mathcal{V}_n}$ denotes the indicator function of \mathcal{V}_n —is L^1 -computable and fails Lévy's Upward Theorem on ω ;
- lastly, we will appeal to Theorem 3.2.6 to show that there is a lower semi-computable function f with computable expectation that fails Lévy's Upward Theorem along ω , as well.

We define the test $\{\mathcal{V}_n\}_{n \in \mathbb{N}}$ inductively as follows. Let $\mathcal{V}_0 = \mathcal{U}_0$. Now, suppose that \mathcal{V}_n has already been defined, so that $\mathcal{V}_n = \bigsqcup_{k \in \mathbb{N}} [\sigma_{n,k}]$ is a Σ_1^0 class. For each $[\sigma_{n,k}]$ enumerated into \mathcal{V}_n , find some $j_k \geq k$ with $2^{-j_k} < 2^{-(n+1)} \mu([\sigma_{n,k}])$ and, if $k > 0$, with $j_k > j_{k-1}$. Such a j_k can be found computably since $\mu([\sigma_{n,k}])$ is a computable real (and μ is strictly positive, so that $\mu([\sigma_{n,k}]) > 0$). Then, enumerate all of the (pairwise disjoint) cylinders included in $\mathcal{U}_{j_k} \cap [\sigma_{n,k}]$ into \mathcal{V}_{n+1} . This ends the construction. Note that, since $\mu(\mathcal{U}_{j_k}) \leq 2^{-j_k} < \mu([\sigma_{n,k}])$, all the cylinders enumerated into \mathcal{V}_{n+1} on behalf of $[\sigma_{n,k}]$ (that is, all the cylinders in $\mathcal{U}_{j_k} \cap [\sigma_{n,k}]$) must be generated by strings τ with $\sigma_{n,k} \sqsubset \tau$. We now show that $\{\mathcal{V}_n\}_{n \in \mathbb{N}}$ is indeed a sequential test for μ -Schnorr randomness. First of all, $\{\mathcal{V}_n\}_{n \in \mathbb{N}}$ is a (nested) sequence of uniformly Σ_1^0 classes. For $n = 0$, we have that $\mu(\mathcal{V}_0) = \mu(\mathcal{U}_0)$. So, $\mu(\mathcal{V}_0)$ is a computable real (and, obviously, we also have that $\mu(\mathcal{V}_0) \leq 2^{-0} = 1$). For

¹⁸The key idea behind this direction of the proof of Theorem 3.2.5 is adapted from the proof of Theorem 5.1 in [Freer et al., 2014], which provides a characterisation of the λ -Schnorr random reals in terms of the Lebesgue Differentiation Theorem and bounded L^1 -computable functions.

$n \geq 1$, on the other hand, we have that

$$\mu(\mathcal{V}_n) \leq \sum_{k \in \mathbb{N}} \mu(\mathcal{U}_{j_k}) \leq \sum_{k \in \mathbb{N}} 2^{-n} \mu([\sigma_{n-1,k}]) = 2^{-n} \mu(\mathcal{V}_{n-1}) \leq 2^{-(2n-1)} \leq 2^{-n},$$

where j_k denotes the index selected on behalf of cylinder $[\sigma_{n-1,k}] \in \mathcal{V}_{n-1}$ during the construction of \mathcal{V}_n . For each $K \in \mathbb{N}$, $\mu(\bigsqcup_{k \leq K} \mathcal{U}_{j_k} \cap [\sigma_{n-1,k}])$ is a computable real, uniformly in K . To see this, let $[\tau_{j_k,0}], [\tau_{j_k,1}], \dots$ be a computable enumeration of the cylinders in \mathcal{U}_{j_k} . Then, $\mathcal{U}_{j_k} \cap [\sigma_{n-1,k}] = \bigsqcup_{\ell \in \mathbb{N}} [\tau_{j_k,\ell}] \cap [\sigma_{n-1,k}]$. Moreover, for each ℓ , $[\tau_{j_k,\ell}] \cap [\sigma_{n-1,k}]$ is either empty or equal to $[\tau_{j_k,\ell}]$, so $\{\mu([\tau_{j_k,\ell}] \cap [\sigma_{n-1,k}])\}_{\ell \in \mathbb{N}}$ is a sequence of uniformly computable non-negative reals. Now, $\sum_{\ell \in \mathbb{N}} \mu([\tau_{j_k,\ell}] \cap [\sigma_{n-1,k}]) = \mu(\bigsqcup_{\ell \in \mathbb{N}} [\tau_{j_k,\ell}] \cap [\sigma_{n-1,k}]) = \mu(\mathcal{U}_{j_k} \cap [\sigma_{n-1,k}])$ and, so, is a computable real, uniformly in k , since $\{\mathcal{U}_m\}_{m \in \mathbb{N}}$ is a sequential test for μ -Schnorr randomness. Let $\{\ell_m\}_{m \in \mathbb{N}}$ be an increasing computable sequence of natural numbers such that, for all m , $\sum_{i=\ell_m+1}^{\infty} \mu([\tau_{j_k,i}]) \leq 2^{-m}$. Then, for all m ,

$$\mu\left(\bigsqcup_{i \geq \ell_m+1} [\tau_{j_k,i}] \cap [\sigma_{n-1,k}]\right) = \sum_{i=\ell_m+1}^{\infty} \mu([\tau_{j_k,i}] \cap [\sigma_{n-1,k}]) \leq \sum_{i=\ell_m+1}^{\infty} \mu([\tau_{j_k,i}]) \leq 2^{-m}.$$

Hence, for all m ,

$$\begin{aligned} \left| \mu(\mathcal{U}_{j_k} \cap [\sigma_{n-1,k}]) - \mu\left(\bigsqcup_{i=0}^{\ell_m} [\tau_{j_k,i}] \cap [\sigma_{n-1,k}]\right) \right| &= \left| \sum_{i \in \mathbb{N}} \mu([\tau_{j_k,i}] \cap [\sigma_{n-1,k}]) - \sum_{i=0}^{\ell_m} \mu([\tau_{j_k,i}] \cap [\sigma_{n-1,k}]) \right| \\ &= \left| \sum_{i=\ell_m+1}^{\infty} \mu([\tau_{j_k,i}] \cap [\sigma_{n-1,k}]) \right| \\ &\leq 2^{-m}. \end{aligned}$$

Since $\{\sum_{i=0}^{\ell_m} \mu([\tau_{j_k,i}] \cap [\sigma_{n-1,k}])\}_{m \in \mathbb{N}}$ is a computable sequence of computable reals, this suffices to conclude that $\mu(\mathcal{U}_{j_k} \cap [\sigma_{n-1,k}])$ is computable, uniformly in k . As a result, $\mu(\bigsqcup_{k \leq K} \mathcal{U}_{j_k} \cap [\sigma_{n-1,k}]) = \sum_{k=0}^K \mu(\mathcal{U}_{j_k} \cap [\sigma_{n-1,k}])$ is also a computable real, uniformly in K . Now, given that $\mathcal{V}_n = \bigsqcup_{k \in \mathbb{N}} (\mathcal{U}_{j_k} \cap [\sigma_{n-1,k}])$, for every K , we have that

$$\left| \mu(\mathcal{V}_n) - \mu\left(\bigsqcup_{k \leq K} \mathcal{U}_{j_k} \cap [\sigma_{n-1,k}]\right) \right| = \mu\left(\bigsqcup_{k=K+1}^{\infty} \mathcal{U}_{j_k} \cap [\sigma_{n-1,k}]\right) \leq 2^{-K},$$

where the inequality holds because, for all k , $j_k \geq k$. Hence, $\mu(\mathcal{V}_n)$ is a computable real,

uniformly in n . This establishes that $\{\mathcal{V}_n\}_{n \in \mathbb{N}}$ is indeed a sequential test for μ -Schnorr randomness.

Note, in addition, that $\bigcap_{n \in \mathbb{N}} \mathcal{U}_n \subseteq \bigcap_{n \in \mathbb{N}} \mathcal{V}_n$. For, suppose that $\alpha \in \bigcap_{n \in \mathbb{N}} \mathcal{U}_n$. Since $\mathcal{V}_0 = \mathcal{U}_0$ and $\alpha \in \mathcal{U}_0$, $\alpha \in \mathcal{V}_0$. Now, suppose that $\alpha \in \mathcal{V}_n$. Then, there is some k such that $\alpha \in [\sigma_{n,k}] \subseteq \mathcal{V}_n$. Let \mathcal{U}_{j_k} be such that $\mathcal{U}_{j_k} \cap [\sigma_{n,k}]$ is enumerated into \mathcal{V}_{n+1} . Given that $\alpha \in \mathcal{U}_{j_k}$, $\alpha \in \mathcal{U}_{j_k} \cap [\sigma_{n,k}]$ and, thus, $\alpha \in \mathcal{V}_{n+1}$. Hence, $\alpha \in \bigcap_{n \in \mathbb{N}} \mathcal{V}_n$. Then, in particular, $\omega \in \bigcap_{n \in \mathbb{N}} \mathcal{V}_n$.

Next, we prove that the function g defined above is L^1 -computable. The functions $\chi_{\mathcal{V}_n}$ are L^1 -computable uniformly in n , since $\{\mathcal{V}_n\}_{n \in \mathbb{N}}$ is a sequential test for μ -Schnorr randomness. For each n , let $\{\xi_{n,k}\}_{k \in \mathbb{N}}$ be a computable sequence of rational-valued step functions witnessing the L^1 -computability of $\chi_{\mathcal{V}_n}$. We begin by proving that, for all m , $g_m = \sum_{n=0}^m (-1)^n \chi_{\mathcal{V}_n}$ is also L^1 -computable, uniformly in m . Fix m . For each k , let $g_{m,k} = \sum_{n=0}^m (-1)^n \xi_{n,k}$. Then, $\{g_{m,k}\}_{k \in \mathbb{N}}$ is a computable sequence of rational-valued step functions, uniformly in m . For every k , let ℓ_k be such that $2^{-\ell_k} \leq 2^{-k}(m+1)^{-1}$. Additionally, for each $k \geq 1$, also make sure that $\ell_k > \ell_{k-1}$. Let $g'_{m,k} = g_{m,\ell_k}$. Then, $\{g'_{m,k}\}_{k \in \mathbb{N}}$, too, is a computable sequence of rational-valued step functions, uniformly in m . Now,

$$\begin{aligned}
\int_{2^{\mathbb{N}}} |g'_{m,k} - g_m| d\mu &= \int_{2^{\mathbb{N}}} \left| \sum_{n=0}^m (-1)^n \xi_{n,\ell_k} - \sum_{n=0}^m (-1)^n \chi_{\mathcal{V}_n} \right| d\mu \\
&= \int_{2^{\mathbb{N}}} \left| \sum_{\substack{n=0 \\ n \text{ even}}}^m (\xi_{n,\ell_k} - \chi_{\mathcal{V}_n}) + \sum_{\substack{n=1 \\ n \text{ odd}}}^m (\chi_{\mathcal{V}_n} - \xi_{n,\ell_k}) \right| d\mu \\
&\leq \int_{2^{\mathbb{N}}} \sum_{\substack{n=0 \\ n \text{ even}}}^m |\xi_{n,\ell_k} - \chi_{\mathcal{V}_n}| + \sum_{\substack{n=1 \\ n \text{ odd}}}^m |\chi_{\mathcal{V}_n} - \xi_{n,\ell_k}| d\mu \\
&= \sum_{n=0}^m \int_{2^{\mathbb{N}}} |\xi_{n,\ell_k} - \chi_{\mathcal{V}_n}| d\mu \\
&\leq (m+1) \cdot 2^{-\ell_k} \\
&\leq 2^{-k}.
\end{aligned}$$

Hence, g_m is L^1 -computable uniformly in m . Let $h_{k,k} = g'_{k+1,k+1}$. We will now show that $\{h_{k,k}\}_{k \in \mathbb{N}}$, which is also a computable sequence of rational-valued step functions, is a

witness to the L^1 -computability of g :

$$\begin{aligned}
\int_{2^{\mathbb{N}}} |h_{k,k} - g| d\mu &= \int_{2^{\mathbb{N}}} |g'_{k+1,k+1} - g| d\mu \\
&\leq \int_{2^{\mathbb{N}}} |g'_{k+1,k+1} - g_{k+1}| d\mu + \int_{2^{\mathbb{N}}} |g_{k+1} - g| d\mu \\
&\leq 2^{-(k+1)} + \int_{2^{\mathbb{N}}} \left| \sum_{n=0}^{k+1} (-1)^n \chi_{\mathcal{V}_n} - \sum_{n \in \mathbb{N}} (-1)^n \chi_{\mathcal{V}_n} \right| d\mu \\
&= 2^{-(k+1)} + \int_{2^{\mathbb{N}}} \left| \sum_{n=k+2}^{\infty} (-1)^n \chi_{\mathcal{V}_n} \right| d\mu \\
&\leq 2^{-(k+1)} + \sum_{n=k+2}^{\infty} \int_{2^{\mathbb{N}}} \chi_{\mathcal{V}_n} d\mu \\
&= 2^{-(k+1)} + \sum_{n=k+2}^{\infty} \mu(\mathcal{V}_n) \\
&\leq 2^{-(k+1)} + 2^{-(k+1)} \\
&= 2^{-k}.
\end{aligned}$$

Now, we argue that the sequence $\{\mathbb{E}_\mu[g \mid \mathcal{F}_i](\omega)\}_{i \in \mathbb{N}}$ does not have a limit. Since $\omega \in \bigcap_{n \in \mathbb{N}} \mathcal{V}_n$, for every n , there is some k_n such that $\omega \in [\sigma_{n,k_n}] \subseteq \mathcal{V}_n$, and $\sigma_{0,k_0} \sqsubset \sigma_{1,k_1} \sqsubset \sigma_{2,k_2} \sqsubset \dots$. We begin by showing that $\lim_{\substack{n \rightarrow \infty \\ n \text{ even}}} \frac{1}{\mu([\sigma_{n,k_n}])} \int_{[\sigma_{n,k_n}]} g d\mu = 1$. Let n be even. For all $i \leq n$, $[\sigma_{n,k_n}] \subseteq \mathcal{V}_i$, since $\{\mathcal{V}_m\}_{m \in \mathbb{N}}$ is by construction a nested sequence. Hence,

$$\frac{1}{\mu([\sigma_{n,k_n}])} \int_{[\sigma_{n,k_n}]} \sum_{i=0}^n (-1)^i \chi_{\mathcal{V}_i} d\mu = 1.$$

Let $[\tau_0], [\tau_1], [\tau_2], \dots$ be an enumeration of all the cylinders in $\mathcal{V}_{n+1} \cap [\sigma_{n,k_n}]$. Then, for each $[\tau_j]$, let $[\tau_{j,0}], [\tau_{j,1}], [\tau_{j,2}], \dots$ be an enumeration of all the cylinders in $\mathcal{V}_{n+2} \cap [\tau_j]$, so that $\mathcal{V}_{n+2} \cap [\sigma_{n,k_n}] = \bigsqcup_{j \in \mathbb{N}} [\tau_{j,\ell}]$. By the same reasoning, for each $i > n$, we have that $\mathcal{V}_i \cap [\sigma_{n,k_n}] = \bigsqcup_{j_1, \dots, j_{(i-n)} \in \mathbb{N}} [\tau_{j_1, \dots, j_{(i-n)}}]$. Therefore,

$$\begin{aligned}
\mu(\mathcal{V}_i \cap [\sigma_{n,k_n}]) &= \mu\left(\bigsqcup_{j_1, \dots, j_{(i-n)} \in \mathbb{N}} [\tau_{j_1, \dots, j_{(i-n)}}] \right) \\
&= \sum_{j_1, \dots, j_{(i-n)} \in \mathbb{N}} \mu([\tau_{j_1, \dots, j_{(i-n)}}])
\end{aligned}$$

$$\begin{aligned}
&\leq 2^{-i} \sum_{j_1, \dots, j_{(i-1)-n} \in \mathbb{N}} \mu([\tau_{j_1, \dots, j_{(i-1)-n}}]) \\
&\leq 2^{-i} \mu([\sigma_{n, k_n}]).
\end{aligned}$$

It then follows that

$$\begin{aligned}
\left| \frac{1}{\mu([\sigma_{n, k_n}])} \int_{[\sigma_{n, k_n}]} g d\mu - 1 \right| &= \left| \frac{\int_{[\sigma_{n, k_n}]} \sum_{i \in \mathbb{N}} (-1)^i \chi_{\mathcal{V}_i} d\mu}{\mu([\sigma_{n, k_n}])} - 1 \right| \\
&= \left| \frac{\int_{[\sigma_{n, k_n}]} (\sum_{i=0}^n (-1)^i \chi_{\mathcal{V}_i} d\mu + \sum_{i=n+1}^{\infty} (-1)^i \chi_{\mathcal{V}_i} d\mu)}{\mu([\sigma_{n, k_n}])} - 1 \right| \\
&= \left| \frac{\int_{[\sigma_{n, k_n}]} \sum_{i=0}^n (-1)^i \chi_{\mathcal{V}_i} d\mu}{\mu([\sigma_{n, k_n}])} + \frac{\int_{[\sigma_{n, k_n}]} \sum_{i=n+1}^{\infty} (-1)^i \chi_{\mathcal{V}_i} d\mu}{\mu([\sigma_{n, k_n}])} - 1 \right| \\
&= \left| \frac{1}{\mu([\sigma_{n, k_n}])} \int_{[\sigma_{n, k_n}]} \sum_{i=n+1}^{\infty} (-1)^i \chi_{\mathcal{V}_i} d\mu \right| \\
&\leq \frac{1}{\mu([\sigma_{n, k_n}])} \int_{[\sigma_{n, k_n}]} \sum_{i=n+1}^{\infty} \chi_{\mathcal{V}_i} d\mu \\
&\leq \frac{1}{\mu([\sigma_{n, k_n}])} \sum_{i=n+1}^{\infty} \mu(\mathcal{V}_i \cap [\sigma_{n, k_n}]) \\
&\leq \frac{1}{\mu([\sigma_{n, k_n}])} \sum_{i=n+1}^{\infty} 2^{-i} \mu([\sigma_{n, k_n}]) \\
&= 2^{-n}.
\end{aligned}$$

Hence, $\lim_{\substack{n \rightarrow \infty \\ n \text{ even}}} \frac{1}{\mu([\sigma_{n, k_n}])} \int_{[\sigma_{n, k_n}]} g d\mu = 1$, which, in turn, entails that $\limsup_{i \rightarrow \infty} \mathbb{E}_\mu[g \mid \mathcal{F}_i](\omega) \geq$

1. Next, we show that $\lim_{\substack{n \rightarrow \infty \\ n \text{ odd}}} \frac{1}{\mu([\sigma_{n, k_n}])} \int_{[\sigma_{n, k_n}]} g d\mu = 0$. Let n be odd. Then, for all $i \leq n$,

$$\frac{1}{\mu([\sigma_{n, k_n}])} \int_{[\sigma_{n, k_n}]} \sum_{i=0}^n (-1)^i \chi_{\mathcal{V}_i} d\mu = 0.$$

We then have that

$$\begin{aligned}
\left| \frac{1}{\mu([\sigma_{n, k_n}])} \int_{[\sigma_{n, k_n}]} g d\mu - 0 \right| &= \left| \frac{\int_{[\sigma_{n, k_n}]} \sum_{i=0}^n (-1)^i \chi_{\mathcal{V}_i} d\mu}{\mu([\sigma_{n, k_n}])} + \frac{\int_{[\sigma_{n, k_n}]} \sum_{i=n+1}^{\infty} (-1)^i \chi_{\mathcal{V}_i} d\mu}{\mu([\sigma_{n, k_n}])} - 0 \right| \\
&= \left| \frac{1}{\mu([\sigma_{n, k_n}])} \int_{[\sigma_{n, k_n}]} \sum_{i=n+1}^{\infty} (-1)^i \chi_{\mathcal{V}_i} d\mu \right|,
\end{aligned}$$

where the last quantity is at most 2^{-n} . Hence, $\lim_{\substack{n \rightarrow \infty \\ n \text{ odd}}} \frac{1}{\mu([\sigma_n, k_n])} \int_{[\sigma_n, k_n]} g d\mu = 0$, which, in turn, entails that $\liminf_{i \rightarrow \infty} \mathbb{E}_\mu[g \mid \mathcal{F}_i](\omega) \leq 0$. Therefore, the sequence $\{\mathbb{E}_\mu[g \mid \mathcal{F}_i](\omega)\}_{i \in \mathbb{N}}$ indeed fails to have a limit.

Lastly, by Theorem 3.2.6, there is a function h that is the difference between two integral tests for μ -Schnorr randomness f_1 and f_2 , and such that h and \hat{g} agree on all μ -Schnorr random sequences. For each i , $\mathbb{E}_\mu[h \mid \mathcal{F}_i](\omega) = \mathbb{E}_\mu[\hat{g} \mid \mathcal{F}_i](\omega) = \mathbb{E}_\mu[g \mid \mathcal{F}_i](\omega)$. Therefore, $\{\mathbb{E}_\mu[h \mid \mathcal{F}_i](\omega)\}_{i \in \mathbb{N}}$ does not have a limit either. Since $\mathbb{E}_\mu[h \mid \mathcal{F}_i](\omega) = \mathbb{E}_\mu[f_1 \mid \mathcal{F}_i](\omega) - \mathbb{E}_\mu[f_2 \mid \mathcal{F}_i](\omega)$, we in turn have that either $\{\mathbb{E}_\mu[f_1 \mid \mathcal{F}_i](\omega)\}_{i \in \mathbb{N}}$ or $\{\mathbb{E}_\mu[f_2 \mid \mathcal{F}_i](\omega)\}_{i \in \mathbb{N}}$ fails to have a limit. Then, let f be whichever one of f_1 or f_2 witnesses this failure. This concludes the proof.¹⁹ \square

Lower semi-computable random variables with finite expectation

Next, we restrict attention to the class of lower semi-computable random variables with finite expectation (as opposed to computable expectation): namely, the collection of integral tests for Martin-Löf randomness. To see that, from a learning-theoretic perspective, this is a natural class to consider, note, for instance, that the indicator functions of Σ_1^0 classes (including the Σ_1^0 classes whose measure is not computable) fall under this category.

Our main goal in what follows is showing that the type of randomness that emerges from the effectivisation of Lévy's Upward Theorem obtained by restricting attention to lower semi-computable random variables with finite expectation is another natural algorithmic randomness notion: density randomness (as given by Definition 1.3.5). Observing a density random data stream guarantees that the beliefs of a computable Bayesian agent will converge to the truth on all inductive problems in this class.

¹⁹Note that this direction of Theorem 3.2.5 can be given a more succinct proof by combining Theorem 3.2.4 and Theorem 3.2.6. Our proof, however, does not rely on Theorem 3.2.4. This is because the aim of the proof is to showcase an alternative route for establishing this direction, which could also be followed to offer a different proof of the backward direction of Theorem 3.2.4. The shorter proof goes as follows. Suppose that ω fails to be μ -Schnorr random. Then, by the proof of the backward direction of Theorem 3.2.4, there is an L^1 -computable function f such that the sequence $\{\mathbb{E}_\mu[f \mid \mathcal{F}_k](\omega)\}_{k \in \mathbb{N}}$ does not have a limit. Crucially, f is defined differently from the function g given above. The rest of the proof is then analogous. By Theorem 3.2.6, there is a function h that is the difference between two integral tests for μ -Schnorr randomness g and ℓ , and such that h and \hat{f} agree on all μ -Schnorr random sequences. For each k , $\mathbb{E}_\mu[h \mid \mathcal{F}_k](\omega) = \mathbb{E}_\mu[\hat{f} \mid \mathcal{F}_k](\omega) = \mathbb{E}_\mu[f \mid \mathcal{F}_k](\omega)$. Hence, $\{\mathbb{E}_\mu[h \mid \mathcal{F}_k](\omega)\}_{k \in \mathbb{N}}$ does not have a limit either. Since $\mathbb{E}_\mu[h \mid \mathcal{F}_k](\omega) = \mathbb{E}_\mu[g \mid \mathcal{F}_k](\omega) - \mathbb{E}_\mu[\ell \mid \mathcal{F}_k](\omega)$, we in turn have that either $\{\mathbb{E}_\mu[g \mid \mathcal{F}_k](\omega)\}_{k \in \mathbb{N}}$ or $\{\mathbb{E}_\mu[\ell \mid \mathcal{F}_k](\omega)\}_{k \in \mathbb{N}}$ fails to have a limit. Thus, either g or ℓ is a witness to the failure of Condition (2).

If f is a lower semi-computable random variable with finite expectation, then the function $d_f : 2^{<\mathbb{N}} \rightarrow \mathbb{R}$, where $d_f(\sigma)$ equals $\frac{\int_{[\sigma]} f d\mu}{\mu([\sigma])}$ if $\mu([\sigma]) > 0$ and is undefined otherwise, is an almost everywhere left-c.e. dyadic μ -martingale. This follows because f is lower semi-computable, which ensures that d_f is (almost everywhere) left-c.e., and because, without loss of generality, f can be assumed to be non-negative. If, in addition, μ is a strictly positive measure, then d_f is total and, thus, a left-c.e. dyadic μ -martingale. Density randomness is defined in terms of the convergence of (almost everywhere) left-c.e. dyadic martingales: this will turn out to be crucial for the characterisation of density randomness in terms of Lévy's Upward Theorem. Before presenting this result, however, we will first see that Martin-Löf randomness is too weak to be characterised via the effectivisation of Lévy's Upward Theorem in terms of lower semi-computable random variables with finite expectation. There are in fact computable strictly positive measures μ and lower semi-computable random variables f with finite expectation such that d_f fails to converge along some μ -Martin-Löf random sequence.

First, we need to introduce the notion of a *Borel normal sequence*.²⁰ For each $m \geq 1$, consider the set of strings 2^m equipped with the lexicographic order. Then, for each $1 \leq i \leq 2^m$ and $\alpha \in (2^m)^{<\mathbb{N}}$, let $N_i^m(\alpha)$ denote the number of occurrences of the i -th string from 2^m in α . Now, for each $\sigma \in 2^{<\mathbb{N}}$ and $m \geq 1$, let $\langle \sigma, m \rangle$ denote the string $\tau_1 \dots \tau_k \in (2^m)^{<\mathbb{N}}$, where $k = \lceil |\sigma| / m \rceil$ and, for each $1 \leq j \leq k$, $\tau_j = \sigma((j-1)m+1) \dots \sigma(jm)$. Then, let $N_i^m(\sigma)$ denote the number of occurrences of the i -th string from 2^m in $\langle \sigma, m \rangle$. Intuitively, we divide σ into blocks of length m and count the number of occurrences of the i -th string from 2^m among these blocks.

Definition 3.2.8 (Borel normal sequence). (a) *Let $m \geq 1$. A sequence $\omega \in 2^{\mathbb{N}}$ is Borel m -normal if, for all $1 \leq i \leq 2^m$,*

$$\lim_{n \rightarrow \infty} \frac{N_i^m(\omega \upharpoonright n)}{\frac{n}{m}} = 2^{-m}.$$

(b) *A sequence $\omega \in 2^{\mathbb{N}}$ is Borel normal if it is Borel m -normal for all $m \geq 1$.*

Take the uniform measure λ . It is a well-known fact that λ -Martin-Löf randomness implies Borel normality (see, for instance, [Becher, 2012]). We make use of this fact in

²⁰Cf. [Calude, 1994], [Becher, 2012], or [Becher and Figueira, 2002].

rehearsing the proof of the result below, from which we can draw an important lesson: observing a Martin-Löf random data stream does not guarantee convergence to the truth in the sense of Lévy's Upward Theorem if the quantity to be estimated is a lower semi-computable random variable with finite expectation.

Observation 3.2.9 (Miyabe et al. [2016]). *There is a λ -Martin-Löf random sequence ω and a lower semi-computable random variable $f : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ with finite expectation such that the sequence $\{\mathbb{E}_\lambda[f \mid \mathcal{F}_k](\omega)\}_{k \in \mathbb{N}}$ fails to have a limit.*

Proof. Let ω be a left-c.e. λ -Martin-Löf random sequence²¹ and $\mathcal{C} = \{\sigma \in 2^{<\mathbb{N}} : \sigma <_L \omega\}$, where $<_L$ denotes the lexicographic order extended to infinite sequences, so that $\sigma <_L \omega$ if and only if there is $\rho \in 2^{<\mathbb{N}}$ with $\rho 0 \sqsubseteq \sigma$ and $\rho 1 \sqsubset \omega$. Since ω is left-c.e., \mathcal{C} is a c.e. set. Now, $\bigcup\{\sigma : \sigma \in \mathcal{C}\} = \{\alpha \in 2^{\mathbb{N}} : \alpha <_L \omega\}$. Call this set \mathcal{U} . We show that

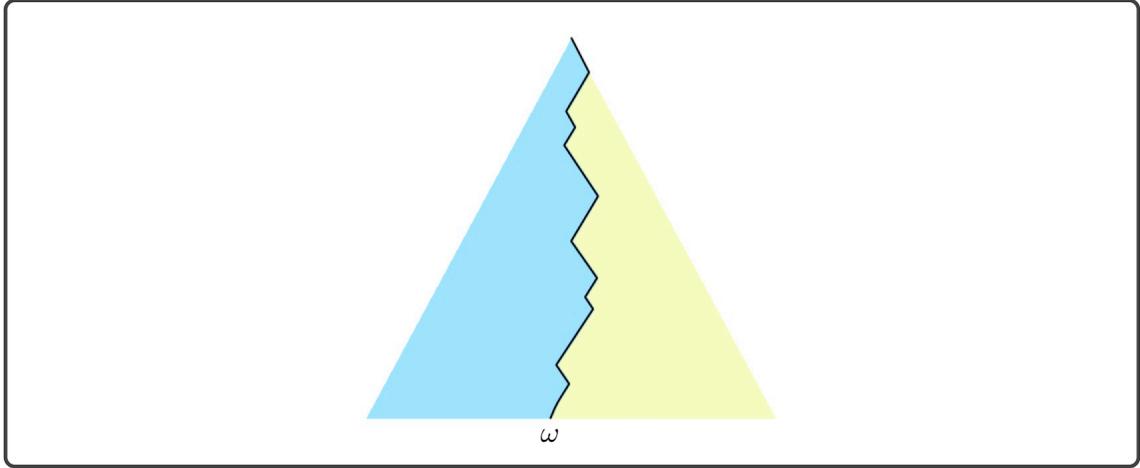


Figure 3.1: A depiction of Cantor space, where the black segmented line represents the sequence ω . The blue-shaded area to the left of ω is \mathcal{U} : the collection of sequences that precede ω in the lexicographic order. The yellow-shaded area to the right of ω , on the other hand, is $\overline{\mathcal{U}}$, the complement of \mathcal{U} .

the indicator function $\chi_{\mathcal{U}}$ of \mathcal{U} provides the desired counterexample. Clearly, $\chi_{\mathcal{U}}$ is lower semi-computable, since \mathcal{U} is a Σ_1^0 class. Moreover, $\int_{2^{\mathbb{N}}} \chi_{\mathcal{U}} d\lambda = \lambda(\mathcal{U}) < \infty$. However,

- (i) $\limsup_{k \rightarrow \infty} \mathbb{E}_\lambda[\chi_{\mathcal{U}} \mid \mathcal{F}_k](\omega) = \limsup_{k \rightarrow \infty} \frac{\lambda(\mathcal{U} \cap [\omega \upharpoonright k])}{\lambda([\omega \upharpoonright k])} = 1$, while
- (ii) $\liminf_{k \rightarrow \infty} \mathbb{E}_\lambda[\chi_{\mathcal{U}} \mid \mathcal{F}_k](\omega) = \liminf_{k \rightarrow \infty} \frac{\lambda(\mathcal{U} \cap [\omega \upharpoonright k])}{\lambda([\omega \upharpoonright k])} = 0$.

²¹E.g., Ω (see Example 1.2.4).

In both cases, the claim follows from the fact that ω , being λ -Martin-Löf random, is Borel normal—and, in particular, Borel normal to base 2. To establish (i), it suffices to show that, for every n , there are infinitely many k such that $\frac{\lambda(\mathcal{U} \cap [\omega \upharpoonright k])}{\lambda([\omega \upharpoonright k])} \geq 1 - 2^{-n}$. So, fix n . By Borel normality to base 2, we have that the string

$$\sigma = \underbrace{1 \dots 1}_{n \text{ 1's}}$$

occurs infinitely often along ω . For each of the infinitely many occurrences of σ , let $\omega \upharpoonright k$ denote the initial segment of ω that precedes σ : i.e., $\omega \upharpoonright k + n = (\omega \upharpoonright k) \frown \sigma$. Now, $\bar{\mathcal{U}} \cap [\omega \upharpoonright k] = \bar{\mathcal{U}} \cap [(\omega \upharpoonright k) \frown \sigma] \subseteq [(\omega \upharpoonright k) \frown \sigma]$. So, $\lambda(\bar{\mathcal{U}} \cap [\omega \upharpoonright k]) \leq \lambda([(\omega \upharpoonright k) \frown \sigma]) = 2^{-(k+n)}$. Since $\mathcal{U} \cap [\omega \upharpoonright k] = [\omega \upharpoonright k] \setminus (\bar{\mathcal{U}} \cap [\omega \upharpoonright k])$, we then have that

$$\begin{aligned} \frac{\lambda(\mathcal{U} \cap [\omega \upharpoonright k])}{\lambda([\omega \upharpoonright k])} &= \frac{\lambda([\omega \upharpoonright k] \setminus (\bar{\mathcal{U}} \cap [\omega \upharpoonright k]))}{\lambda([\omega \upharpoonright k])} \\ &= \frac{\lambda([\omega \upharpoonright k])}{\lambda([\omega \upharpoonright k])} - \frac{\lambda(\bar{\mathcal{U}} \cap [\omega \upharpoonright k])}{\lambda([\omega \upharpoonright k])} \\ &\geq 1 - 2^{-n}. \end{aligned}$$

To establish (ii), on the other hand, it suffices to show that, for every n , there are infinitely many k such that $\frac{\lambda(\mathcal{U} \cap [\omega \upharpoonright k])}{\lambda([\omega \upharpoonright k])} \leq 2^{-n}$. The argument is analogous to the previous one. Fix n . Borel normality to base 2 implies that the string

$$\tau = \underbrace{0 \dots 0}_{n \text{ 0's}}$$

occurs infinitely often along ω . For each occurrence of τ , let $\omega \upharpoonright k$ be the initial segment of ω that precedes τ . Now, $\mathcal{U} \cap [\omega \upharpoonright k] = \mathcal{U} \cap [(\omega \upharpoonright k) \frown \tau] \subseteq [(\omega \upharpoonright k) \frown \tau]$. So,

$$\frac{\lambda(\mathcal{U} \cap [\omega \upharpoonright k])}{\lambda([\omega \upharpoonright k])} \leq \frac{\lambda([(\omega \upharpoonright k) \frown \tau])}{\lambda([\omega \upharpoonright k])} = 2^{-n}.$$

Hence, the sequence $\{\mathbb{E}_\lambda[f \mid \mathcal{F}_k](\omega)\}_{k \in \mathbb{N}}$ does not have a limit. \square

As once again shown by Miyabe et al. [2016], Martin-Löf randomness does however suffice to guarantee that the limit of the conditional expectations of a lower semi-computable

random variable with finite expectation is equal to the correct value of that function, provided that this limit exists. Their argument is given in the context of the uniform measure, but, as shown below, it can be easily generalised to arbitrary computable measures.

Lemma 3.2.10 (Miyabe et al. [2016]). *Let $f : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ be a lower semi-computable random variable with finite expectation and $\omega \in 2^{\mathbb{N}}$ a μ -Martin-Löf random sequence, with μ a computable measure. If $\lim_{k \rightarrow \infty} \mathbb{E}_{\mu}[f \mid \mathcal{F}_k](\omega)$ exists, then $\lim_{k \rightarrow \infty} \mathbb{E}_{\mu}[f \mid \mathcal{F}_k](\omega) = f(\omega)$.*

Proof. Without loss of generality, f can be assumed to be non-negative. Moreover, since ω is μ -Martin-Löf random, $\mu([\omega \upharpoonright k]) > 0$ for all k . We begin by showing that $\lim_{k \rightarrow \infty} \mathbb{E}_{\mu}[f \mid \mathcal{F}_k](\omega) \geq f(\omega)$. Since f is lower semi-computable, by Definition 1.2.12, there is a sequence of uniformly computable rational-valued functions $g_n : 2^{<\mathbb{N}} \rightarrow \mathbb{Q}$ such that

- (a) $g_{n+1}(\sigma) \geq g_n(\sigma)$ for all $n \geq 0$ and $\sigma \in 2^{<\mathbb{N}}$;
- (b) $g_n(\sigma\tau) \geq g_n(\sigma)$ for all $n \geq 0$ and all $\sigma, \tau \in 2^{<\mathbb{N}}$;
- (c) $f(\alpha) = \sup\{g_n(\alpha \upharpoonright k) : n, k \geq 0\}$ for all $\alpha \in 2^{\mathbb{N}}$.

Hence, for all $\alpha \in 2^{\mathbb{N}}$, $\lim_{n \rightarrow \infty} g_n(\alpha \upharpoonright n) = f(\alpha)$. Now, let $\sigma_1, \dots, \sigma_{2^n}$ be an enumeration of all strings in $2^{<\mathbb{N}}$ of length n , and, for each $\alpha \in 2^{\mathbb{N}}$, let $f_n(\alpha) = \sum_{i=1}^{2^n} g_n(\sigma_i) \cdot \chi_{[\sigma_i]}(\alpha) = g_n(\alpha \upharpoonright n)$. Then, $\{f_n\}_{n \in \mathbb{N}}$ is a computable non-decreasing sequence of rational-valued step functions such that $\lim_{n \rightarrow \infty} f_n(\alpha) = f(\alpha)$ for all $\alpha \in 2^{\mathbb{N}}$. Clearly, for every $\alpha \in 2^{\mathbb{N}}$ with $\mu([\alpha \upharpoonright k]) > 0$ for all k , $f_n(\alpha) = \lim_{k \rightarrow \infty} \mathbb{E}_{\mu}[f_n \mid \mathcal{F}_k](\alpha)$, as $\mathbb{E}_{\mu}[f_n \mid \mathcal{F}_k](\alpha)$ is eventually constant (past $k = n$), with final value $g_n(\alpha \upharpoonright n) \in \mathbb{Q}$. By assumption, $\lim_{k \rightarrow \infty} \mathbb{E}_{\mu}[f \mid \mathcal{F}_k](\omega)$ exists. Call this limit ℓ . Since $\lim_{n \rightarrow \infty} f_n = f$ everywhere, and $f_0 \leq f_1 \leq f_2 \leq \dots$, for every $\sigma \in 2^{<\mathbb{N}}$ we have that

$$\int_{[\sigma]} f_0 d\mu \leq \int_{[\sigma]} f_1 d\mu \leq \int_{[\sigma]} f_2 d\mu \leq \dots \leq \int_{[\sigma]} f d\mu.$$

In particular, for all $n, k \geq 0$,

$$\frac{1}{\mu([\omega \upharpoonright k])} \int_{[\omega \upharpoonright k]} f_n d\mu \leq \frac{1}{\mu([\omega \upharpoonright k])} \int_{[\omega \upharpoonright k]} f d\mu,$$

from which it follows that

$$f_n(\omega) = \lim_{k \rightarrow \infty} \frac{1}{\mu([\omega \upharpoonright k])} \int_{[\omega \upharpoonright k]} f_n d\mu \leq \lim_{k \rightarrow \infty} \frac{1}{\mu([\omega \upharpoonright k])} \int_{[\omega \upharpoonright k]} f d\mu.$$

Since $\lim_{n \rightarrow \infty} f_n(\omega) = f(\omega)$, we have that

$$f(\omega) = \lim_{n \rightarrow \infty} f_n(\omega) \leq \lim_{k \rightarrow \infty} \frac{1}{\mu([\omega \upharpoonright k])} \int_{[\omega \upharpoonright k]} f d\mu = \lim_{k \rightarrow \infty} \mathbb{E}_\mu[f \mid \mathcal{F}_k](\omega) = \ell.$$

Now, suppose towards a contradiction that $f(\omega) < \ell$. Then, there is some $q \in \mathbb{Q}$ with $f(\omega) < q < \ell$. In particular, $q > 0$, since f is by assumption non-negative. We will construct another lower semi-computable random variable $\xi : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ with finite expectation such that $\xi(\omega) = \infty$. This will yield a contradiction because, by the characterisation of Martin-Löf randomness in terms of integral tests (Theorem 1.2.14), the existence of such a function contradicts the assumption that ω is μ -Martin-Löf random.

To define ξ , we first define inductively an auxiliary sequence $\{S_n\}_{n \in \mathbb{N}}$ of uniformly c.e. subsets of $2^{<\mathbb{N}} \times \mathbb{N}$. First, let $S_0 = \{\langle \varepsilon, 0 \rangle\}$. For $n \geq 1$, suppose that S_{n-1} has already been defined and is a c.e. set. Uniformly in $\langle \sigma, s \rangle \in S_{n-1}$, define the set $B_{\langle \sigma, s \rangle}$ as

$$\left\{ \tau \in 2^{<\mathbb{N}} : \sigma \sqsubset \tau, |\tau| \geq s, \mu([\tau]) > 0, \frac{1}{\mu([\tau])} \int_{[\tau]} f_s d\mu \leq q, \text{ and } (\exists t) \frac{1}{\mu([\tau])} \int_{[\tau]} f_t d\mu > q \right\}.$$

Clearly, $B_{\langle \sigma, s \rangle}$ is c.e. and, without loss of generality, it can be assumed to be prefix-free. Now, for each $\tau \in B_{\langle \sigma, s \rangle}$, let $t > s$ be least with $\frac{1}{\mu([\tau])} \int_{[\tau]} f_t d\mu > q$ and put $\langle \tau, t \rangle$ in S_n . This ends the construction of the S_n 's.

We now show that none of the S_n 's is empty: in particular, for each $n \in \mathbb{N}$, there is $\langle \tau_n, t_n \rangle \in S_n$ such that $\tau_n \sqsubset \omega$. Let $\tau_0 = \varepsilon$ and $t_0 = 0$. Then, $\langle \varepsilon, 0 \rangle \in S_0$ by definition. For $n \geq 1$, suppose that $\langle \tau_{n-1}, t_{n-1} \rangle \in S_{n-1}$ and $\tau_{n-1} \sqsubset \omega$. Since

$$\lim_{k \rightarrow \infty} \frac{1}{\mu([\omega \upharpoonright k])} \int_{[\omega \upharpoonright k]} f d\mu = \ell > q,$$

there is some M such that, for all $m \geq M$, $\frac{1}{\mu([\omega \upharpoonright m])} \int_{[\omega \upharpoonright m]} f d\mu > q$. Moreover, since $\lim_{n \rightarrow \infty} f_n = f$ everywhere and $f_0 \leq f_1 \leq f_2 \leq \dots$, the Monotone Convergence Theorem

implies that, for any $\sigma \in 2^{<\mathbb{N}}$ with $\mu([\sigma]) > 0$,

$$\lim_{n \rightarrow \infty} \frac{1}{\mu([\sigma])} \int_{[\sigma]} f_n d\mu = \frac{1}{\mu([\sigma])} \int_{[\sigma]} f d\mu.$$

In particular, for all $m \geq M$,

$$\lim_{n \rightarrow \infty} \frac{1}{\mu([\omega \upharpoonright m])} \int_{[\omega \upharpoonright m]} f_n d\mu = \frac{1}{\mu([\omega \upharpoonright m])} \int_{[\omega \upharpoonright m]} f d\mu > q.$$

So, for all $m \geq M$, there is some N such that, for all $n \geq N$, $\frac{1}{\mu([\omega \upharpoonright m])} \int_{[\omega \upharpoonright m]} f_n d\mu > q$. Let $\langle \tau_n, t_n \rangle$ be the pair where τ_n is the shortest string with $\tau_{n-1} \sqsubset \tau_n \sqsubset \omega$, $|\tau_n| \geq t_{n-1}$ and $|\tau_n| \geq M$, while t_n is least with $\frac{1}{\mu([\tau_n])} \int_{[\tau_n]} f_{t_n} d\mu > q$ (such a t_n exists by our previous argument). We also have that

$$\frac{1}{\mu([\tau_n])} \int_{[\tau_n]} f_{t_{n-1}} d\mu \leq q,$$

and, so, $t_n > t_{n-1}$. This is because $|\tau_n| \geq t_{n-1}$, which means that

$$\frac{1}{\mu([\tau_n])} \int_{[\tau_n]} f_{t_{n-1}} d\mu = f_{t_{n-1}}(\omega) \leq f(\omega) < q.$$

Hence, $\langle \tau_n, t_n \rangle \in S_n$.

Next, we proceed to the construction of $\xi : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$. Let $\xi_\varepsilon = q \cdot \chi_{[\varepsilon]}$ and, for each $n \geq 1$ and $\langle \tau, t \rangle$ in S_n , let

$$\xi_\tau = \left(q - \frac{1}{\mu([\tau])} \int_{[\tau]} f_s d\mu \right) \cdot \chi_{[\tau]},$$

where the index s of f_s comes from the pair $\langle \sigma, s \rangle \in S_{n-1}$ which prompted $\langle \tau, t \rangle$ to be put in S_n . The ξ_τ 's (including ξ_ε) are uniformly computable and, since $\frac{1}{\mu([\tau])} \int_{[\tau]} f_s d\mu \leq q$ by construction, non-negative. Then, let

$$\xi = \sum_{n \in \mathbb{N}} \sum_{\langle \tau, t \rangle \in S_n} \xi_\tau.$$

We now show that ξ is indeed lower semi-computable with finite expectation.

To see that ξ is lower semi-computable, let $q' \in \mathbb{Q}$ and let $S_{n,s}$ be the set of all pairs in S_n

that are enumerated within s steps. Then,

$$\begin{aligned} \{\alpha \in 2^{\mathbb{N}} : \xi(\alpha) > q'\} &= \{\alpha \in 2^{\mathbb{N}} : \sum_{n \in \mathbb{N}} \sum_{\langle \tau, t \rangle \in S_n} \xi_{\tau}(\alpha) > q'\} \\ &= \{\alpha \in 2^{\mathbb{N}} : (\exists j)(\exists s) \sum_{n=0}^j \sum_{\langle \tau, t \rangle \in S_{n,s}} \xi_{\tau}(\alpha) > q'\}. \end{aligned}$$

Hence, $\{\alpha \in 2^{\mathbb{N}} : \xi(\alpha) > q'\}$ is a Σ_1^0 class, uniformly in q' . By Proposition 1.2.13, this suffices to establish that ξ is lower semi-computable.

Next, we show that $\int_{2^{\mathbb{N}}} \xi d\mu$ is finite. Since

$$\begin{aligned} \int_{2^{\mathbb{N}}} \sum_{n \in \mathbb{N}} \sum_{\langle \tau, t \rangle \in S_n} \xi_{\tau} d\mu &= \int_{2^{\mathbb{N}}} \lim_{N \rightarrow \infty} \sum_{n=0}^N \sum_{\langle \tau, t \rangle \in S_n} \xi_{\tau} d\mu \\ &= \lim_{N \rightarrow \infty} \int_{2^{\mathbb{N}}} \sum_{n=0}^N \sum_{\langle \tau, t \rangle \in S_n} \xi_{\tau} d\mu \\ &= \lim_{N \rightarrow \infty} \sum_{n=0}^N \sum_{\langle \tau, t \rangle \in S_n} \int_{2^{\mathbb{N}}} \xi_{\tau} d\mu \end{aligned}$$

by the Monotone Convergence Theorem (as well as the linearity of expectations), it is sufficient to prove that

$$\sum_{n=0}^N \sum_{\langle \tau, t \rangle \in S_n} \int_{2^{\mathbb{N}}} \xi_{\tau} d\mu \leq q + \int_{2^{\mathbb{N}}} f d\mu < \infty$$

for every $N \geq 0$. To see that this is indeed the case, note that, for any $n > 0$ and $\langle \tau, t \rangle \in S_n$, the following holds:

$$\begin{aligned} \int_{2^{\mathbb{N}}} \xi_{\tau} d\mu &= \int_{2^{\mathbb{N}}} \left(q - \frac{1}{\mu([\tau])} \int_{[\tau]} f_s d\mu \right) \cdot \chi_{[\tau]} d\mu \\ &\leq \int_{2^{\mathbb{N}}} \left(\frac{1}{\mu([\tau])} \int_{[\tau]} f_t d\mu - \frac{1}{\mu([\tau])} \int_{[\tau]} f_s d\mu \right) \cdot \chi_{[\tau]} d\mu \\ &= \left(\frac{1}{\mu([\tau])} \int_{[\tau]} f_t - f_s d\mu \right) \int_{2^{\mathbb{N}}} \chi_{[\tau]} d\mu \\ &= \int_{[\tau]} f_t - f_s d\mu \end{aligned}$$

$$\leq \int_{[\tau]} f - f_s d\mu.$$

Hence,

$$\begin{aligned} \sum_{\langle \tau, t \rangle \in S_n} \int_{2^{\mathbb{N}}} \xi_\tau d\mu &\leq \sum_{\langle \tau, t \rangle \in S_n} \int_{[\tau]} f - f_s d\mu \\ &\leq \sum_{\langle \sigma, s \rangle \in S_{n-1}} \int_{[\sigma]} f - f_s d\mu, \end{aligned}$$

where the second inequality follows from the fact that $\bigcup_{\tau \in B_{\langle \sigma, s \rangle}} [\tau] \subseteq [\sigma]$. Now, for $N \geq 2$,

$$\begin{aligned} \sum_{n=N-1}^N \sum_{\langle \tau, t \rangle \in S_n} \int_{2^{\mathbb{N}}} \xi_\tau d\mu &= \sum_{\langle \sigma, s \rangle \in S_{N-1}} \int_{2^{\mathbb{N}}} \xi_\sigma d\mu + \sum_{\langle \tau, t \rangle \in S_N} \int_{2^{\mathbb{N}}} \xi_\tau d\mu \\ &\leq \sum_{\langle \sigma, s \rangle \in S_{N-1}} \int_{[\sigma]} (f_s - f_r) d\mu + \sum_{\langle \sigma, s \rangle \in S_{N-1}} \int_{[\sigma]} (f - f_s) d\mu \\ &\leq \sum_{\langle \sigma, s \rangle \in S_{N-1}} \int_{[\sigma]} (f - f_r) d\mu \\ &\leq \sum_{\langle \rho, r \rangle \in S_{N-2}} \int_{[\rho]} (f - f_r) d\mu, \end{aligned}$$

where $\langle \rho, r \rangle$ is the pair in S_{N-2} that prompted $\langle \sigma, s \rangle$ to be put in S_{N-1} , which in turn prompted the pair $\langle \tau, t \rangle$ to be put in S_N . This argument can be repeated for sums starting at $N-2, N-3, \dots, 1$, so that

$$\begin{aligned} \sum_{n=0}^N \sum_{\langle \tau, t \rangle \in S_n} \int_{2^{\mathbb{N}}} \xi_\tau d\mu &\leq \sum_{\langle \tau, t \rangle \in S_0} \int_{2^{\mathbb{N}}} \xi_\tau d\mu + \sum_{\langle \tau, t \rangle \in S_0} \int_{[\tau]} (f - f_0) d\mu \\ &= \int_{2^{\mathbb{N}}} \xi_\varepsilon d\mu + \int_{2^{\mathbb{N}}} (f - f_0) d\mu \\ &\leq q + \int_{2^{\mathbb{N}}} f d\mu \\ &< \infty. \end{aligned}$$

This establishes that the expectation of ξ is finite.

The last thing to show is that $\xi(\omega) = \infty$. Let $\{a_n\}_{n \in \mathbb{N}}$ be the constant sequence where

each $a_n = q - f(\omega) > 0$. We have already argued that, for each n , there is $\langle \tau_n, t_n \rangle \in S_n$ with $\tau_n \sqsubset \omega$. Hence,

$$\begin{aligned} \xi(\omega) &= q + \sum_{n \geq 1} \left(q - \frac{1}{\mu([\tau_n])} \int_{[\tau_n]} f_{t_{n-1}} d\mu \right) \\ &= q + \sum_{n \geq 1} \left(q - f_{t_{n-1}}(\omega) \right) \\ &\geq \sum_{n \in \mathbb{N}} a_n \\ &= \infty, \end{aligned}$$

which contradicts the assumption that ω is μ -Martin-Löf random. \square

We are now ready to prove that the density random data streams guarantee that a computable Bayesian agent's beliefs will converge to the truth across all inductive problems that can be modelled as lower semi-computable random variables with finite expectation.

Theorem 3.2.11 below is a straightforward analogue of Theorem 5.8 by Miyabe et al. [2016], who provide a characterisation of density randomness via the Lebesgue Differentiation Theorem in the context of the uniform measure. The significance of Theorem 3.2.11 chiefly lies with its interpretation. In the context of Lévy's Upward Theorem, this characterisation of density randomness shows that the correspondence between truth-conduciveness and algorithmic randomness is not restricted to the case of Schnorr randomness: it extends to another natural algorithmic randomness notion, which suggests that the connection between algorithmic randomness and convergence to the truth is indeed a robust one.

Theorem 3.2.11. *Let $\omega \in 2^{\mathbb{N}}$. Consider the following statements:*

- (1) ω is μ -density random;
- (2) for all lower semi-computable $f : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ with finite expectation,

$$\lim_{k \rightarrow \infty} \mathbb{E}_{\mu}[f \mid \mathcal{F}_k](\omega) = f(\omega) < \infty.$$

For any computable measure μ , (1) implies (2). When μ is the uniform measure λ , (1)

and (2) are equivalent.

Proof. First, let μ be an arbitrary computable measure and ω a μ -density random sequence. Then, for all k , $\mu([\omega \upharpoonright k]) > 0$. Let f be a lower semi-computable random variable with finite expectation. Without loss of generality, f can be taken to be non-negative. Then, the function d_f , where $d_f(\sigma)$ equals $\frac{\int_{[\sigma]} f d\mu}{\mu([\sigma])}$ if $\mu([\sigma]) > 0$ and is undefined otherwise, is an almost everywhere left-c.e. dyadic μ -martingale. By Definition 1.3.5, $\lim_{k \rightarrow \infty} \mathbb{E}_\mu[f \mid \mathcal{F}_k](\omega)$ exists and is finite. By Lemma 3.2.10 and the fact that μ -density randomness implies μ -Martin-Löf randomness, we can then conclude that $\lim_{k \rightarrow \infty} \mathbb{E}_\mu[f \mid \mathcal{F}_k](\omega) = f(\omega)$.

Now, fix the uniform measure λ . By the previous argument, we clearly have that (1) entails (2). For the other direction, suppose that Condition (2) holds along ω . Then, $f(\omega)$ is finite for all lower semi-computable random variables with finite expectation. Hence, by Theorem 1.2.14, ω is λ -Martin-Löf random. Now, suppose towards a contradiction that ω is not a λ -dyadic density-one point (and, so, by Theorem 1.3.6, that it is not λ -density random). Let \mathcal{C} be a Π_1^0 class such that $\omega \in \mathcal{C}$ but $\underline{\rho}(\mathcal{C} \mid \omega) = \liminf_{k \rightarrow \infty} \frac{\lambda(\mathcal{C} \cap [\omega \upharpoonright k])}{\lambda([\omega \upharpoonright k])} = \liminf_{k \rightarrow \infty} \lambda(\mathcal{C} \mid [\omega \upharpoonright k]) < 1$. Since \mathcal{C} is a Π_1^0 class, its complement $\bar{\mathcal{C}}$ is a Σ_1^0 class and, so, the indicator function $\chi_{\bar{\mathcal{C}}}$ of $\bar{\mathcal{C}}$ is lower semi-computable. Moreover, the expectation of $\chi_{\bar{\mathcal{C}}}$ is clearly finite. Given that $\omega \notin \bar{\mathcal{C}}$, we have that $\chi_{\bar{\mathcal{C}}}(\omega) = 0$. Yet, $\limsup_{k \rightarrow \infty} \frac{\lambda(\bar{\mathcal{C}} \cap [\omega \upharpoonright k])}{\lambda([\omega \upharpoonright k])} = \limsup_{k \rightarrow \infty} \lambda(\bar{\mathcal{C}} \mid [\omega \upharpoonright k]) = \limsup_{k \rightarrow \infty} (1 - \lambda(\mathcal{C} \mid [\omega \upharpoonright k])) > 0$. Therefore, even if the sequence $\{\mathbb{E}_\lambda[\chi_{\bar{\mathcal{C}}} \mid \mathcal{F}_k](\omega)\}_{k \in \mathbb{N}}$ does have a limit, this limit cannot be 0 and, so, it has to be different from $\chi_{\bar{\mathcal{C}}}(\omega)$. This, however, contradicts our initial assumption that Condition (2) holds along ω . □

Almost-everywhere finite computable random variables

The last family of integral tests for randomness that we shall consider here is the collection of integral tests for weak 1-randomness: namely, computable functions that are finite almost everywhere (cf. §1.2.2). The effectivisation of Lévy's Upward Theorem that results from restricting attention to computable almost-everywhere finite random variables yields a characterisation of weak 1-randomness.²² Hence, the weakly 1-random data streams are precisely the ones along which beliefs converge to the truth when the inductive problems

²²Also see Miyabe's characterisation of weak 1-randomness in terms of the Lebesgue Differentiation Theorem [Miyabe, 2013b].

to be solved are computable. As we will see, both directions of this result hold for all computable measures. For the forward direction, we have that, for any computable measure μ , if ω is μ -weakly 1-random, then $\mu([\omega \upharpoonright n]) > 0$ for all n : if there were some k with $\mu([\omega \upharpoonright k]) = 0$, there would in fact be a Σ_1^0 class of μ -measure one—the complement of $[\omega \upharpoonright k]$ —to which ω would fail to belong.

Theorem 3.2.12. *Let μ be a computable measure and $\omega \in 2^{\mathbb{N}}$. Then, the following are equivalent:*

- (1) ω is μ -weakly 1-random;
- (2) for all computable functions $f : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ that are finite almost everywhere,

$$\lim_{k \rightarrow \infty} \mathbb{E}_{\mu}[f \mid \mathcal{F}_k](\omega) = f(\omega) < \infty.$$

Proof. (1) \Rightarrow (2) Let f be a computable random variable that is finite almost everywhere. Let $d : 2^{\mathbb{N}} \times 2^{\mathbb{N}} \rightarrow [0, +\infty)$ denote the canonical metric on $2^{\mathbb{N}}$ given by

$$d(\alpha, \alpha') = \begin{cases} 2^{-n} & \text{if } \alpha \neq \alpha' \text{ and } n \text{ is least with } \alpha(n) \neq \alpha'(n); \\ 0 & \text{if } \alpha = \alpha'. \end{cases}$$

For any $\alpha \in 2^{\mathbb{N}}$ with $f(\alpha) < \infty$ and for any $\epsilon > 0$, there is $\delta > 0$ such that, for all $\alpha' \in 2^{\mathbb{N}}$, if $d(\alpha, \alpha') < \delta$, then $|f(\alpha) - f(\alpha')| < \epsilon$. To see this, take $\alpha \in 2^{\mathbb{N}}$ with $f(\alpha) < \infty$ and $\epsilon > 0$. Let $q, p \in (f(\alpha) - \epsilon, f(\alpha) + \epsilon)$ be rationals with $q < f(\alpha) < p$. Then, $f^{-1}((q, p))$ is a Σ_1^0 class, uniformly in q, p . Thus, $f^{-1}((q, p)) = \bigsqcup_{n \in \mathbb{N}} [\sigma_n]$ and there is an n such that $\alpha \in [\sigma_n]$. Let $\delta = 2^{-|\sigma_n|}$. Then, for all $\alpha' \in [\alpha \upharpoonright |\sigma_n| + 1] \subset [\sigma_n]$, $d(\alpha, \alpha') < 2^{-|\sigma_n|}$ and, for all $\alpha' \notin [\alpha \upharpoonright |\sigma_n| + 1]$, $d(\alpha, \alpha') \geq 2^{-|\sigma_n|}$. Moreover, for all $\alpha' \in [\alpha \upharpoonright |\sigma_n| + 1]$, $|f(\alpha) - f(\alpha')| < \epsilon$. Now, since ω is μ -weakly 1-random, $f(\omega) < \infty$ by Theorem 1.2.17. Fix $\epsilon > 0$. Then, as just shown, there is $\delta > 0$ such that, for all $\alpha \in 2^{\mathbb{N}}$, if $d(\alpha, \omega) < \delta$, then $|f(\alpha) - f(\omega)| < \epsilon$. Fix m with $2^{-m} < \delta$. Then, for all $n \geq m$,

$$\left| \frac{\int_{[\omega \upharpoonright n]} f d\mu}{\mu([\omega \upharpoonright n])} - f(\omega) \right| \leq \frac{\int_{[\omega \upharpoonright n]} |f - f(\omega)| d\mu}{\mu([\omega \upharpoonright n])} < \frac{\int_{[\omega \upharpoonright n]} \epsilon d\mu}{\mu([\omega \upharpoonright n])} = \epsilon.$$

All of the ratios above are well-defined because $\mu([\omega \upharpoonright n]) > 0$ for all n , since ω is μ -weakly

1-random. The first inequality, on the other hand, holds because $f(\omega)$ is a constant. Hence,

$$\lim_{k \rightarrow \infty} \mathbb{E}_\mu[f \mid \mathcal{F}_k](\omega) = \lim_{k \rightarrow \infty} \frac{1}{\mu([\omega \upharpoonright k])} \int_{[\omega \upharpoonright k]} f d\mu = f(\omega) < \infty.$$

(2) \Rightarrow (1) This direction follows immediately from Miyabe's characterisation of μ -weak 1-randomness in terms of integral tests. Suppose that ω is not μ -weakly 1-random. Then, by Theorem 1.2.17, there is a computable random variable f that is finite almost everywhere and such that $f(\omega) = \infty$. This fact, by itself, contradicts (2). \square

3.2.3 Weakly L^1 -computable functions

Next, we focus on a computability concept for functions in L^1 , *weak L^1 -computability*,²³ that is weaker than L^1 -computability, and that is the analogue of the notion of a weakly computable real number [Ambos-Spies et al., 2000].

The main goal of this section is proving the following effectivisation of Lévy's Upward Theorem:

Theorem 3.2.13. *Let $\omega \in 2^{\mathbb{N}}$. Consider the following statements:*

- (1) ω is μ -density random;
- (2) for all weakly L^1 -computable functions $f : \subseteq 2^{\mathbb{N}} \rightarrow \mathbb{R}$ with witness $\{f_n\}_{n \in \mathbb{N}}$, $\hat{f}(\omega)$ is defined, and

$$\lim_{k \rightarrow \infty} \mathbb{E}_\mu[f \mid \mathcal{F}_k](\omega) = \hat{f}(\omega).$$

For any computable measure μ , (1) entails (2). When μ is the uniform measure λ , (1) and (2) are equivalent.

We begin by defining weakly computable real numbers.

Definition 3.2.14 (Weakly computable real number). *A real number r_0 is weakly computable if there are two left-c.e. real numbers r_1 and r_2 such that $r_0 = r_1 - r_2$.*

²³Weak L^1 -computability was introduced by Miyabe [2013a]. Our presentation of weakly L^1 -computable functions is slightly different from Miyabe's, and it mirrors the definition of L^1 -computable functions given in §3.2.1.

Theorem 3.2.15 (Ambos-Spies et al. [2000]). *A real number r is weakly computable if and only if there is a computable sequence $\{q_n\}_{n \in \mathbb{N}}$ of rationals such that $q_n \rightarrow r$ and $\sum_{n \in \mathbb{N}} |q_{n+1} - q_n| < \infty$.*

Now, the definition of a weakly L^1 -computable function mirrors the characterisation of weakly computable reals given in Theorem 3.2.15:

Definition 3.2.16 (Weak L^1 -computability). *A function $f : \subseteq 2^{\mathbb{N}} \rightarrow \mathbb{R}$ is weakly L^1 -computable, relative to a computable measure μ , if there is a computable sequence $\{f_n\}_{n \in \mathbb{N}}$ of rational-valued step functions such that $f_n \rightarrow f$ in L^1 and $\sum_{n \in \mathbb{N}} \|f_{n+1} - f_n\|_1 < \infty$ (i.e., the sequence $\{f_n\}_{n \in \mathbb{N}}$ has bounded L^1 -variation).*

Weakly L^1 -computable functions and their properties

In what follows, we will discuss some of the basic properties of weak L^1 -computability and see how it relates to the notion of L^1 -computability presented in §3.2.1. All of the results discussed here hold for all computable measures.

First, we show that, in the definition of a weakly L^1 -computable function, the computable sequence of rational-valued step functions that serves as a witness may be replaced, without loss of generality, by a computable sequence of real-valued step functions, as long as the reals involved are uniformly computable.²⁴

Observation 3.2.17. *Let $f : \subseteq 2^{\mathbb{N}} \rightarrow \mathbb{R}$ be such that there is a computable sequence $f_n = \sum_{i=1}^{k_n} r_i \chi_{[\tau_i]}$ of real-valued step functions, where the r_i 's are uniformly computable, with $f_n \rightarrow f$ in L^1 and $\sum_{n \in \mathbb{N}} \|f_{n+1} - f_n\|_1 < \infty$. Then, f is weakly L^1 -computable.*

Proof. For each n and each r_i , with $1 \leq i \leq k_n$, there is a computable sequence of rationals $q_{i,0}, q_{i,1}, \dots \rightarrow r_i$ such that $|q_{i,n} - r_i| \leq 2^{-n}$ for all n . For each n , define $g_n : 2^{\mathbb{N}} \rightarrow \mathbb{Q}$ in terms of f_n as follows: $g_n = \sum_{i=1}^{k_n} q_{i,n} \chi_{[\tau_i]}$. Then, $\{g_n\}_{n \in \mathbb{N}}$ is a computable sequence of rational-valued step functions. For every $n \in \mathbb{N}$, we then have that

$$\int_{2^{\mathbb{N}}} |g_n - f_n| d\mu = \int_{2^{\mathbb{N}}} \left| \sum_{i=1}^{k_n} (q_{i,n} - r_i) \cdot \chi_{[\tau_i]} \right| d\mu$$

²⁴The same also holds for L^1 -computable functions.

$$\begin{aligned}
&\leq \int_{2^{\mathbb{N}}} \sum_{i=1}^{k_n} |q_{i,n} - r_i| \cdot \chi_{[\tau_i]} d\mu \\
&\leq \int_{2^{\mathbb{N}}} 2^{-n} d\mu \\
&= 2^{-n}.
\end{aligned}$$

Let $\epsilon > 0$. Since $f_n \rightarrow f$ in L^1 , there is some N such that, for all $n \geq N$, $\|f_n - f\|_1 < \frac{\epsilon}{2}$. Let K be such that $2^{-K} < \frac{\epsilon}{2}$. Then, for all $k \geq K$, $\|g_k - f_k\|_1 \leq 2^{-K} < \frac{\epsilon}{2}$. Let $M = \max\{N, K\}$. Then, for all $m \geq M$, $\|g_m - f\|_1 \leq \|g_m - f_m\|_1 + \|f_m - f\|_1 < \epsilon$. Hence, $g_n \rightarrow f$ in L^1 . Moreover,

$$\sum_{n \in \mathbb{N}} \|g_{n+1} - g_n\|_1 \leq \sum_{n \in \mathbb{N}} \|g_{n+1} - f_{n+1}\|_1 + \sum_{n \in \mathbb{N}} \|f_{n+1} - f_n\|_1 + \sum_{n \in \mathbb{N}} \|f_n - g_n\|_1 < \infty.$$

Hence, $\{g_n\}_{n \in \mathbb{N}}$ witnesses the weak L^1 -computability of f . \square

L^1 -computability can be characterised in a way analogous to the above definition of weak L^1 -computability. The following also establishes that L^1 -computability entails weak L^1 -computability.

Observation 3.2.18. *A function $f : \subseteq 2^{\mathbb{N}} \rightarrow \mathbb{R}$ is L^1 -computable if and only if there is a computable sequence $\{f_n\}_{n \in \mathbb{N}}$ of rational-valued step functions such that $f_n \rightarrow f$ in L^1 and $\sum_{n \in \mathbb{N}} \|f_{n+1} - f_n\|_1$ is finite and computable.*

Proof. (\Rightarrow) Let $\{f_n\}_{n \in \mathbb{N}}$ be a witness to the L^1 -computability of f . For each $n \geq 0$, since f_{n+1} and f_n are both rational-valued step functions, so are $f_{n+1} - f_n$ and $|f_{n+1} - f_n|$. Moreover, the expectations of all these functions are computable, uniformly in n . Further, to see that $\sum_{m \in \mathbb{N}} \|f_{m+1} - f_m\|_1$ is a computable real, note that, for all $n \geq 0$, we have

$$\left| \sum_{m \in \mathbb{N}} \|f_{m+1} - f_m\|_1 - \sum_{m=0}^{n+1} \|f_{m+1} - f_m\|_1 \right| = \sum_{m=n+2}^{\infty} \|f_{m+1} - f_m\|_1 \leq 2^{-n},$$

where the inequality follows from Proposition 3.2.3(ii). Therefore, $\sum_{n \in \mathbb{N}} \|f_{n+1} - f_n\|_1$ is computable, in addition to being finite.

(\Leftarrow) Since the sum $\sum_{n \in \mathbb{N}} \|f_{n+1} - f_n\|_1$ is finite and computable, choose an increasing computable sequence $\{\alpha_n\}_{n \in \mathbb{N}}$ of natural numbers such that $\sum_{m=\alpha_n}^{\infty} \|f_{m+1} - f_m\|_1 \leq 2^{-n}$

for each $n \geq 0$. Then, for all $n \geq 0$ and $k \geq n$, we have that

$$\|f_{\alpha_k} - f_{\alpha_n}\|_1 \leq \sum_{m=\alpha_n}^{\alpha_k-1} \|f_{m+1} - f_m\|_1 \leq \sum_{m=\alpha_n}^{\infty} \|f_{m+1} - f_m\|_1 \leq 2^{-n}.$$

Therefore, $\{g_n\}_{n \in \mathbb{N}}$, where $g_n = f_{\alpha_n}$ is a fast computable L^1 -Cauchy sequence. Moreover, since $f_n \rightarrow f$ in L^1 , we also have that $g_n \rightarrow f$ in L^1 . Hence, by Proposition 3.2.3(i), $\{g_{n+1}\}_{n \in \mathbb{N}}$ is a witness to the L^1 -computability of f . \square

In fact, as its name suggests, weak L^1 -computability is strictly weaker than L^1 -computability.

Observation 3.2.19. *There is a weakly L^1 -computable function that is not L^1 -computable.*

To prove Observation 3.2.19, we will make use of the following lemma:

Lemma 3.2.20. *Let $f : \subseteq 2^{\mathbb{N}} \rightarrow \mathbb{R}$ be an L^1 -computable function with witness $\{f_n\}_{n \in \mathbb{N}}$. Then, for all $\sigma \in 2^{<\mathbb{N}}$, $\int_{[\sigma]} f d\mu$ is a computable real number.*

Proof of Lemma 3.2.20. Take $\sigma \in 2^{<\mathbb{N}}$ and $n \in \mathbb{N}$. Then,

$$\int_{[\sigma]} f_n d\mu = \int_{[\sigma]} \sum_{i=1}^{k_n} q_i \chi_{[\tau_i]} d\mu = \sum_{i=1}^{k_n} q_i \cdot \mu([\sigma] \cap [\tau_i]).$$

Hence, $\int_{[\sigma]} f_n d\mu$ is computable, uniformly in σ and n , and $\{\int_{[\sigma]} f_n d\mu\}_{n \in \mathbb{N}}$ is a sequence of uniformly computable reals. Moreover, the following holds:

$$\left| \int_{[\sigma]} f_n d\mu - \int_{[\sigma]} f d\mu \right| = \left| \int_{[\sigma]} (f_n - f) d\mu \right| \leq \int_{[\sigma]} |f_n - f| d\mu \leq \int_{2^{\mathbb{N}}} |f_n - f| d\mu \leq 2^{-n}.$$

Hence, $\int_{[\sigma]} f d\mu$ is indeed computable. \square

Proof of Observation 3.2.19. Let $\{\mathcal{U}_n\}_{n \in \mathbb{N}}$ be a μ -Solovay test (cf. Theorem 1.2.5) that is not a total μ -Solovay test (cf. Theorem 1.2.7). Then, each \mathcal{U}_n can be written as $\bigsqcup_{i \in \mathbb{N}} [\tau_{n,i}]$, for some uniformly computable sequence of cylinders. Let $g = \sum_{n \in \mathbb{N}} \chi_{\mathcal{U}_n}$. Then, g is an integral test for μ -Martin-Löf randomness—i.e., it is lower semi-computable and its expectation is finite. To see this, first note that, by the Monotone Convergence Theorem, $\int_{2^{\mathbb{N}}} g d\mu = \sum_{n \in \mathbb{N}} \mu(\mathcal{U}_n)$. And, since $\{\mathcal{U}_n\}_{n \in \mathbb{N}}$ is a μ -Solovay test, $\sum_{n \in \mathbb{N}} \mu(\mathcal{U}_n) < \infty$. To see

that g is also lower semi-computable, observe that, for any $q \in \mathbb{Q}$,

$$\begin{aligned} \{\omega \in 2^{\mathbb{N}} : g(\omega) > q\} &= \left\{ \omega \in 2^{\mathbb{N}} : \sum_{n \in \mathbb{N}} \chi_{\mathcal{U}_n}(\omega) > q \right\} \\ &= \left\{ \omega \in 2^{\mathbb{N}} : (\exists m) \sum_{n=0}^m \sum_{i=0}^m \chi_{[\tau_{n,i}]}(\omega) > q \right\}. \end{aligned}$$

Hence, $\{\omega \in 2^{\mathbb{N}} : g(\omega) > q\}$ is a Σ_1^0 class, uniformly in q , and so, by Proposition 1.2.13, g is lower semi-computable. Now, let h be such that $h(\omega)$ equals $g(\omega)$ if $g(\omega) < \infty$ and is undefined otherwise. We will show that h is weakly L^1 -computable but not L^1 -computable. Since $h = g$ on all μ -Martin-Löf random sequences, $h = g$ μ -almost everywhere and, so, $\int_{2^{\mathbb{N}}} h \, d\mu = \int_{2^{\mathbb{N}}} g \, d\mu$. Moreover, since $\{\mathcal{U}_n\}_{n \in \mathbb{N}}$ is not a total μ -Solovay test, $\sum_{n \in \mathbb{N}} \mu(\mathcal{U}_n)$ is not computable. Hence, $\int_{2^{\mathbb{N}}} h \, d\mu$ is finite but not computable, which, by Lemma 3.2.20, implies that h is not an L^1 -computable function. To show that h is weakly L^1 -computable, for each m , let $g_m = \sum_{n=0}^m \sum_{i=0}^m \chi_{[\tau_{n,i}]}$. Then, $\{g_m\}_{m \in \mathbb{N}}$ is a computable sequence of non-negative non-decreasing rational-valued step functions and, for all $\omega \in 2^{\mathbb{N}}$, $\lim_{m \rightarrow \infty} g_m(\omega) = g(\omega)$. By the Monotone Convergence Theorem, we then have that $g_m \rightarrow g$ in L^1 . Moreover, since $h = g$ μ -almost everywhere, $g_n \rightarrow h$ in L^1 , as well. Lastly, we have that

$$\begin{aligned} \sum_{m \in \mathbb{N}} \int_{2^{\mathbb{N}}} |g_{m+1} - g_m| \, d\mu &= \sum_{m \in \mathbb{N}} \int_{2^{\mathbb{N}}} (g_{m+1} - g_m) \, d\mu \\ &= \lim_{k \rightarrow \infty} \sum_{m=0}^k \left(\int_{2^{\mathbb{N}}} g_{m+1} \, d\mu - \int_{2^{\mathbb{N}}} g_m \, d\mu \right) \\ &= \left(\lim_{k \rightarrow \infty} \int_{2^{\mathbb{N}}} g_{k+1} \, d\mu \right) - \int_{2^{\mathbb{N}}} g_0 \, d\mu \\ &= \int_{2^{\mathbb{N}}} g \, d\mu - \int_{2^{\mathbb{N}}} g_0 \, d\mu \\ &\leq \int_{2^{\mathbb{N}}} g \, d\mu \\ &< \infty, \end{aligned}$$

where the first identity follows from the fact that the g_m 's are non-decreasing, the second identity from the linearity of expectation, and the first inequality from the fact that g_0 is non-negative and $g_0 \leq g$ everywhere. This establishes that h is weakly L^1 -computable. \square

Even though the effectivisation of Lévy's Upward Theorem in terms of weakly L^1 -computable functions does not yield Martin-Löf randomness, being Martin-Löf random is equivalent to guaranteeing the convergence of any computable sequence of rational-valued step functions witnessing the weak L^1 -computability of a function.

Proposition 3.2.21. *Let μ be a computable measure and $\omega \in 2^{\mathbb{N}}$. Then, the following are equivalent:*

- (1) ω is μ -Martin-Löf random;
- (2) for all weakly L^1 -computable functions $f : \subseteq 2^{\mathbb{N}} \rightarrow \mathbb{R}$ with witness $\{f_n\}_{n \in \mathbb{N}}$, $\lim_{n \rightarrow \infty} f_n(\omega)$ exists and is finite.

Proof. (1) \Rightarrow (2) Let $f : \subseteq 2^{\mathbb{N}} \rightarrow \mathbb{R}$ be a weakly L^1 -computable function with witness $\{f_n\}_{n \in \mathbb{N}}$. Then, $\sum_{n \in \mathbb{N}} \|f_{n+1} - f_n\|_1 < \infty$. Define $\xi : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ as $\xi(\alpha) = \sum_{n \in \mathbb{N}} |f_{n+1}(\alpha) - f_n(\alpha)| = \lim_{k \rightarrow \infty} \sum_{n=0}^k |f_{n+1}(\alpha) - f_n(\alpha)|$ for all $\alpha \in 2^{\mathbb{N}}$. Then, by the Monotone Convergence Theorem, ξ has finite expectation. Moreover, ξ is lower semi-computable, since it is the pointwise limit of a computable sequence of non-negative non-decreasing rational-valued step functions. Hence, ξ is an integral test for μ -Martin-Löf randomness. Theorem 1.2.14 then entails that $\xi(\omega) < \infty$, since ω is by assumption μ -Martin-Löf random. In turn, this implies that $\{f_n(\omega)\}_{n \in \mathbb{N}}$ is Cauchy. For, suppose that $\epsilon > 0$. Choose N such that $\sum_{n \geq N} |f_{n+1}(\omega) - f_n(\omega)| < \epsilon$. Then, for all $m \geq n \geq N$, we have that

$$|f_m(\omega) - f_n(\omega)| \leq \sum_{k=n}^m |f_{k+1}(\omega) - f_k(\omega)| < \epsilon.$$

Thus, $\{f_n(\omega)\}_{n \in \mathbb{N}}$ is Cauchy and, so, $\lim_{n \rightarrow \infty} f_n(\omega)$ exists and is finite.²⁵

(2) \Rightarrow (1) Suppose that ω is not μ -Martin-Löf random. Then, there is a μ -Martin-Löf test $\{\mathcal{U}_n\}_{n \in \mathbb{N}}$ such that $\omega \in \bigcap_{n \in \mathbb{N}} \mathcal{U}_n$ and each $\mathcal{U}_n = \bigsqcup_{i \in \mathbb{N}} [\tau_{n,i}]$. Let $g = \sum_{n \in \mathbb{N}} \chi_{\mathcal{U}_n}$ and, for each $n \geq 0$, define $g_n = \sum_{m=0}^n \sum_{i=0}^m \chi_{[\tau_{m,i}]}$. Then, $\{g_n\}_{n \in \mathbb{N}}$ is a computable sequence of non-negative non-decreasing rational-valued step functions. Since $\lim_{n \rightarrow \infty} g_n(\alpha) = g(\alpha)$ for all $\alpha \in 2^{\mathbb{N}}$, $g_n \rightarrow g$ in L^1 by the Monotone Convergence Theorem. Now, let f be such that $f(\alpha)$ equals $g(\alpha)$ if $g(\alpha) < \infty$ and is undefined otherwise. Then, $f = g$ on all μ -Martin-Löf

²⁵This direction can also be proved by appealing to Theorem 3.2.25.

random sequences and, so, $g_n \rightarrow f$ in L^1 . By the same argument used in the proof of Observation 3.2.19,

$$\sum_{n \in \mathbb{N}} \int_{2^{\mathbb{N}}} |g_{n+1} - g_n| d\mu \leq \lim_{n \rightarrow \infty} \int_{2^{\mathbb{N}}} g_n d\mu = \int_{2^{\mathbb{N}}} g d\mu = \sum_{n \in \mathbb{N}} \mu(\mathcal{U}_n) < \infty.$$

Therefore, $\{g_n\}_{n \in \mathbb{N}}$ is a witness to the weak L^1 -computability of f . Moreover, $\omega \in \bigcap_{n \in \mathbb{N}} \mathcal{U}_n$ implies that $\lim_{n \rightarrow \infty} g_n(\omega) = g(\omega) = \infty$. \square

Given a weakly L^1 -computable function $f : \subseteq 2^{\mathbb{N}} \rightarrow \mathbb{R}$ with witness $\{f_n\}_{n \in \mathbb{N}}$, let the function $\hat{f} : \subseteq 2^{\mathbb{N}} \rightarrow \mathbb{R}$ be defined as $\hat{f}(\omega) = \lim_{n \rightarrow \infty} f_n(\omega)$ if this limit exists and is finite, and let it be undefined otherwise. Then, it is easy to see that the following analogue of Lemma 3.2.7 holds for all computable measures:

Observation 3.2.22. *Let $f : \subseteq 2^{\mathbb{N}} \rightarrow \mathbb{R}$ be weakly L^1 -computable with witness $\{f_n\}_{n \in \mathbb{N}}$. Then, $\|\hat{f} - f\|_1 = 0$.*

Proof. By Proposition 3.2.21, $\lim_{n \rightarrow \infty} f_n(\omega)$ exists and is finite for all μ -Martin-Löf random sequences $\omega \in 2^{\mathbb{N}}$. Therefore, $\{f_n\}_{n \in \mathbb{N}}$ converges pointwise—and, thus, \hat{f} is defined— μ -almost everywhere. Since $\sum_{n \in \mathbb{N}} \|f_{n+1} - f_n\|_1 < \infty$, $f_n \rightarrow \hat{f}$ in L^1 . But then, since $f_n \rightarrow f$ in L^1 , too, we have that $\|\hat{f} - f\|_1 = 0$. \square

Finally, the following holds for μ -Martin-Löf random sequences:

Observation 3.2.23. *Let $f : \subseteq 2^{\mathbb{N}} \rightarrow \mathbb{R}$ and $g : \subseteq 2^{\mathbb{N}} \rightarrow \mathbb{R}$ be weakly L^1 -computable functions. Then, $\|f - g\|_1 = 0$ if and only if \hat{f} and \hat{g} are defined and equal on all μ -Martin-Löf random sequences.*

Proof. (\Rightarrow) Since f and g are both weakly L^1 -computable, there are computable sequences $\{f_n\}_{n \in \mathbb{N}}$ and $\{g_n\}_{n \in \mathbb{N}}$ of rational-valued step functions witnessing the weak L^1 -computability of f and g , respectively. Since $\|f - g\|_1 = 0$, $\lim_{n \rightarrow \infty} \|f_n - g_n\|_1 = 0$. For each k , computably find $n_k \geq k$ such that $\|f_{n_k} - g_{n_k}\|_1 \leq 2^{-k}$. Then, define $\xi : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ as $\xi(\omega) = \sum_{k \in \mathbb{N}} |f_{n_k}(\omega) - g_{n_k}(\omega)|$ for all $\omega \in 2^{\mathbb{N}}$. Clearly, ξ is lower semi-computable and has finite expectation. Hence, ξ is an integral test for μ -Martin-Löf randomness, and thus, by Theorem 1.2.14, $\xi(\omega) < \infty$ for all μ -Martin-Löf random sequences $\omega \in 2^{\mathbb{N}}$. By Proposition 3.2.21, both $\lim_{n \rightarrow \infty} f_n(\omega)$ and $\lim_{n \rightarrow \infty} g_n(\omega)$ exist and are finite for all μ -Martin-Löf random

$\omega \in 2^{\mathbb{N}}$. Hence, for all μ -Martin-Löf random $\omega \in 2^{\mathbb{N}}$, $\lim_{k \rightarrow \infty} f_{n_k}(\omega) = \lim_{n \rightarrow \infty} f_n(\omega) < \infty$ and $\lim_{k \rightarrow \infty} g_{n_k}(\omega) = \lim_{n \rightarrow \infty} g_n(\omega) < \infty$. Let ω be μ -Martin-Löf random: we show that $\hat{f}(\omega) = \lim_{n \rightarrow \infty} f_n(\omega) = \lim_{n \rightarrow \infty} g_n(\omega) = \hat{g}(\omega)$. Let $\epsilon > 0$. Since $\xi(\omega) = \sum_{k \in \mathbb{N}} |f_{n_k}(\omega) - g_{n_k}(\omega)| < \infty$, there is some K such that $\sum_{k > K} |f_{n_k}(\omega) - g_{n_k}(\omega)| < \epsilon$. Hence, for all $k > K$, $|f_{n_k}(\omega) - g_{n_k}(\omega)| < \epsilon$, which implies that $\lim_{k \rightarrow \infty} f_{n_k}(\omega) - g_{n_k}(\omega) = \lim_{k \rightarrow \infty} f_{n_k}(\omega) - \lim_{k \rightarrow \infty} g_{n_k}(\omega) = 0$. Hence, $\lim_{n \rightarrow \infty} f_n(\omega) = \lim_{n \rightarrow \infty} g_n(\omega)$: i.e., $\hat{f}(\omega) = \hat{g}(\omega)$.

(\Leftarrow) Suppose that $\hat{f}(\omega) = \hat{g}(\omega)$ for all μ -Martin-Löf random $\omega \in 2^{\mathbb{N}}$. Then, $\int_{2^{\mathbb{N}}} |\hat{f} - \hat{g}| d\mu = 0$. By Observation 3.2.22, $\int_{2^{\mathbb{N}}} |\hat{f} - f| d\mu = 0$ and $\int_{2^{\mathbb{N}}} |\hat{g} - g| d\mu = 0$. Hence,

$$\int_{2^{\mathbb{N}}} |f - g| d\mu \leq \int_{2^{\mathbb{N}}} |f - \hat{f}| d\mu + \int_{2^{\mathbb{N}}} |\hat{f} - \hat{g}| d\mu + \int_{2^{\mathbb{N}}} |\hat{g} - g| d\mu = 0,$$

which concludes the proof. \square

Weakly L^1 -computable functions and density randomness

Just as in the context of lower semi-computable random variables with finite expectation (see §3.2.2), Martin-Löf randomness is too weak to be characterisable via the effectivisation of Lévy's Upward Theorem in terms of weakly L^1 -computable functions. To see this, take the uniform measure λ . Then, the following holds:

Observation 3.2.24. *There is a λ -Martin-Löf random sequence ω and a weakly L^1 -computable function $f : \subseteq 2^{\mathbb{N}} \rightarrow \mathbb{R}$ such that $\lim_{k \rightarrow \infty} \mathbb{E}_{\mu}[f | \mathcal{F}_k](\omega)$ does not exist.*

Proof. Just as in the proof of Proposition 3.2.9, consider the indicator function $\chi_{\mathcal{U}}$ of the set $\mathcal{U} = \{\alpha \in 2^{\mathbb{N}} : \alpha <_L \omega\}$, where ω is a left-c.e. λ -Martin-Löf random sequence. Let $\mathcal{C} \subseteq 2^{<\mathbb{N}}$ denote the c.e. set of strings that are lexicographically prior to ω . Without loss of generality, \mathcal{C} can be taken to be prefix-free. For each $n \geq 0$, let $f_n = \sum_{i=0}^n \chi_{[\sigma_i]}$, where $\sigma_0, \dots, \sigma_n$ are the first $n + 1$ strings enumerated into \mathcal{C} . Then, $\{f_n\}_{n \in \mathbb{N}}$ is a computable sequence of rational-valued step functions. Moreover, $f_n \rightarrow \chi_{\mathcal{U}}$ in L^1 , as

$$\lim_{n \rightarrow \infty} \int_{2^{\mathbb{N}}} |f_n - \chi_{\mathcal{U}}| d\lambda = \lim_{n \rightarrow \infty} \int_{2^{\mathbb{N}}} \sum_{i=n+1}^{\infty} \chi_{[\sigma_i]} d\lambda = \lim_{n \rightarrow \infty} \sum_{i=n+1}^{\infty} \lambda([\sigma_i]) = 0,$$

and

$$\sum_{n \in \mathbb{N}} \int_{2^{\mathbb{N}}} |f_{n+1} - f_n| d\lambda = \sum_{n \in \mathbb{N}} \int_{2^{\mathbb{N}}} \chi_{[\sigma_{n+1}]} d\lambda = \sum_{n \in \mathbb{N}} \lambda([\sigma_{n+1}]) < \lambda(\mathcal{U}) < \infty.$$

However, just as in the case of Proposition 3.2.9, $\lim_{k \rightarrow \infty} \mathbb{E}[\chi_{\mathcal{U}} \mid \mathcal{F}_k](\omega)$ does not exist because ω , by virtue of being λ -Martin-Löf random, is Borel normal. \square

Now, say that a function $h : \subseteq 2^{\mathbb{N}} \rightarrow \mathbb{R}$ is the difference between two integral tests for Martin-Löf randomness if there are two integral tests for Martin-Löf randomness g and ℓ such that $h(\omega) = g(\omega) - \ell(\omega)$ if $g(\omega) < \infty$ and $\ell(\omega) < \infty$, and $h(\omega)$ is undefined otherwise. The proof of Theorem 3.2.13 relies on the following result by Miyabe, which holds for all computable measures.²⁶

Theorem 3.2.25 (Miyabe [2013b]). (i) *Let h be the difference between two integral tests for μ -Martin-Löf randomness g and ℓ . Then, there is a weakly L^1 -computable function f with witness $\{f_n\}_{n \in \mathbb{N}}$ such that $\hat{f}(\omega) = h(\omega)$ for all μ -Martin-Löf random sequences ω .*

(ii) *Let f be a weakly L^1 -computable function with witness $\{f_n\}_{n \in \mathbb{N}}$. Then, there is a function h that is the difference between two integral tests for μ -Martin-Löf randomness such that $h(\omega) = \hat{f}(\omega)$ for all μ -Martin-Löf random sequences ω .*

We are now ready to prove Theorem 3.2.13, which shows that, for weakly L^1 -computable inductive problems, density randomness guarantees that the beliefs of a computable Bayesian agent will eventually align with the truth.

Theorem 3.2.13. *Let $\omega \in 2^{\mathbb{N}}$. Consider the following statements:*

- (1) *ω is μ -density random;*
- (2) *for all weakly L^1 -computable functions $f : \subseteq 2^{\mathbb{N}} \rightarrow \mathbb{R}$ with witness $\{f_n\}_{n \in \mathbb{N}}$, $\hat{f}(\omega)$ is defined, and*

$$\lim_{k \rightarrow \infty} \mathbb{E}_{\mu}[f \mid \mathcal{F}_k](\omega) = \hat{f}(\omega).$$

For any computable measure μ , (1) entails (2). When μ is the uniform measure λ , (1) and (2) are equivalent.

²⁶This is the analogue of Theorem 3.2.6. Note that Observation 3.2.24 above can also be proved by appealing to this result and Proposition 3.2.9.

Proof. First, let μ be a computable measure and $\omega \in 2^{\mathbb{N}}$ a μ -density random sequence. Let f be a weakly L^1 -computable function with witness $\{f_n\}_{n \in \mathbb{N}}$. Since ω is also μ -Martin-Löf random, by Proposition 3.2.21, $\lim_{n \rightarrow \infty} f_n(\omega)$ exists and is finite. Hence, $\hat{f}(\omega)$ is defined. Moreover, $\mu([\omega \upharpoonright k]) > 0$ for all k . Now, suppose towards a contradiction that the sequence $\{\mathbb{E}_\mu[f \mid \mathcal{F}_k](\omega)\}_{k \in \mathbb{N}}$ does not have a limit. By Theorem 3.2.25, there is a function h that is the difference between two integral tests for μ -Martin-Löf randomness g and ℓ such that h and \hat{f} agree on all μ -Martin-Löf random sequences. Then, in particular, they agree on ω . Now, for each $k \geq 0$, $\mathbb{E}_\mu[f \mid \mathcal{F}_k](\omega) = \mathbb{E}_\mu[\hat{f} \mid \mathcal{F}_k](\omega) = \mathbb{E}_\mu[h \mid \mathcal{F}_k](\omega) = \mathbb{E}_\mu[g \mid \mathcal{F}_k](\omega) - \mathbb{E}_\mu[\ell \mid \mathcal{F}_k](\omega)$. Thus, if $\{\mathbb{E}_\mu[f \mid \mathcal{F}_k](\omega)\}_{k \in \mathbb{N}}$ does not have a limit, neither does the sequence $\{\mathbb{E}_\mu[h \mid \mathcal{F}_k](\omega)\}_{k \in \mathbb{N}}$. Hence, either $\{\mathbb{E}_\mu[g \mid \mathcal{F}_k](\omega)\}_{k \in \mathbb{N}}$ or $\{\mathbb{E}_\mu[\ell \mid \mathcal{F}_k](\omega)\}_{k \in \mathbb{N}}$ fails to have a limit. Whichever of these two sequences does not have a limit contradicts Theorem 3.2.11. Therefore, $\lim_{k \rightarrow \infty} \mathbb{E}_\mu[f \mid \mathcal{F}_k](\omega)$ does exist. Moreover,

$$\begin{aligned}
\lim_{k \rightarrow \infty} \mathbb{E}_\mu[f \mid \mathcal{F}_k](\omega) &= \lim_{k \rightarrow \infty} \mathbb{E}_\mu[h \mid \mathcal{F}_k](\omega) \\
&= \lim_{k \rightarrow \infty} \mathbb{E}_\mu[g - \ell \mid \mathcal{F}_k](\omega) \\
&= \lim_{k \rightarrow \infty} \mathbb{E}_\mu[g \mid \mathcal{F}_k](\omega) - \lim_{k \rightarrow \infty} \mathbb{E}_\mu[\ell \mid \mathcal{F}_k](\omega) \\
&= g(\omega) - \ell(\omega) \\
&= h(\omega) \\
&= \hat{f}(\omega),
\end{aligned}$$

where the third from last identity follows from Theorem 3.2.11.

Now, fix the uniform measure λ . The previous argument establishes that (1) entails (2). So, assume that Condition (2) holds along ω . Then, by Proposition 3.2.21, ω is λ -Martin-Löf random. Now, suppose towards a contradiction that ω is not a λ -dyadic density-one point. Then, there is a Π_1^0 class \mathcal{C} such that $\omega \in \mathcal{C}$ and $\liminf_{k \rightarrow \infty} \frac{\lambda(\mathcal{C} \cap [\omega \upharpoonright k])}{\lambda([\omega \upharpoonright k])} < 1$. As in the proof of Theorem 3.2.11, the indicator function $\chi_{\overline{\mathcal{C}}}$ is an integral test for λ -Martin-Löf randomness. Hence, by Theorem 3.2.25, there is a weakly L^1 -computable function f with witness $\{f_n\}_{n \in \mathbb{N}}$ such that $\hat{f}(\alpha) = \chi_{\overline{\mathcal{C}}}(\alpha)$ for all λ -Martin-Löf random $\alpha \in 2^{\mathbb{N}}$. In particular, $\hat{f}(\omega) = \chi_{\overline{\mathcal{C}}}(\omega) = 0$. Moreover, for all k , $\mathbb{E}_\lambda[f \mid \mathcal{F}_k](\omega) = \mathbb{E}_\lambda[\chi_{\overline{\mathcal{C}}} \mid \mathcal{F}_k](\omega)$. Hence, even if the sequence $\{\mathbb{E}_\lambda[f \mid \mathcal{F}_k](\omega)\}_{k \in \mathbb{N}}$ does have a limit, this limit cannot be 0. This, however,

contradicts our initial assumption that Condition (2) holds along ω . \square

Thus, in addition to being the property that characterises the truth-conducive data streams when the inductive problem at hand is a lower semi-computable random variable with finite expectation, density randomness also guarantees convergence to the truth when the quantities to be estimated are weakly L^1 -computable.

This result concludes our present investigation of the connections between algorithmic randomness and Lévy's Upward Martingale Convergence Theorem.

3.3 Discussion

In what follows, we offer a discussion of the results presented above, as well as some of their philosophical ramifications.

3.3.1 Why is algorithmic randomness truth-conducive?

The characterisation results in this chapter establish a robust connection between algorithmic randomness and successful Bayesian learning: for many natural classes of effective inductive problems, the algorithmically random data streams are exactly the ones that ensure that the beliefs of a computable (open-minded) Bayesian agent will asymptotically converge to the truth.

The finding that the most irregular data streams turn out to coincide with the sequences of observations that lead to successful learning might, at first, appear counter-intuitive. After all, inductive learning is often equated with being able to successfully project into the future the patterns detected in past observational data. So, when trying to identify the conditions under which inductive success is attainable, one might conjecture that observing an algorithmically random data stream is the worst possible evidential situation to be finding oneself in.

Yet, as already argued in both Chapter 1 and Chapter 2, patterns and uniformity come in a great many guises. The algorithmically random data streams do not possess any identifying patterns, they do not stand out, and, as evinced by the unpredictability paradigm, observing their initial segments does not provide any useful information for consistently guessing what the next observations are going to be. However, the algorithmically random

data streams are unruly, irregular, and lawless only when considered locally. When considered globally, they are, to the contrary, regular and lawful: they satisfy all effectively specifiable statistical laws—where the type of effective laws involved varies depending on the particular algorithmic randomness notion in question, as well as the computable prior with respect to which randomness is defined.

We saw in Chapter 1 that the measure-theoretic typicality paradigm is the approach that most clearly brings out this feature of algorithmic randomness. Consider, for instance, the notion of Martin-Löf randomness defined with respect to the uniform measure λ . A λ -Martin-Löf random data stream does not possess any atypical properties that can be expressed as Π_2^0 classes of effective λ -measure zero. This ensures that the λ -Martin-Löf random data streams lack any local patterns that would make them atypical and conspicuous in the above sense. On the other hand, by their very definition, the λ -Martin-Löf random data streams are also Borel normal, satisfy the Strong Law of Large Numbers, the Law of the Iterated Logarithm, and all other statistical laws whose satisfaction can be captured in terms of membership in Σ_2^0 classes of effective λ -measure one. Hence, from a statistical point of view, the λ -Martin-Löf random data streams are extremely uniform and regular.

It is this type of global, statistical uniformity, as judged from the Bayesian learner's standpoint (i.e., from the perspective of their prior), that is ultimately responsible for the truth-conduciveness of the algorithmically random data streams. As evinced by the classical version of Lévy's Upward Theorem, Bayesian convergence to the truth is in fact itself a statistical law: it holds with probability one, no matter what the underlying probability measure is, and it embodies a specific type of global uniformity. Effective versions of Lévy's Upward Theorem therefore correspond to specific families of effectively specifiable statistical laws. For example, the effectivisation of Lévy's Upward Theorem that yields a characterisation of μ -weak 1-randomness (Theorem 3.2.12) specifies a special class of epistemically significant statistical laws: all and only the laws that can be expressed as the satisfaction of Lévy's Upward Theorem, where the quantity to be estimated is a computable (extended computable, to be precise) random variable.

From this viewpoint, it makes perfect sense that the algorithmically random data streams should be truth-conducive. It is in their very nature to be so. However, our

characterisation results go well beyond this observation. First of all, they demonstrate that the particular collections of effective statistical laws determined by various effectivisations of Lévy's Upward Theorem suffice to yield characterisations of canonical algorithmic randomness notions. Secondly, given some algorithmic randomness notion R , they establish a precise correspondence between, on the one hand, the types of effectively specifiable statistical laws that are standardly used to define R within the measure-theoretic typicality paradigm and, on the other hand, the types of effectively specifiable statistical laws induced by Lévy's Upward Theorem that also yield a characterisation of R . For instance, going back to our previous example, Theorem 3.2.12 shows that μ -weak 1-randomness is not only characterisable in terms of the satisfaction of all statistical laws that can be expressed as Σ_1^0 classes of μ -measure one; it is also characterisable in terms of the satisfaction of all and only the statistical laws that express convergence to the truth in the sense of Lévy's Upward Theorem, whenever the underlying inductive problem is computable.

3.3.2 Probabilities over rich languages

Gaifman and Snir [1982] were the first to suggest a bridge between the theory of algorithmic randomness and Bayesian convergence to the truth.

In keeping with Carnap's inductive logic programme (cf., for instance, [Carnap, 1952]), Gaifman and Snir develop a comprehensive treatment of probability and probabilistic learning in the setting of a rich logical language over which probabilities are defined. The main advantage of this logical framework is that it allows to factor in the analysis of learning considerations pertaining to the definability and logical complexity of both probability measures and inductive problems.

The language in question, \mathcal{L} , is the standard language of first-order Peano arithmetic \mathcal{L}_0 , augmented with a collection \mathcal{L}_E of (finitely many) relation and function symbols. Possible worlds correspond to models for this language \mathcal{L} : namely, models with \mathbb{N} as their domain. The basic idea is that \mathcal{L}_0 represents our mathematical language: all functions and relations in \mathcal{L}_0 have fixed interpretations over the standard model \mathbb{N} of the natural numbers, so that mathematics is the same in all possible worlds. By contrast, \mathcal{L}_E is the empirical language. The relation and function symbols in \mathcal{L}_E do not have a fixed interpretation over the standard model, and the goal of a Bayesian learner is finding out various facts about

the interpretations of these empirical symbols.

In this very general setting, Gaifman and Snir prove both a version of Lévy's Upward Theorem and a version of the Blackwell-Dubins Theorem (and offer a unified philosophical interpretation of them). In this framework, a Bayesian agent begins their epistemic life with a prior \Pr . The agent does not know what the true world (model) w is, but gathers data about it, thereby obtaining a sequence $\varphi_1^{(w)}, \dots, \varphi_n^{(w)}, \dots$ of \mathcal{L} -sentences—where $\varphi_i^{(w)} = \varphi_i$ if $w \models \varphi_i$ and $\varphi_i^{(w)} = \neg\varphi_i$ otherwise. At each stage n of the learning process, the total evidence amounts to $\bigwedge_{i \leq n} \varphi_i^{(w)}$, and the agent updates their subjective probability to $\Pr(\cdot \mid \bigwedge_{i \leq n} \varphi_i^{(w)})$. Given some hypothesis ψ , the agent's beliefs are then said to converge to the truth about ψ in the limit if $\lim_{n \rightarrow \infty} \Pr(\psi \mid \bigwedge_{i \leq n} \varphi_i^{(w)})$ exists and is equal to the actual truth value $\llbracket \psi \rrbracket(w)$ of ψ at w (namely, 1 if $w \models \psi$, and 0 otherwise).

In the same paper, Gaifman and Snir also provide a general analysis of randomness. They define randomness as follows: given a prior \Pr and a set of sentences Φ , a world w is defined as Φ -random if, for all $\varphi \in \Phi$, $\Pr(\varphi) = 1$ implies that $w \models \varphi$. On the other hand, a world is random *simpliciter* if it is random with respect to the entire collection of sentences of \mathcal{L} . The notion of a randomness test can also be given a general definition within this framework: given a set of sentences Ψ , a Ψ -test for randomness with respect to \Pr is a sentence $\psi \in \Psi$ such that $\Pr(\psi) = 0$. A world w passes test ψ if $w \not\models \psi$. The standard definitions of algorithmic randomness and tests can then be recovered by restricting attention to computable metric spaces, such as Cantor space, computable measures (Δ_1^0 -definable measures), and appropriate sets of sentences to define randomness tests.

Gaifman and Snir then connect randomness and Bayesian convergence to the truth by showing that if both the agent's prior and the procedure adopted by the agent to gather data are definable, then the random worlds (the worlds that are random relative to the entire collection of sentences of \mathcal{L}) are exactly the worlds w on which convergence to $\llbracket \psi \rrbracket(w)$ occurs for all sentences ψ .

So, how is Gaifman and Snir's theorem related to the characterisation results in this dissertation? Along one axis, their theorem is more general, since our characterisation results are proven in the context of Cantor space (so, possible worlds correspond to infinite binary sequences) and computable measures. Along a different axis, however, our results

are more general. We in fact show that algorithmic randomness and truth-conduciveness coincide in the general setting of Lévy's Upward Theorem, where the inductive problems to be solved are arbitrary random variables. As evinced by the above discussion, Gaifman and Snir's result, on the other hand, is only concerned with the case where the inductive problem under investigation can be represented as the characteristic function of a definable set.²⁷

Most importantly, Gaifman and Snir's theorem does not offer a characterisation of any algorithmic randomness notion: the randomness concept featured in their result is the one defined with respect to the entire collection of sentences of \mathcal{L} . To the contrary, our results demonstrate that, when the underlying prior is computable, the type of randomness studied within the theory of algorithmic randomness is precisely the one that robustly characterises the truth-conducive data streams. Moreover, our results establish a precise relationship between the complexity of the inductive problems to be solved and the algorithmic randomness notions that characterise the collection of truth-conducive data streams.

3.3.3 In the long run we are all dead

One of the most common objections to the epistemic significance of Lévy's Upward Theorem (and other convergence-to-the-truth results, both Bayesian and non-Bayesian) targets the asymptotic nature of convergence. Lévy's Upward Theorem guarantees that, with probability one, a Bayesian agent's beliefs will align with the truth, but not at any finite stage of the learning process: in the limit. However, as John Maynard Keynes pointedly put it, "in the long run we are all dead." So, what kind of reassurance can asymptotic convergence to the truth provide to agents like us who are inherently finite and would like to have a guarantee that correct beliefs will be attained within the span of a human lifetime?

A familiar response to this objection is that, while legitimate, this kind of worry does not strip convergence-to-the-truth results of their epistemic relevance. Asymptotic results offer a proof of concept: they demonstrate that, at least in the ideal case where all the evidence is eventually observed, inductive success is guaranteed. Were these results not to

²⁷Cf. the discussion in §3.1.1.

hold, we would be left in an epistemically disconcerting situation: the inductive methods that we so confidently rely upon could fail to lead us to correct beliefs even when all the evidence necessary to settle the question under consideration eventually becomes available.

This response is to some extent assuaging, but we might actually be able to squeeze more philosophical juice out of convergence-to-the-truth theorems, in spite of their asymptotic character. In particular, the computability-theoretic perspective adopted here may provide a promising avenue for achieving a more informative type of convergence. Some of the results presented in this chapter offer a modest step in this direction. In spite of failing to ensure convergence to the truth in the short run, the classical version of Lévy's Upward Theorem does ensure that, as the evidence accumulates, a Bayesian agent's beliefs will get closer and closer to the truth (with probability one). Crucially, in the effective setting discussed here, it is possible to provide a more fine-grained analysis of this phenomenon: we can get a more concrete grasp of the rate at which convergence to the truth is expected to occur. More precisely, when the inductive problem faced by the computable Bayesian agent can be modelled as a random variable that is either (i) L^1 -computable, (ii) an integral test for Schnorr randomness, or (iii) an integral test for weak 1-randomness (see Theorem 3.2.4, Theorem 3.2.5 and Theorem 3.2.12, respectively), then convergence to the truth is expected to happen at a computable rate. In other words, in this setting, for any distance from the truth that a Bayesian agent might be interested in reaching, they will be able to computably determine how many observations are required before their beliefs (their estimates) can achieve that level of closeness to the truth in expectation.

To show that this is indeed the case, we will prove it in the context of L^1 -computable functions (the other two cases are analogous):

Proposition 3.3.1. *If $f : \subseteq 2^{\mathbb{N}} \rightarrow \mathbb{R}$ is an L^1 -computable function with witness $\{f_n\}_{n \in \mathbb{N}}$, then there is a computable increasing sequence $\ell_0, \ell_1, \ell_2, \dots$ of natural numbers such that $\|\mathbb{E}_\mu[f \mid \mathcal{F}_{\ell_m}] - f\|_1 \leq 2^{-m}$ for all $m \in \mathbb{N}$.*

Proof. First note that, for all n, ℓ ,

$$\begin{aligned} \int_{2^{\mathbb{N}}} |\mathbb{E}_\mu[f_n \mid \mathcal{F}_\ell] - \mathbb{E}_\mu[f \mid \mathcal{F}_\ell]| d\mu &= \int_{2^{\mathbb{N}}} |\mathbb{E}_\mu[f_n - f \mid \mathcal{F}_\ell]| d\mu \\ &\leq \int_{2^{\mathbb{N}}} \mathbb{E}_\mu[|f_n - f| \mid \mathcal{F}_\ell] d\mu \end{aligned}$$

$$\begin{aligned}
&= \int_{2^{\mathbb{N}}} |f_n - f| d\mu \\
&\leq 2^{-n}.
\end{aligned}$$

Now, for each m , take f_{m+1} and let ℓ_m be the smallest number greater than ℓ_{m-1} such that $\mathbb{E}_\mu[f_{m+1} \mid \mathcal{F}_{\ell_m}] = f_{m+1}$ (for $m = 0$, simply let ℓ_0 be least with $\mathbb{E}_\mu[f_1 \mid \mathcal{F}_{\ell_0}] = f_1$). Such an ℓ_m can be found computably. Then,

$$\begin{aligned}
\|\mathbb{E}_\mu[f \mid \mathcal{F}_{\ell_m}] - f\|_1 &\leq \|\mathbb{E}_\mu[f \mid \mathcal{F}_{\ell_m}] - \mathbb{E}_\mu[f_{m+1} \mid \mathcal{F}_{\ell_m}]\|_1 + \|\mathbb{E}_\mu[f_{m+1} \mid \mathcal{F}_{\ell_m}] - f_{m+1}\|_1 + \|f_{m+1} - f\|_1 \\
&\leq 2^{-(m+1)} + 0 + 2^{-(m+1)} \\
&= 2^{-m},
\end{aligned}$$

which establishes the result. □

3.3.4 Epistemic immodesty

Rather than seeing Lévy's Upward Theorem as a strength of the Bayesian framework, a number of authors have argued that it constitutes the Achilles heel of Bayesianism (see, for instance, [Glymour, 1980], [Earman, 1992], [Kelly, 1996], and [Belot, 2013]). Belot [2013], in particular, contends that the almost sure asymptotic convergence to the truth ensured by Lévy's Upward Theorem implies that Bayesian reasoners are plagued by a pernicious type of epistemic immodesty:

Some have seen in the tendency of Bayesian agents to converge to the truth—and in related results concerning the eventual merger of opinion between Bayesian agents whose initial credences share a certain amount of common ground—the materials for acquitting personalist Bayesianism of the charge of excessive subjectivity. But recent philosophical commentators (some Bayesians among them) have tended to downplay the significance of these results, pointing out that what they guarantee is that Bayesian agents think that there is no chance that their own future opinions will fail to converge to the truth, which is not the same thing as saying that the opinions of each Bayesian agent are in fact certain to converge to the truth. The truth concerning Bayesian convergence-to-the-truth

results is significantly worse than has been generally allowed—they constitute a real liability for Bayesianism by forbidding a reasonable epistemological modesty. [Belot, 2013, p. 502]

Bayesian convergence-to-the-truth theorems tell us that Bayesian agents are forbidden to think that there is any chance that they will be fooled in the long run, even when they know that their credence function is defined on a space that includes many hypotheses that would frustrate their desire to reach the truth. [Belot, 2013, p. 500]

By the very nature of the Bayesian framework, Bayesian agents are bound to invariably expect that their beliefs will converge to the truth. Because of this, Belot argues, they are forced to ignore the fact that, for many hypotheses, failure, rather than success, is the “typical” outcome of the learning process—where, crucially, the notion of typicality employed in his argument is topological, rather than probabilistic.

We have seen that an event is probabilistically, or measure-theoretically, atypical if it has measure zero and typical if it has measure one (relative to a given probability measure). In topology, on the other hand, typicality is defined qualitatively. Recall that a *nowhere dense* set is one for which the smallest closed set that contains it does not itself contain any non-trivial open set. Roughly speaking, a nowhere dense set is such that its elements are not tightly clustered. An event is then topologically atypical if it is *meagre*: that is, if it is expressible as a countable union of nowhere dense subsets of the given topological space. Conversely, an event is topologically typical if it is *co-meagre*: that is, if it is the complement of a meagre set. Belot’s criticism of Bayesian convergence to the truth is motivated by the following observation: there are several learning situations where the set of data streams along which convergence to the truth occurs turns out to be meagre (topologically atypical); yet, Bayesian agents must nonetheless assign probability one to this event, since Lévy’s Upward Theorem establishes that, from the perspective of the agent, convergence to the truth is measure-theoretically typical.

The observation that measure-theoretic typicality and topological typicality often come apart is not new (see, for instance, [Oxtoby, 1980]). However, Belot’s argument highlights that this dichotomy can also occur in an epistemically relevant context, where the event

witnessing the coming apart of these two notions of typicality is the collection of data streams on which the beliefs of a Bayesian agent converge to the truth about the hypothesis under investigation. In this setting, according to Belot, this dichotomy is particularly alarming, for he takes meagreness to embody an objective notion of typicality—in that it does not depend on any particular agent or their beliefs—while the measure-theoretic notion of typicality, at least when the measure corresponds to a subjective prior, merely reflects a particular agent’s opinion. These considerations are what leads him to conclude that Bayesian agents suffer from an irrational over-confidence in their ability to be inductively successful.²⁸

Recently, Belot’s objection has received a lot of attention in the literature. Most responses so far have focused on either one of two strategies: criticising some of the premises in Belot’s argument (see, for example, [Cisewski et al., 2018]) or substantially modifying the Bayesian framework in order to evade his conclusion. For instance, Huttegger [2015b] proposes to use *metric Boolean algebras*²⁹ to avoid drawing distinctions between events that can only be made by infinite observations, Weatherson [2015] advocates passing to imprecise Bayesianism,³⁰ while Elga [2016] and Nielsen and Stewart [2019] suggest dropping countable additivity in favour of finite additivity.

Another way to address Belot’s worry, however, consists in asking whether there are any natural restrictions that may be imposed on subjective priors and the random variables featuring in Lévy’s Upward Theorem which allow to circumvent Belot’s observation while retaining the standard Bayesian (and measure-theoretic) apparatus. The effectivisation of Lévy’s Upward Theorem we presented in §3.2.2, it turns out, allows to do precisely this.

Recall that the following holds:

²⁸A structurally similar argument was antecedently put forward by Kelly (see, for instance, [Kelly, 1996, Chapter 13]). Kelly’s argument relies on cardinality, rather than topological considerations. In particular, he points out that there are learning situations where, even though the collection of data streams along which convergence to the truth occurs has probability one, the collection of data streams on which convergence to the truth instead fails is an uncountable set. Kelly locates the culprit of Bayesian immodesty in the axiom of countable additivity for probability measures.

²⁹A metric Boolean algebra over the Cantor space of infinite binary sequences (endowed with the Borel σ -algebra $\mathcal{B}(2^{\mathbb{N}})$) is the algebra resulting from factoring out measure-zero sets, relative to a Borel probability measure that is strictly positive (i.e., that assigns positive probability to every open set) and that assigns probability zero to each particular infinite sequence.

³⁰According to imprecise Bayesianism, an agent’s degrees of belief should be represented as a set of credence functions, rather than as a single credence function.

Theorem 3.2.12. *Let μ be a computable measure and $\omega \in 2^{\mathbb{N}}$. Then, the following are equivalent:*

- (1) ω is μ -weakly 1-random;
- (2) for all computable functions $f : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ that are finite almost everywhere,

$$\lim_{k \rightarrow \infty} \mathbb{E}_{\mu}[f \mid \mathcal{F}_k](\omega) = f(\omega) < \infty.$$

Theorem 3.2.12 establishes that, for computable Bayesian agents trying to solve computable inductive problems, the truth-conducive data streams are exactly the weakly 1-random ones.

Crucially, weak 1-randomness is the only algorithmic randomness notion that satisfies both the measure-theoretic and the topological notion of typicality: the collection of weakly 1-random sequences has not only measure one (relative to the prior with respect to which randomness is defined), but it is also a co-meagre set,³¹ provided that the underlying measure is strictly positive. This means that, for computable open-minded Bayesian agents trying to solve computable inductive problems, believing that the truth is within reach does not entail the type of epistemic immodesty that Belot is concerned about. In this setting, Bayesian agents are never forced to *ex ante* exclude from consideration a topologically typical failure set. Belot's challenge can thus be met without deviating from the standard framework by way of introducing computability-theoretic restrictions that naturally apply to more realistic, less-than-ideal Bayesian agents.

Though very simple, this observation raises a rather general question: what restrictions on priors and inductive problems are sufficient for the attainment of almost everywhere convergence in both the measure-theoretic and the topological sense? Put differently, when is convergence to the truth a typical phenomenon both probabilistically and topologically?

3.4 Conclusion

We conclude with a brief summary of the results presented here, and by discussing some possible avenues for future work.

³¹See, for instance, [Nies, 2009, Fact 3.5.4, p. 128].

In this chapter, we investigated the effects of observing an algorithmically random data stream on the learning performance of computationally limited Bayesian agents. Our main finding was that the algorithmically random data streams coincide with the ones that ensure that a computable Bayesian agent’s beliefs will converge to the truth (in the sense of Lévy’s Upward Theorem). In particular, we saw that (1) Schnorr randomness characterises the truth-conducive data streams whenever the quantities to be estimated are either lower semi-computable random variables with computable expectation or L^1 -computable random variables (cf. Theorem 3.2.5 and Theorem 3.2.4, respectively), (2) density randomness characterises the truth-conducive data streams whenever the quantities to be estimated are either lower semi-computable random variables with finite expectation or weakly L^1 -computable random variables (cf. Theorem 3.2.11 and Theorem 3.2.13, respectively), and (3) weak 1-randomness characterises the truth-conducive data streams whenever the quantities to be estimated are computable almost-everywhere finite random variables (cf. Theorem 3.2.12).

This robust correspondence between algorithmic randomness and truth-conduciveness suggests a broader research programme involving the systematic study of the connections between algorithmic randomness and convergence results that are relevant to epistemology.

From the standpoint of Bayesian epistemology, an immediate question is whether there are any other philosophically significant convergence results which, when appropriately effectivised, give rise to the same phenomenon evinced in the context of Lévy’s Upward Theorem—that is, from which it emerges that the data streams along which the relevant type of convergence occurs are exactly the algorithmically random ones. A first step in this direction is taken in [Huttegger et al., 2021], where the Blackwell-Dubins Theorem that took center stage in Chapter 2, as well as Lévy’s Downward Theorem³² are studied from this perspective. Further natural candidates for this kind of analysis are other merging-of-opinions theorems—such as the ones proven by Diaconis and Freedman [1986], D’Aristotile et al. [1988], and Kalai and Lehrer [1994]—de Finetti’s Theorem [1929; 1937], as well as

³²Lévy’s Downward Theorem, also due to Lévy, is the following result. Let $(\Omega, \mathcal{E}, \mu)$ be a probability space and f an integrable random variable. Moreover, let $\{\mathcal{E}_n\}_{n \in \mathbb{N}}$ be any *decreasing* sequence of sub- σ -algebras of \mathcal{E} and define \mathcal{E}_∞ as $\bigcap_{n \in \mathbb{N}} \mathcal{E}_n$. Then, $\lim_{n \rightarrow \infty} \mathbb{E}[f \mid \mathcal{E}_n] = \mathbb{E}[f \mid \mathcal{E}_\infty]$ both μ -almost everywhere and in the L^1 -norm. From an epistemology point of view, Lévy’s Downward Theorem can be seen as elucidating the asymptotic behaviour of a Bayesian agent’s beliefs in the setting of gradual information loss

Solomonoff's result concerning the convergence of his universal prior [Solomonoff, 1964] (which we briefly discussed in the Introduction). Some negative results pertaining to Solomonoff induction may be found in [Hutter and Muchnik, 2004] and [Lattimore and Hutter, 2015]. Lattimore and Hutter [2015], in particular, show that there does not exist a universal prior that converges to the true computable measure on all Martin-Löf random sequences.

This question may also be addressed by combining the results in this chapter with the results from Chapter 2. In Chapter 2, we considered several alternatives to absolute continuity defined in terms of algorithmic randomness: namely, several notions of compatibility between measures determined by various forms of agreement about which data streams are algorithmically random. These notions of compatibility, as we have seen, lead to merging of opinions. In this richer setting, one may therefore ask: if two Bayesian agents are in agreement relative to some algorithmic randomness notion R , do the data streams along which merging of opinions occurs coincide with some (possibly different) algorithmic randomness notion R' ?

Convergence-to-the-truth results are not the prerogative of Bayesian epistemology. Another field concerned with the phenomenon of convergence to the truth is formal learning theory (to which we will return in Chapter 4, though from a different angle). Formal learning theory encompasses a broad class of mathematical frameworks for modelling inductive learning that focus on the concept of reliable inquiry. In this setting, a learning problem consists in (i) a collection of possible worlds or hypotheses, (ii) the data streams that are compatible with each possible world, (iii) a learning method, and (iv) a notion of success, which specifies the conditions under which the learning method may be said to converge to the truth along a data stream. Traditionally, a method is taken to solve the learning problem at hand if it is reliable: that is, if it succeeds (in the specified sense) no matter what the true world turns out to be (i.e., if it is logically guaranteed to succeed). This analysis allows to provide a precise taxonomy of inductive problems, whose complexity can be determined by identifying the notions of success (or convergence) with respect to which these problems are solvable. In addition, formal learning theory is well-suited for studying the learning performance of computationally limited learners, which can be modelled as computable learning methods. In fact, ever since the pioneering work of Putnam [1963,

1965] and Gold [1965, 1967], formal learning theory has been developed in a computability-theoretic direction,³³ and the field is also sometimes referred to as computational learning theory.

Just as in computability theory one may ask whether a problem is decidable given certain assumptions, in formal learning theory one can ask if an inductive problem is solvable on the basis of some particular inductive assumptions. One such possible inductive assumption is algorithmic randomness: as noted by Kelly, “[s]ince randomness is a property of data streams, it is yet another empirical assumption that is subject to empirical scrutiny. [...] The effect of randomness assumptions upon logical reliability is an important issue for further study” [Kelly, 1996, p. 63].

Some interesting results pertaining to the connections between algorithmic randomness assumptions and reliable learning can be found in [Vitányi and Chater, 2017], [Bienvenu et al., 2018], and [Barmpalias et al., 2018]. The questions that they consider are the following: is there a single algorithm that succeeds on all data streams that are random with respect to some computable measure (i.e., an algorithm that eventually correctly guesses the true computable measure upon observing an algorithmically random data stream generated by it), no matter what that measure is? If not, are there any natural classes of computable measures for which such an algorithm exists? In all three papers, these questions are approached from the perspective of Martin-Löf randomness. In particular, Bienvenu et al. [2018] prove that such an algorithm exists only under an extremely weak notion of success.³⁴ In light of this result, an immediate question is what happens in the context of stronger randomness assumptions: that is, if one assumes that the observable data streams are algorithmically random in the sense of some algorithmic randomness notion stronger than Martin-Löf randomness. Do stronger assumptions allow to attain a more demanding type of convergence to the truth?

Algorithmic randomness can be viewed as a theory of (effective) typicality. From this perspective, the results in this chapter establish that observing a typical data stream brings about a certain type of inductive success. Typicality, however, comes in many

³³For a survey, see, for instance, [Osherson et al., 1986].

³⁴This notion of success does not require the learning algorithm to guess exactly the true generating distribution: rather, it requires it to guess measures that are “sufficiently similar” to the correct one. No algorithm guaranteed to guess the correct distribution exists, since, as we have seen in Chapter 2, different computable measures can generate the same collection of algorithmically random data streams.

stripes. For instance, as discussed in §§3.3.4, in addition to measure-theoretic typicality there is also the notion of topological typicality. In fact, the theory of algorithmic randomness itself also studies concepts of effective topological typicality—which, much like their measure-theoretic counterparts, give rise to infinite hierarchies: the n -genericity and the weak n -genericity hierarchy (see, for example, [Downey and Hirschfeldt, 2010, §8.20]). This suggests a very broad question: what are the effects of typicality (effective and non), when taken to be a property of data streams, on inductive learning? The frameworks for which convergence-to-the-truth results are important are numerous. As we have seen, they include Bayesian learning (and, more generally, statistics) and formal learning theory. Another framework worth mentioning is the theory of belief revision, which subsumes numerous logical frameworks aimed at modelling in qualitative terms the process of rational belief change triggered by new pieces of information, and which has more recently been studied from the perspective of convergence to the truth.³⁵ Addressing this question in full generality would thus involve a comprehensive taxonomical effort, aimed at identifying the learning scenarios and the inductive problems for which observing a typical data stream is conducive to learning.

³⁵See, for instance, [Kelly, 1998], [Baltag and Smets, 2011], and [Baltag et al., 2019].

Chapter 4

Algorithmic randomness and unlearnability

But success is not supposed to be a matter of mere luck or accident. A reliable method is in some sense guaranteed to converge to the truth, given the scientist's assumptions. Guarantees come in various grades.

Kelly, *The Logic of Reliable Inquiry*

The results from Chapter 2 and Chapter 3 establish that, for learning tasks such as merging of opinions and convergence to the truth, algorithmic randomness is conducive to learning (either because agreeing on which data streams are algorithmically random is a type of doxastic compatibility that leads to merging of opinions, or because the algorithmically random data streams are precisely the ones that ensure that an agent's beliefs will eventually converge to the truth). These results, as we have seen, hinge on the fact that the algorithmically random data streams, by their very nature, have to display important statistical regularities and constitute the most typical outcomes of the underlying probability measure.

Yet, inductive learning is often described as the process of extrapolating patterns into the future from past empirical data and, since algorithmic randomness amounts to local patternlessness and irregularity, there is a sense in which it is natural to regard randomness

as antithetical to inductive learning. Intuitively, observing an algorithmically random data stream should be detrimental for the learning process whenever the the learning task at hand crucially relies on the presence of patterns in the data that can be used to set them apart from other possible sequences of observations.

This chapter is devoted to studying algorithmic randomness from this perspective: that is, rather than exploring the applications of algorithmic randomness in formal models of learning, in what follows we study algorithmic randomness itself from a learning-theoretic perspective that hinges on the local irregularity of algorithmically random data streams. In a nutshell, the goal of this chapter is to offer novel characterisations of standard algorithmic randomness notions in terms of unlearnability.

As explained in Chapter 1, algorithmic randomness notions are customarily defined in terms of either incompressibility, measure-theoretic typicality, or unpredictability. In a more recent paper, however, Osherson and Weinstein [2008] propose an alternative framework for modelling algorithmic randomness that builds upon intuitions from the field of formal (or computational) learning theory—the mathematical approach to learning spearheaded by Putnam [1963, 1965] and Gold [1965, 1967]. The basic idea behind their approach is that a sequence is random if it does not possess any patterns whose presence can be detected by an effective learning function. More specifically, Osherson and Weinstein define two success criteria for learning functions, which specify under what conditions an infinite sequence of observations can be said to possess an effectively detectable pattern. Then, they use these criteria to offer learning-theoretic characterisations of two well-known algorithmic randomness notions: weak 1-randomness and weak 2-randomness, respectively. In a nutshell, each of these two randomness notions is shown to correspond to the collection of data streams on which all computable learning functions fail to meet the relevant success criterion.¹

This learning-theoretic approach affords an intuitive perspective on algorithmic randomness which conforms to the well-entrenched intuition that patternlessness is detrimental

¹Note that this approach is importantly different from the possible bridge between formal learning theory and algorithmic randomness suggested in the conclusion of Chapter 3. The goal of Osherson and Weinstein’s framework is not to clarify the effects of randomness assumptions on logical reliability; rather, as already hinted at in the previous paragraph, their main idea is to use tools from formal learning theory to define randomness itself.

for learning. Moreover, it invites the question of whether restricting attention to learning-theoretic success criteria comes at an expressivity cost. In other words, to what extent is this approach generalisable and capable of yielding natural characterisations of other algorithmic randomness notions? Osherson and Weinstein’s characterisation results arguably target the weakest (weak 1-randomness) and one of the strongest (weak 2-randomness) core algorithmic randomness notions. Thus, the most immediate question is whether all of the most well-behaved and well-studied algorithmic randomness notions are characterisable in this learning-theoretic setting. An especially pressing issue is whether Martin-Löf randomness—arguably, the most prominent algorithmic randomness notion in the literature—is amenable to a learning-theoretic characterisation. Indeed, being able to capture Martin-Löf randomness seems to be a minimum requirement that any framework aimed at modelling algorithmic randomness should meet.

The purpose of this chapter is to further explore this learning-theoretic framework and gauge its expressivity. In particular, we will answer the latter question in the affirmative by offering a novel, learning-theoretic characterisation of Martin-Löf randomness. Our second main result is a learning-theoretic characterisation of Schnorr randomness—as we have seen, another central algorithmic randomness notion. These characterisation theorems constitute a first step towards an in-depth study of algorithmic randomness from a learning-theoretic perspective, both in the setting of Osherson and Weinstein’s framework and, more generally, through the prism of computational learning theory at large.

The remainder of this chapter is organised as follows. In §4.1, we present the learning-theoretic approach to algorithmic randomness that shall take centre stage here, as well as Osherson and Weinstein’s characterisations of weak 1-randomness and weak 2-randomness. Our interpretation diverges from Osherson and Weinstein’s original formulation of the framework in terms of a memorisation game (see [Osherson and Weinstein, 2008, §3]); this translates into a slightly different presentation of the two success criteria proposed by Osherson and Weinstein (which they use in their characterisations of weak 1-randomness and weak 2-randomness). In §4.2, we then prove our two main results: a learning-theoretic characterisation of Martin-Löf randomness (§§4.2.2) and a learning-theoretic characterisation of Schnorr randomness (§§4.2.3). Both proofs rely on the identification of a natural bridge between this learning-theoretic approach and the variant of the standard measure-theoretic

typicality paradigm built upon integral tests (cf. §1.2.2).

All of the results will be presented in terms of the uniform measure λ , but they are readily generalisable to all other computable measures on Cantor space. Since no ambiguity will arise, we will write Martin-Löf randomness, Schnorr randomness, etc., rather than λ -Martin-Löf randomness, λ -Schnorr randomness, and so on.

4.1 Algorithmic randomness and detectability

All three standard algorithmic randomness paradigms rely on a common intuition: a sequence fails to be random if some effective method can detect the presence of a distinguishing pattern in it. Each paradigm hinges on a different formalisation of the notion of an effective pattern-detection method: the incompressibility paradigm formalises it in terms of compression algorithms, the measure-theoretic typicality paradigm in terms of effective statistical tests for randomness, and the unpredictability paradigm in terms of effective betting strategies.

Another simple way to capture the notion of an effective pattern-detection method is in terms of computable functions of the form $\ell : 2^{<\mathbb{N}} \rightarrow \{\text{yes, no}\}$, which take as input finite binary strings and output either **yes** or **no**. A method of this kind may be viewed as a qualitative, coarse-grained, black-box testing procedure, whose goal is to ascertain whether the world, modelled as an infinite binary sequence, is patterned or not. The method is fed longer and longer initial segments of some infinite sequence and, at each stage, it has to output a conjecture as to whether said sequence exhibits any of the patterns that the method is looking for. Whenever the method outputs **yes**, its current conjecture is that the observed sequence indeed displays a pattern (and therefore fails to be random); when it outputs **no**, on the other hand, the method is guessing that the sequence does not exhibit any of its target patterns. Methods of this kind will be referred to as (computable) learning functions.

In Osherson and Weinstein's framework, whether a learning function counts as having detected the presence of a pattern in a sequence depends on how high the standards for detectability are. In particular, pattern detection amounts to the satisfaction of some *success criterion*, which, to a first approximation, specifies how often a learning function

has to answer yes while observing a sequence in order to count as having spotted at least one of the patterns that it is looking for. The patterns that a learning function is geared towards, in turn, are defined extensionally: they are taken to coincide with the set of sequences on which the relevant success criterion is met by the given learning function. This means that, once a success criterion has been fixed, what counts as a pattern is determined by the discriminatory power of computable learning functions. Different success criteria may then be seen as tracking the complexity of the patterns involved: the weaker, or the more permissive, the success criterion, the more subtle and difficult to detect the patterns defined in terms of such criterion. A set that meets the requirements for being a learning-theoretic pattern with respect to a certain success criterion may amount to mere noise from the perspective of a stronger success criterion that imposes higher standards for detectability—and, thus, for patternhood.

A further constraint imposed on learning-theoretic patterns in this setting is that they have to be measure-theoretically rare. In other words, to qualify as a relevant pattern, a set need not only coincide with the collection of sequences on which a computable learning function satisfies a certain success criterion, but it must also have measure zero. One way to think about this requirement is that it restricts attention to patterns that allow a learning function to truly single out sequences displaying them, in the sense that a sequence possessing a pattern of this kind can be told apart from most other sequences (to be precise, measure-one many sequences) by a computable learning function.²

As we shall see, algorithmically random sequences (for some suitable notions of randomness) are precisely the ones that no computable learning function can single out, for they do not possess any patterns whose presence is detectable by a computable learning function. The learning-theoretic patterns in question, determined by more or less demanding success criteria, will vary with the algorithmic randomness notion under consideration. The weaker the success criterion, the stronger the randomness notion such criterion yields.

The remainder of §4.1 will be devoted to discussing the two success criteria introduced by Osherson and Weinstein (renamed in accordance with our interpretation of the framework), as well as their characterisations of weak 1-randomness and weak 2-randomness in terms of said criteria.

²As a result, this constraint also rules out trivial learning functions that ignore the input.

4.1.1 Strong detectability

The first success criterion proposed by Osherson and Weinstein echoes the notion of *identifiability in the limit* [Gold, 1967] from formal learning theory, according to which a learning function is successful if there is a finite number of observations after which the guesses of the learning function are always the same and correct. In light of our interpretation of the framework, we call this success criterion *strong detectability*.

Definition 4.1.1 (Strong detectability). *A learning function $\ell : 2^{<\mathbb{N}} \rightarrow \{\text{yes, no}\}$ is said to strongly detect that a sequence $\omega \in 2^{\mathbb{N}}$ is patterned if and only if*

- (1) $\ell(\omega \upharpoonright m) = \text{yes}$ for cofinitely many $m \in \mathbb{N}$, and
- (2) $\lambda(\{\alpha \in 2^{\mathbb{N}} : \ell(\alpha \upharpoonright m) = \text{yes for cofinitely many } m \in \mathbb{N}\}) = 0$, where λ denotes the uniform measure on the Borel σ -algebra $\mathcal{B}(2^{\mathbb{N}})$.

Strong detectability places constraints on the type of “convergence to the truth” (roughly speaking, the amount of affirmative answers) that a learning function has to achieve in order to count as having detected a pattern, as well as on how “common” the patterns involved have to be. The relevant notion of convergence to the truth—the success criterion that a learning function has to meet to qualify for strong detectability—is given by Condition (1) in Definition 4.1.1: learning requires that a learning function output *yes* on all but finitely many initial segments of the observed data stream. This is a demanding success criterion: it calls for the eventual stabilisation on the conjecture that a pattern is indeed present. More precisely, strong detectability requires that there be a finite number of observations after which the learning function conjectures that the sequence displays the target pattern and then never retracts this conjecture on the basis of further observations.

Condition (2), on the other hand, requires that, to qualify for strong detectability, a learning function must be looking for some property, or collection of properties, that is rare: i.e., that only measure-zero many sequences possess. On the technical side, Condition (2) ensures that the notion of strong detectability does not trivialise: more precisely, it guarantees that the concept of pattern employed in Definition 4.1.1 is not so liberal that all sequences end up possessing a pattern that can be strongly detected. To see this, let \mathcal{P}_ℓ denote the success set of ℓ —that is, let $\mathcal{P}_\ell = \{\alpha \in 2^{\mathbb{N}} : \ell(\alpha \upharpoonright m) = \text{yes for cofinitely many } m \in \mathbb{N}\}$.

\mathbb{N} —and suppose that Condition (2) were replaced by the weaker constraint $\lambda(\mathcal{P}_\ell) \leq r$, for $r > 0$. Then, for every $\omega \in 2^{\mathbb{N}}$, we could take the cylinder $[\omega \upharpoonright n]$, for some n large enough to ensure that $\lambda([\omega \upharpoonright n]) = 2^{-n} \leq r$. The learning function $\ell_{\omega \upharpoonright n} : 2^{<\mathbb{N}} \rightarrow \{\text{yes, no}\}$ given by $\ell_{\omega \upharpoonright n}(\sigma) = \text{yes}$ if and only if $\omega \upharpoonright n \sqsubseteq \sigma$ would then strongly detect that ω is patterned, because $\omega \in [\omega \upharpoonright n] = \mathcal{P}_{\ell_{\omega \upharpoonright n}} = \{\alpha \in 2^{\mathbb{N}} : \ell_{\omega \upharpoonright n}(\alpha \upharpoonright m) = \text{yes} \text{ for cofinitely many } m \in \mathbb{N}\}$. Clearly, this argument no longer works if \mathcal{P}_ℓ is required to be a null set, as all cylinders have positive measure (under the uniform measure).³ More generally, as we shall see shortly, Theorem 4.1.2 below establishes that there are indeed sequences (in fact, measure-one many sequences) that do not display any measure-zero patterns detectable by a computable learning function in the sense of Definition 4.1.1.

The best way to ensure that strong detectability by a learning function captures a reasonable notion of “absence of randomness” consists in characterising a standard algorithmic randomness notion in terms of this success criterion. This is precisely what Osherson and Weinstein do in their paper: they prove that weak 1-randomness can be characterised via strong detectability by restricting attention to computable learning functions.⁴

Weak 1-randomness is often described as the weakest algorithmic randomness notion (we have seen that, in general, it does not even imply the Strong Law of Large Numbers). The success criterion with which it can be characterised in this learning-theoretic setting, strong detectability, is correspondingly very stringent.

Theorem 4.1.2 (Osherson and Weinstein [2008]). *A sequence $\omega \in 2^{\mathbb{N}}$ is weakly 1-random if and only if there is no computable learning function that strongly detects that ω is patterned.*

The left-to-right direction of the proof of Theorem 4.1.2 is immediate: restricting attention to the class of computable learning functions entails that the patterns targeted by strong detectability are countable unions of measure-zero Π_1^0 classes. Given a computable learning function ℓ , we in fact have that

$$\begin{aligned} & \{\alpha \in 2^{\mathbb{N}} : \ell(\alpha \upharpoonright m) = \text{yes} \text{ for cofinitely many } m \in \mathbb{N}\} \\ &= \bigcup_{n \in \mathbb{N}} \{\alpha \in 2^{\mathbb{N}} : (\forall m > n) \ell(\alpha \upharpoonright m) = \text{yes}\}. \end{aligned}$$

³Clearly, this argument generalises to any (computable) regular measure.

⁴Recall that a sequence $\omega \in 2^{\mathbb{N}}$ is *weakly 1-random* if and only if it belongs to every Σ_1^0 class of measure one (cf. Definition 1.2.9).

Hence, if a computable learning function strongly detects that a sequence ω is patterned, then there is a measure-one Σ_1^0 class to which ω fails to belong.

Both directions offer a specifically learning-theoretic perspective on the kinds of patterns ruled out by weak 1-randomness. The left-to-right direction says that if a sequence is weakly 1-random, then no finite amount of observations will ever suffice to convince a computable learning function that its target pattern is present. The right-to-left direction, on the other hand, reveals that any sequence that fails to be weakly 1-random displays some rare pattern whose presence can be detected by a computable learning function—in the sense that there is some computable learning function and some initial segment of the sequence that will make this learning function conjecture that its target pattern is present, and after which said learning function will never be prompted to change its mind about this conjecture by further observations. A failure of weak 1-randomness can thus always be witnessed via this relatively simple behaviour of computable learning functions.

4.1.2 Weak detectability

The second success criterion introduced by Osherson and Weinstein is what we call *weak detectability*:

Definition 4.1.3 (Weak detectability). *A learning function $\ell : 2^{<\mathbb{N}} \rightarrow \{\text{yes}, \text{no}\}$ is said to weakly detect that a sequence $\omega \in 2^{\mathbb{N}}$ is patterned if and only if*

- (1) $\ell(\omega \upharpoonright m) = \text{yes}$ for infinitely many $m \in \mathbb{N}$, and
- (2) $\lambda(\{\alpha \in 2^{\mathbb{N}} : \ell(\alpha \upharpoonright m) = \text{yes for infinitely many } m \in \mathbb{N}\}) = 0$.

Weak detectability differs from strong detectability only with respect to Condition (1): both notions are tailored to the detection of measure-zero patterns, but they rely on different success criteria. In particular, the success criterion used in the context of weak detectability, which echoes the notion of *partial learning* [Osherson et al., 1986] from formal learning theory, is a rather permissive one. To see this, suppose that a given learning function ℓ weakly detects that some sequence $\omega \in 2^{\mathbb{N}}$ is patterned. Upon sequentially observing longer and longer initial segments of ω , ℓ may change its mind as to whether ω displays its target pattern infinitely often—i.e., $\ell(\omega \upharpoonright m) = \text{no}$ may occur for infinitely

many m —and yet count as having (weakly) detected that ω is patterned, provided that ℓ outputs **yes** infinitely often on ω . Thus, ℓ weakly detecting that ω is patterned only ensures that there is no finite stage (that is, no initial segment of ω) after which ℓ deems the hypothesis that ω is patterned falsified. Put differently, there is no finite stage beyond which ℓ stops considering it possible that ω displays its target pattern.

In their paper, Osherson and Weinstein show that, when one restricts attention to computable learning functions, weak detectability can be employed to characterise another standard algorithmic randomness notion: weak 2-randomness.⁵

Weak 2-randomness is much stronger than weak 1-randomness. And indeed its learning-theoretic characterisation operates via the more lenient success criterion of weak detectability. This highlights the fact that the patterns which reveal a failure of weak 2-randomness can be more complex than the ones which reveal a failure of weak 1-randomness.

Theorem 4.1.4 (Osherson and Weinstein [2008]). *A sequence $\omega \in 2^{\mathbb{N}}$ is weakly 2-random if and only if there is no computable learning function that weakly detects that ω is patterned.*

Just as in the case of Osherson and Weinstein’s learning-theoretic characterisation of weak 1-randomness, it is immediate that weak 2-randomness implies that weak detectability by computable learning functions is impossible. This is due to the fact that the patterns targeted by weak detectability in the context of computable learning functions are measure-zero Π_2^0 classes. Given a computable learning function ℓ and a sequence ω , ω is patterned according to ℓ if and only if

$$\omega \in \{\alpha \in 2^{\mathbb{N}} : (\forall n)(\exists m > n) \ell(\alpha \upharpoonright m) = \text{yes}\},$$

and this set has measure zero. Thus, if the computable learning function ℓ weakly detects that ω is patterned, then there is a Σ_2^0 class of measure one to which ω does not belong.

From a learning-theoretic point of view, both directions of Theorem 4.1.4 are revealing. The right-to-left direction shows that a failure of weak 2-randomness guarantees the existence of a computable learning function forever entertaining the possibility that its target pattern is present. The left-to-right direction, on the other hand, tells us that if a sequence

⁵Recall that a sequence $\omega \in 2^{\mathbb{N}}$ is *weakly 2-random* if and only if it belongs to every Σ_2^0 class of measure one.

is weakly 2-random, then any computable learning function looking for a measure-zero pattern will answer no cofinitely often: namely, it will eventually stabilise on the conjecture that its target pattern is absent.

4.2 Martin-Löf randomness and Schnorr randomness

An immediate question prompted by Osherson and Weinstein’s results is whether this learning-theoretic framework is expressive enough to afford characterisations of at least some of the algorithmic randomness notions that lie between weak 2-randomness and weak 1-randomness. Being able to capture Martin-Löf randomness, in particular, seems to be a benchmark requirement for any framework aimed at modelling algorithmic randomness (on account of the central role of Martin-Löf randomness in the field). In what follows, we will show that, by taking the notion of weak detectability from §4.1.2 and strengthening the measure-theoretic requirements imposed on the target patterns in a natural way, it is possible to give a learning-theoretic characterisation of not only Martin-Löf randomness, but also of Schnorr randomness.

Our results rely on bridging the learning-theoretic approach and the measure-theoretic typicality paradigm in the context of integral tests for randomness (§3.2.2). To elucidate the connections between these two frameworks, we will begin by offering a direct proof of the equivalence between the characterisation of weak 2-randomness via integral tests (Theorem 1.2.16) and its characterisation in terms of weak detectability (Theorem 4.1.4). Then, we will see how similar ideas can be employed to characterise Martin-Löf randomness and Schnorr randomness learning-theoretically.

4.2.1 A connection with integral tests

Recall that integral tests for randomness are classes of effectively approximable functions of the form $f : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ which meet certain measure-theoretic conditions—where the relevant notion of effective approximability is captured in terms of lower semi-computable functions.

The reason why integral tests for randomness offer an intuitive bridge between the measure-theoretic and the learning-theoretic approach to randomness is that, as we will see, integral tests may be naturally viewed as tracking the number of affirmative answers

of a learning function, while learning functions can be seen as tracking the growth of the computable approximation witnessing the lower semi-computability of an integral test.

To build some intuitions, here is a direct proof of the equivalence between the learning-theoretic characterisation of weak 2-randomness from Theorem 4.1.4 and its characterisation in terms of integral tests.

Proposition 4.2.1. *Let $\omega \in 2^{\mathbb{N}}$. Then, the following are equivalent:*

- (a) *there is a computable learning function that weakly detects that ω is patterned;*
- (b) *there is a lower semi-computable function $f : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ with $\lambda(\{\alpha \in 2^{\mathbb{N}} : f(\alpha) < \infty\}) = 1$ (i.e., an integral test for weak 2-randomness) such that $f(\omega) = \infty$.*

Proof. ((a) \Rightarrow (b)) Let ℓ denote the computable learning function that weakly detects that ω is patterned. Then, $\omega \in \{\alpha \in 2^{\mathbb{N}} : \ell(\alpha \upharpoonright n) = \text{yes for infinitely many } n\}$ and $\lambda(\{\alpha \in 2^{\mathbb{N}} : \ell(\alpha \upharpoonright n) = \text{yes for infinitely many } n\}) = 0$. Define the function $f : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ as $f(\alpha) = \#\{n \in \mathbb{N} : \ell(\alpha \upharpoonright n) = \text{yes}\}$ for every $\alpha \in 2^{\mathbb{N}}$. It then immediately follows that $f(\omega) = \infty$ and $\lambda(\{\alpha \in 2^{\mathbb{N}} : f(\alpha) < \infty\}) = 1$. So, all that is left to show is that f is lower semi-computable. Let $h : 2^{<\mathbb{N}} \rightarrow \mathbb{N}$ be given by $h(\sigma) = \#\{n \leq |\sigma| : \ell(\sigma \upharpoonright n) = \text{yes}\}$ for every $\sigma \in 2^{<\mathbb{N}}$. Then, for each $k \in \mathbb{N}$ and $\sigma \in 2^{<\mathbb{N}}$, let $g_k(\sigma) = h(\sigma)$. Since ℓ is by assumption computable, so is h . Hence, the g_k 's trivially form a sequence of uniformly computable functions. Moreover, for each k, σ and τ , $g_{k+1}(\sigma) = g_k(\sigma)$ and $g_k(\sigma\tau) \geq g_k(\sigma)$. Finally, for each $\alpha \in 2^{\mathbb{N}}$, $f(\alpha) = \sup\{h(\alpha \upharpoonright n) : n \geq 0\} = \sup\{g_k(\alpha \upharpoonright n) : k, n \geq 0\}$. Hence, by Definition 1.2.12, f is indeed lower semi-computable.

((b) \Rightarrow (a)) Since f is lower semi-computable, there is a sequence of uniformly computable functions $g_k : 2^{<\mathbb{N}} \rightarrow \mathbb{Q}$ satisfying conditions (1)-(3) from Definition 1.2.12. Without loss of generality, we can assume that f is non-negative and that $g_0(\varepsilon) = 0$. First, we prove the following simple auxiliary lemma.

Lemma 4.2.2. *Let $\alpha \in 2^{\mathbb{N}}$ and $g : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ a non-negative lower semi-computable function with approximation $\{g_k\}_{k \in \mathbb{N}}$. Then, $g(\alpha) = \infty$ if and only if, for every $i \in \mathbb{N}^+$, there is some $m_0 \in \mathbb{N}^+$ such that $g_{m_0}(\alpha \upharpoonright m_0) \geq i > g_{m_0-1}(\alpha \upharpoonright m_0 - 1)$.*

Proof of Lemma 4.2.2. (\Rightarrow) Suppose that $g(\alpha) = \infty$ and $i > 0$. Since $g(\alpha) = \sup\{g_k(\alpha \upharpoonright n) : k, n \geq 0\}$, there must be a pair k, n of natural numbers such that $g_k(\alpha \upharpoonright n) \geq i$.

Let $m = \max\{k, n\}$. Then, $g_m(\alpha \upharpoonright m) \geq g_k(\alpha \upharpoonright n) \geq i$. Now, for all $j \leq m$, $g_m(\alpha \upharpoonright m) \geq g_j(\alpha \upharpoonright j)$. So, let m_0 be the least natural number such that $g_{m_0}(\alpha \upharpoonright m_0) \geq i$. Again, without loss of generality, we can assume that $g_0(\varepsilon) = 0$. Then, since $i > 0$, we have that $m_0 > 0$. Hence, $g_{m_0}(\alpha \upharpoonright m_0) \geq i > g_{m_0-1}(\alpha \upharpoonright m_0 - 1)$.

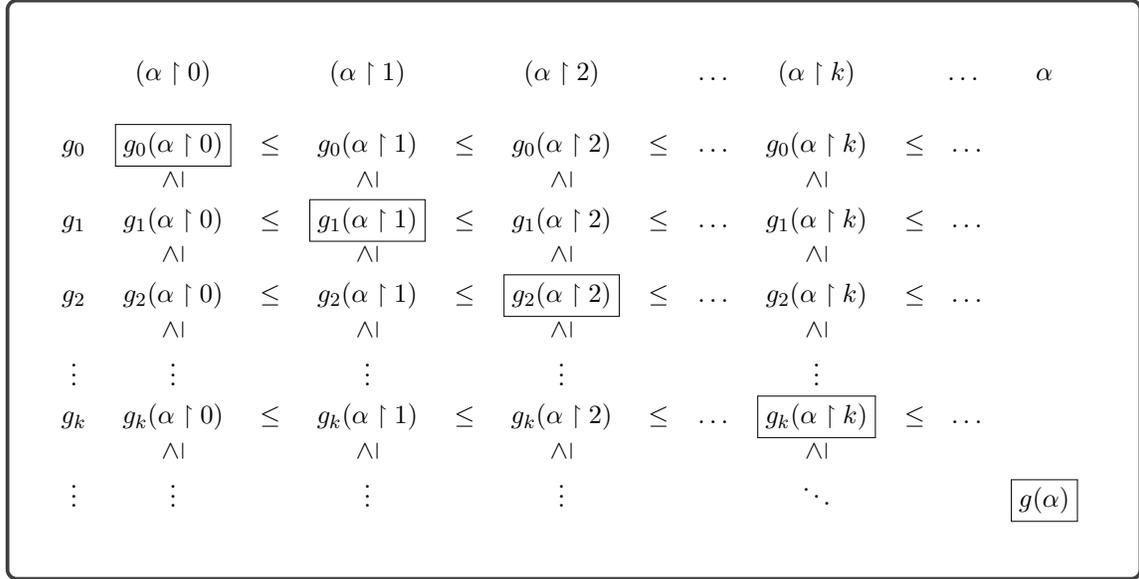


Figure 4.1: The sequence $\{g_k\}_{k \in \mathbb{N}}$ of uniformly computable functions that approximate g from the proof of Lemma 4.2.2. For each $k \in \mathbb{N}$ and $\alpha \in 2^{\mathbb{N}}$, $g_k(\alpha \upharpoonright k)$ is the best approximation of $g(\alpha)$ from among the values appearing in the quadrant whose bottom right corner is occupied by $g_k(\alpha \upharpoonright k)$.

(\Leftarrow) Immediate. +

Now, for each string $\sigma \neq \varepsilon$, let σ^- denote the initial segment of σ of length $|\sigma| - 1$. Define a learning function ℓ as follows: $\ell(\varepsilon) = \text{yes}$ and, for all $\sigma \neq \varepsilon$,

$$\ell(\sigma) = \begin{cases} \text{yes} & \text{if there is some } i \in \mathbb{N}^+ \text{ such that } g_{|\sigma|}(\sigma) \geq i > g_{|\sigma|-1}(\sigma^-); \\ \text{no} & \text{otherwise.} \end{cases}$$

Since the g_k 's are uniformly computable functions, ℓ is computable, as well. Additionally, we claim that

$$\{\alpha \in 2^{\mathbb{N}} : f(\alpha) = \infty\} = \{\alpha \in 2^{\mathbb{N}} : \ell(\alpha \upharpoonright n) = \text{yes for infinitely many } n \in \mathbb{N}\}.$$

For the left-to-right inclusion, suppose that $f(\alpha) = \infty$. Then, by Lemma 4.2.2, for every $i \in \mathbb{N}^+$, there is some $m_0 \in \mathbb{N}^+$ with $g_{m_0}(\alpha \upharpoonright m_0) \geq i > g_{m_0-1}(\alpha \upharpoonright m_0 - 1)$. Hence, $\ell(\alpha \upharpoonright n) = \text{yes}$ for infinitely many $n \in \mathbb{N}$. For the right-to-left inclusion, suppose that $f(\alpha) < \infty$. Let $i \in \mathbb{N}^+$ be least with $g_n(\alpha \upharpoonright n) < i$ for all $n \in \mathbb{N}$. Then, $\#\{n \in \mathbb{N} : \ell(\alpha \upharpoonright n) = \text{yes}\} \leq i$.

Given that $f(\omega) = \infty$, we have that $\omega \in \{\alpha \in 2^{\mathbb{N}} : \ell(\alpha \upharpoonright n) = \text{yes} \text{ for infinitely many } n \in \mathbb{N}\}$. Moreover, since $\lambda(\{\alpha \in 2^{\mathbb{N}} : f(\alpha) < \infty\}) = 1$, it follows that $\lambda(\{\alpha \in 2^{\mathbb{N}} : \ell(\alpha \upharpoonright n) = \text{yes} \text{ for infinitely many } n \in \mathbb{N}\}) = 0$. Hence, ℓ is a computable learning function that weakly detects that ω is patterned. \square

Next, we will see how the basic ideas underlying the above proof can be adapted to obtain characterisations of Martin-Löf randomness and Schnorr randomness.

4.2.2 Uniform weak detectability and Martin-Löf randomness

First, recall that, for any computable measure μ , Martin-Löf randomness can be characterised via integral tests for randomness as follows:

Theorem 1.2.14 (Levin [1976]). *Let $\omega \in 2^{\mathbb{N}}$. The following are equivalent:*

- (1) ω is μ -Martin-Löf random;
- (2) $f(\omega) < \infty$ for all lower semi-computable functions $f : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ with finite expectation: i.e., such that $\int_{2^{\mathbb{N}}} f d\mu < \infty$;
- (3) $f(\omega) < \infty$ for all lower semi-computable functions $f : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ such that $\int_{2^{\mathbb{N}}} f d\mu \leq 1$.

Now, the key to providing a learning-theoretic characterisation of Martin-Löf randomness lies in appropriately strengthening the notion of weak detectability:

Definition 4.2.3 (Uniform weak detectability). *A learning function $\ell : 2^{<\mathbb{N}} \rightarrow \{\text{yes}, \text{no}\}$ is said to uniformly weakly detect that a sequence $\omega \in 2^{\mathbb{N}}$ is patterned if and only if*

- (1) $\ell(\omega \upharpoonright m) = \text{yes}$ for infinitely many $m \in \mathbb{N}$, and
- (2) $\lambda(\{\alpha \in 2^{\mathbb{N}} : \#\{m \in \mathbb{N} : \ell(\alpha \upharpoonright m) = \text{yes}\} \geq n\}) \leq 2^{-n}$ for all $n \in \mathbb{N}$.

What kind of patterns does the notion of uniform weak detectability pick out? Let \mathcal{P}_ℓ denote the success set $\{\alpha \in 2^{\mathbb{N}} : \ell(\alpha \upharpoonright m) = \text{yes} \text{ for infinitely many } m \in \mathbb{N}\}$ of ℓ . Since

$$\mathcal{P}_\ell = \bigcap_{n \in \mathbb{N}} \{\alpha \in 2^{\mathbb{N}} : \#\{m \in \mathbb{N} : \ell(\alpha \upharpoonright m) = \text{yes}\} \geq n\},$$

Condition (2) above implies that \mathcal{P}_ℓ is a measure-zero pattern. However, it also tells us something more. A sequence $\alpha \notin \mathcal{P}_\ell$ such that $\#\{m \in \mathbb{N} : \ell(\alpha \upharpoonright m) = \text{yes}\} \geq n$ is one that “fools” the learning function ℓ into thinking that $\alpha \in \mathcal{P}_\ell$ up to some initial segment of α . Condition (2) then says that, as n increases, fooling ℓ becomes increasingly “difficult”—fewer and fewer sequences can fool ℓ at least n times—and that this increase in difficulty can be effectively bounded. More precisely, the difficulty increases at a computable rate uniformly in n . Thus, \mathcal{P}_ℓ is a measure-zero pattern on which ℓ zeroes in at a computable rate.⁶

Next, we put the notion of uniform weak detectability to work and show that it indeed yields Martin-Löf randomness. First, note that, given a computable learning function ℓ that uniformly weakly detects that a sequence $\omega \in 2^{\mathbb{N}}$ is patterned, the sequence $\{\mathcal{U}_n\}_{n \in \mathbb{N}}$ —where, for all $n \in \mathbb{N}$, $\mathcal{U}_n = \{\alpha \in 2^{\mathbb{N}} : \#\{m \in \mathbb{N} : \ell(\alpha \upharpoonright m) = \text{yes}\} \geq n\}$ —is a sequential Martin-Löf test (see Definition 1.2.3). It is therefore immediate that Martin-Löf randomness implies that uniform weak detectability by computable learning functions is impossible: if a computable learning function uniformly weakly detects that some sequence is patterned, then there is a (learning-theoretic) sequential Martin-Löf test that said sequence fails. Once again, the more interesting implication is the converse one, which establishes that a failure of Martin-Löf randomness can be converted into an instance of uniform weak detectability.

To highlight the connections between the learning-theoretic framework and integral tests for algorithmic randomness, both directions of our learning-theoretic characterisation of Martin-Löf randomness will proceed by appealing to its characterisation in terms of

⁶Note that fooling a learning function becomes more and more difficult in the setting of weak detectability, as well. In this case, however, the increase in difficulty cannot be effectively bounded. The patterns targeted by weak detectability are measure-zero Π_2^0 classes: that is, countable intersections of uniformly Σ_1^0 classes whose measures converge to zero, albeit not effectively. This means that, given a learning function ℓ that weakly detects that a certain sequence is patterned, for all n , there is some N such that, for all $m \geq N$, the collection of sequences that fool ℓ at least m times has measure at most 2^{-n} . However, in general, this N cannot be found computably in n .

integral tests, as given in Theorem 1.2.14.

Theorem 4.2.4. *A sequence $\omega \in 2^{\mathbb{N}}$ is Martin-Löf random if and only if there is no computable learning function that uniformly weakly detects that ω is patterned.*

Proof. (\Rightarrow) Suppose that there is a computable learning function ℓ which uniformly weakly detects that ω is patterned. Just as in the ((a) \Rightarrow (b))-direction of the proof of Proposition 4.2.1, define $f : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ as $f(\alpha) = \#\{m \in \mathbb{N} : \ell(\alpha \upharpoonright m) = \text{yes}\}$ for all $\alpha \in 2^{\mathbb{N}}$. Then, f is lower semi-computable and $f(\omega) = \infty$. Moreover, the fact that

$$\lambda(\{\alpha \in 2^{\mathbb{N}} : \#\{m \in \mathbb{N} : \ell(\alpha \upharpoonright m) = \text{yes}\} \geq n\}) \leq 2^{-n}$$

for each $n \in \mathbb{N}$ implies that $\lambda(\{\alpha \in 2^{\mathbb{N}} : f(\alpha) \geq n\}) \leq 2^{-n}$ for all n . By definition, f can take only countably many values. In particular, since the set of sequences $\alpha \in 2^{\mathbb{N}}$ with $f(\alpha) = \infty$ is a null set, it follows that

$$\begin{aligned} \int_{2^{\mathbb{N}}} f d\lambda &= \sum_{n \in \mathbb{N}} n \cdot \lambda(\{\alpha \in 2^{\mathbb{N}} : f(\alpha) = n\}) \\ &\leq \sum_{n \in \mathbb{N}} n \cdot \lambda(\{\alpha \in 2^{\mathbb{N}} : f(\alpha) \geq n\}) \\ &\leq \sum_{n \in \mathbb{N}} n \cdot 2^{-n}. \end{aligned}$$

Hence, $\int_{2^{\mathbb{N}}} f d\lambda \leq 2$. By Theorem 1.2.14—and, in particular, the equivalence of (1) and (2)— ω is not Martin-Löf random.

(\Leftarrow) Suppose that ω is not Martin-Löf random. Then, by the equivalence of (1) and (3) from Theorem 1.2.14, there is a lower semi-computable function $f : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ with $\int_{2^{\mathbb{N}}} f d\lambda \leq 1$ and such that $f(\omega) = \infty$. Since f is lower semi-computable, there is a sequence of uniformly computable functions $g_k : 2^{<\mathbb{N}} \rightarrow \mathbb{Q}$ satisfying conditions (1)-(3) from Definition 1.2.12. Without loss of generality, we can assume that f is non-negative and that $g_0(\varepsilon) = 0$. We define a learning function $\ell : 2^{<\mathbb{N}} \rightarrow \{\text{yes}, \text{no}\}$ analogously to the way we defined the learning function in the ((b) \Rightarrow (a))-direction of the proof of Proposition 4.2.1, except that now ℓ will be required to answer *yes* more sparingly. Let $\ell(\varepsilon) = \text{no}$. For all $\sigma \neq \varepsilon$, recall that σ^- denotes the initial segment of σ of length $|\sigma| - 1$, and set

$$\ell(\sigma) = \begin{cases} \text{yes} & \text{if there is some } i \in \mathbb{N}^+ \text{ such that } g_{|\sigma|}(\sigma) \geq 2^i > g_{|\sigma|-1}(\sigma^-); \\ \text{no} & \text{otherwise.} \end{cases}$$

Since the g_k 's are uniformly computable, the learning function ℓ is computable, as well.

Moreover, the fact that $f(\omega) = \infty$ implies that $\omega \in \{\alpha \in 2^{\mathbb{N}} : \ell(\alpha \upharpoonright m) = \text{yes} \text{ for infinitely many } m \in \mathbb{N}\}$, as the proof of Lemma 4.2.2 goes through with i replaced by 2^i . Now, for each n ,

$$\begin{aligned} & \{\alpha \in 2^{\mathbb{N}} : \#\{m \in \mathbb{N} : \ell(\alpha \upharpoonright m) = \text{yes}\} \geq n\} \\ &= \{\alpha \in 2^{\mathbb{N}} : \#\{m \in \mathbb{N} : \exists i \in \mathbb{N}^+ \text{ with } g_m(\alpha \upharpoonright m) \geq 2^i > g_{m-1}(\alpha \upharpoonright m-1)\} \geq n\} \\ &\subseteq \{\alpha \in 2^{\mathbb{N}} : \exists m \in \mathbb{N} \text{ with } g_m(\alpha \upharpoonright m) \geq 2^n\} \\ &\subseteq \{\alpha \in 2^{\mathbb{N}} : f(\alpha) \geq 2^n\}. \end{aligned}$$

Since $\int_{2^{\mathbb{N}}} f d\lambda \leq 1$, Markov's inequality entails that, for every n ,

$$\lambda(\{\alpha \in 2^{\mathbb{N}} : f(\alpha) \geq 2^n\}) \leq 2^{-n} \int_{2^{\mathbb{N}}} f d\lambda \leq 2^{-n},$$

which, in turn, implies that

$$\lambda(\{\alpha \in 2^{\mathbb{N}} : \#\{m \in \mathbb{N} : \ell(\alpha \upharpoonright m) = \text{yes}\} \geq n\}) \leq 2^{-n} \text{ for all } n.$$

Hence, ℓ is a computable learning function that uniformly weakly detects that ω is patterned, which concludes the proof. \square

The above result establishes that Martin-Löf randomness indeed admits a characterisation in learning-theoretic terms. In particular, Theorem 4.2.4 shows that Martin-Löf randomness coincides with the collection of sequences along which no computable learning function is able to forever entertain the possibility that its target pattern is present—where the patterns involved are rare, in the sense that they correspond to certain effective measure-zero sets that depend on the given learning function.

4.2.3 Computably uniform weak detectability and Schnorr randomness

Our second result is a learning-theoretic characterisation of Schnorr randomness (which, as we have seen, is strictly weaker than Martin-Löf randomness but strictly stronger than weak 1-randomness).

Recall that Schnorr randomness has the following natural characterisation in terms of integral tests (where μ denotes an arbitrary computable measure):

Theorem 1.2.15 (Miyabe [2013a]). *Let $\omega \in 2^{\mathbb{N}}$. The following are equivalent:*

- (1) ω is μ -Schnorr random;
- (2) $f(\omega) < \infty$ for all lower semi-computable functions $f : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ with computable expectation: i.e., such that $\int_{2^{\mathbb{N}}} f d\mu$ is a computable real;
- (3) $f(\omega) < \infty$ for all lower semi-computable functions $f : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ such that $\int_{2^{\mathbb{N}}} f d\mu = 1$.

In proving Theorem 1.2.15, Miyabe relies on the result below, which will turn out to be useful for the proof of our learning-theoretic characterisation of Schnorr randomness.⁷

Lemma 4.2.5 (Miyabe [2013a]). *Let $f : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ be a lower semi-computable function with $\int_{2^{\mathbb{N}}} f d\lambda = 1$ (i.e., an integral test for Schnorr randomness). Then, there is an unbounded sequence $\{r_n\}_{n \in \mathbb{N}}$ of uniformly computable reals such that, for all n , $\lambda(\{\alpha \in 2^{\mathbb{N}} : f(\alpha) = r_n\}) = 0$ and the measures $\lambda(\{\alpha \in 2^{\mathbb{N}} : f(\alpha) \geq r_n\})$ are computable reals, uniformly in n .*

Schnorr randomness, as we have seen, is a variant on Martin-Löf randomness that relies on imposing stronger effectivity constraints on tests. The key to obtaining a learning-theoretic characterisation of Schnorr randomness thus consists in employing a success criterion that strengthens uniform weak detectability along these lines:

Definition 4.2.6 (Computably uniform weak detectability).

A learning function $\ell : 2^{<\mathbb{N}} \rightarrow \{\text{yes, no}\}$ is said to computably uniformly weakly detect that a sequence $\omega \in 2^{\mathbb{N}}$ is patterned if and only if

- (1) $\ell(\omega \upharpoonright m) = \text{yes}$ for infinitely many $m \in \mathbb{N}$,

⁷We state the result in terms of the uniform measure λ , but Lemma 4.2.5 holds for all computable measures.

- (2) $\lambda(\{\alpha \in 2^{\mathbb{N}} : \#\{m \in \mathbb{N} : \ell(\alpha \upharpoonright m) = \text{yes}\} \geq n\}) \leq 2^{-n}$ for all $n \in \mathbb{N}$, and
- (3) the values $\lambda(\{\alpha \in 2^{\mathbb{N}} : \#\{m \in \mathbb{N} : \ell(\alpha \upharpoonright m) = \text{yes}\} \geq n\})$ are uniformly computable reals.

The only difference between computably uniform weak detectability and uniform weak detectability is Condition (3), which imposes further computability-theoretic constraints. Let \mathcal{P}_ℓ denote the success set $\{\alpha \in 2^{\mathbb{N}} : \ell(\alpha \upharpoonright m) = \text{yes} \text{ for infinitely many } m \in \mathbb{N}\}$ of ℓ . One way to think about Condition (3) is that it ensures that, for each $n \in \mathbb{N}$, it is possible to precisely determine the size (i.e., the measure) of the collection of sequences that are not in \mathcal{P}_ℓ and yet manage to fool ℓ at least n times.

Considerations analogous to the ones applicable to uniform weak detectability are pertinent in this context, as well. In particular, given a computable learning function ℓ that computably uniformly weakly detects that a sequence ω is patterned, if \mathcal{U}_n is defined as the set $\{\alpha \in 2^{\mathbb{N}} : \#\{m \in \mathbb{N} : \ell(\alpha \upharpoonright m) = \text{yes}\} \geq n\}$ for all $n \in \mathbb{N}$, then the sequence $\{\mathcal{U}_n\}_{n \in \mathbb{N}}$ is a sequential Schnorr test (Definition 1.2.6). Once again, to illustrate the connections between the learning-theoretic framework and the measure-theoretic typicality paradigm in the setting of integral tests, both directions of the proof of Theorem 4.2.7 below rely on the characterisation of Schnorr randomness via integral tests, rather than sequential tests.

The right-to-left direction of the proof of Theorem 4.2.7 is more involved than its counterpart in the learning-theoretic characterisation of Martin-Löf randomness. Given a sequence ω that fails to be Schnorr random, the main challenge consists in defining a computable learning function that outputs *yes* sufficiently often to count as having detected that ω is patterned, but that, at the same time, outputs *yes* sufficiently sparingly to guarantee that Condition (2) and Condition (3) from Definition 4.2.6 are satisfied.

Theorem 4.2.7. *A sequence $\omega \in 2^{\mathbb{N}}$ is Schnorr random if and only if there is no computable learning function that computably uniformly weakly detects that ω is patterned.*

Proof. (\Rightarrow) Suppose that there is a computable learning function ℓ which computably uniformly weakly detects that ω is patterned. Once again, let $f : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ be the function given by $f(\alpha) = \#\{m \in \mathbb{N} : \ell(\alpha \upharpoonright m) = \text{yes}\}$ for all $\alpha \in 2^{\mathbb{N}}$. Since $\omega \in \{\alpha \in 2^{\mathbb{N}} : \ell(\alpha \upharpoonright m) = \text{yes} \text{ for infinitely many } m \in \mathbb{N}\}$, $f(\omega) = \infty$. Moreover, by the same argument

used in the proof of Theorem 4.2.4, f is lower semi-computable and $\int_{2^{\mathbb{N}}} f d\lambda = \sum_{n \in \mathbb{N}} n \cdot \lambda(\{\alpha \in 2^{\mathbb{N}} : f(\alpha) = n\}) \leq 2$.

Next, we show that $\int_{2^{\mathbb{N}}} f d\lambda$ is computable. For each $n \in \mathbb{N}$, let λ_n be an abbreviation for $\lambda(\{\alpha \in 2^{\mathbb{N}} : f(\alpha) = n\})$ —or, equivalently, for $\lambda(\{\alpha \in 2^{\mathbb{N}} : \#\{m \in \mathbb{N} : \ell(\alpha \upharpoonright m) = \text{yes}\} = n\})$. Now, let $(\vartheta_n)_{n \in \mathbb{N}}$ be an increasing computable sequence of natural numbers such that

$$\sum_{m=\vartheta_n+1}^{\infty} m \cdot 2^{-m} < 2^{-n}$$

for all $n \in \mathbb{N}$. Since, for all $m \in \mathbb{N}$, $m \cdot \lambda_m \leq m \cdot 2^{-m}$,

$$\sum_{m=\vartheta_n+1}^{\infty} m \cdot \lambda_m < 2^{-n}.$$

This implies that, for all $n \in \mathbb{N}$,

$$\left| \sum_{m \in \mathbb{N}} m \cdot \lambda_m - \sum_{m=1}^{\vartheta_n} m \cdot \lambda_m \right| < 2^{-n}.$$

For each n , $\lambda_n = \lambda(\{\alpha \in 2^{\mathbb{N}} : \#\{m \in \mathbb{N} : \ell(\alpha \upharpoonright m) = \text{yes}\} \geq n\}) - \lambda(\{\alpha \in 2^{\mathbb{N}} : \#\{m \in \mathbb{N} : \ell(\alpha \upharpoonright m) = \text{yes}\} \geq n+1\})$. Hence, $\{n \cdot \lambda_n\}_{n \in \mathbb{N}}$ is a sequence of uniformly computable reals. In turn, this implies that $\{\sum_{m=1}^{\vartheta_n} m \cdot \lambda_m\}_{n \in \mathbb{N}}$ is an increasing sequence of uniformly computable reals. Hence, $\sum_{m \in \mathbb{N}} m \cdot \lambda_m$ is computable. Thus, $\int_{2^{\mathbb{N}}} f d\lambda$ is not only finite, but also computable. By the equivalence of (1) and (2) from Theorem 1.2.15, it follows that ω is not Schnorr random.

(\Leftarrow) Suppose that ω is not Schnorr random. Our ultimate goal is defining a computable learning function that computably uniformly weakly detects that ω is patterned. The equivalence of (1) and (3) from Theorem 1.2.15 implies that there is a lower semi-computable function $f : 2^{\mathbb{N}} \rightarrow \overline{\mathbb{R}}$ with $\int_{2^{\mathbb{N}}} f d\lambda = 1$ and such that $f(\omega) = \infty$. Since f is lower semi-computable, there is a sequence of uniformly computable functions $g_k : 2^{<\mathbb{N}} \rightarrow \mathbb{Q}$ satisfying conditions (1)-(3) from Definition 1.2.12. Without loss of generality, we can assume that f is non-negative and that $g_0(\varepsilon) = 0$. Moreover, by Lemma 4.2.5, there is an unbounded sequence $\{r_n\}_{n \in \mathbb{N}}$ of uniformly computable reals such that, for all n , $\lambda(\{\alpha \in 2^{\mathbb{N}} : f(\alpha) = r_n\}) = 0$ and the measures $\lambda(\{\alpha \in 2^{\mathbb{N}} : f(\alpha) \geq r_n\})$ are computable

reals, uniformly in n . Let $\{s_n\}_{n \in \mathbb{N}}$ be a computable, increasing subsequence of $\{r_n\}_{n \in \mathbb{N}}$ such that $s_n \geq 2^n$ for all n . Now, let the following functions be defined by simultaneous induction (due to the interdependence of the functions).

- (1) Let $j : 2^{<\mathbb{N}} \rightarrow \mathbb{N}$ be such that $j(\varepsilon) = 0$ and, for $\sigma \neq \varepsilon$,

$$j(\sigma) = \begin{cases} \max\{i \in \mathbb{N}^+ : g_{|\sigma|}(\sigma) \geq s_i > g_{|\sigma|-1}(\sigma^-)\} & \text{if } \{i \in \mathbb{N}^+ : g_{|\sigma|}(\sigma) \geq s_i > \\ & g_{|\sigma|-1}(\sigma^-)\} \text{ is non-empty;} \\ j(\sigma^-) & \text{otherwise.} \end{cases}$$

Then, j is non-decreasing: if $\sigma \sqsubseteq \tau$, then $j(\sigma) \leq j(\tau)$. Since $s_n \geq 2^n$ for all n , $j(\sigma)$ essentially keeps track of the highest power of two that has been jumped over by the computable approximation witnessing the lower semi-computability of f after having observed σ .

- (2) Let $c : 2^{<\mathbb{N}} \rightarrow \mathbb{Z}$ be such that $c(\varepsilon) = 0$ and, for $\sigma \neq \varepsilon$,

$$c(\sigma) = \begin{cases} j(\sigma) - y(\sigma) & \text{if } j(\sigma) > j(\sigma^-) \text{—function } y \text{ is defined in (4) below;} \\ c(\sigma^-) & \text{if } j(\sigma) = j(\sigma^-) \text{ and } c(\sigma^-) = 0; \\ c(\sigma^-) - 1 & \text{if } j(\sigma) = j(\sigma^-) \text{ and } c(\sigma^-) \neq 0. \end{cases}$$

Function c works as a counter: for a given σ , $c(\sigma)$ represents the number of yes's that a learning function still needs to output on extensions of σ in order to match $j(\sigma)$.

- (3) Let $\ell : 2^{<\mathbb{N}} \rightarrow \{\text{yes, no}\}$ be the learning function given by $\ell(\varepsilon) = \text{no}$ and, for $\sigma \neq \varepsilon$,

$$\ell(\sigma) = \begin{cases} \text{yes} & \text{if } j(\sigma) > j(\sigma^-) \text{ or } c(\sigma^-) > 0; \\ \text{no} & \text{otherwise.} \end{cases}$$

- (4) Let $y : 2^{<\mathbb{N}} \rightarrow \mathbb{N}$ be such that $y(\varepsilon) = 0$ and, for $\sigma \neq \varepsilon$,

$$y(\sigma) = \begin{cases} y(\sigma^-) + 1 & \text{if } \ell(\sigma) = \text{yes;} \\ y(\sigma^-) & \text{otherwise.} \end{cases}$$

The purpose of function y is keeping track of the number of yes's output by the learning function ℓ up to the current stage.

Now, since the g_k 's are uniformly computable, j is computable. This implies that c , ℓ and y are computable, as well. Next, we need to show that

- (a) $\ell(\omega \upharpoonright m) = \text{yes}$ for infinitely many $m \in \mathbb{N}$,
- (b) $\lambda(\{\alpha \in 2^{\mathbb{N}} : \#\{m \in \mathbb{N} : \ell(\alpha \upharpoonright m) = \text{yes}\} \geq n\}) \leq 2^{-n}$ for all $n \in \mathbb{N}$, and
- (c) the measures $\lambda(\{\alpha \in 2^{\mathbb{N}} : \#\{m \in \mathbb{N} : \ell(\alpha \upharpoonright m) = \text{yes}\} \geq n\})$ are computable reals, uniformly in n .

For (a), recall that $f(\omega) = \infty$. Lemma 4.2.2 still holds if i is replaced by s_i . Thus, for every $i \geq 1$, there is some $m_0 \geq 1$ with $g_{m_0}(\omega \upharpoonright m_0) \geq s_i > g_{m_0-1}(\omega \upharpoonright m_0 - 1)$. For each of these infinitely many m_0 's, $j(\omega \upharpoonright m_0) > j(\omega \upharpoonright m_0 - 1)$. Hence, $\ell(\omega \upharpoonright m) = \text{yes}$ for infinitely many m .

To prove (b) and (c), we rely on two intermediate lemmas. (Lemma 4.2.8 below also establishes that function c , the counter, is in fact non-negative.)

Lemma 4.2.8. *For all $\sigma \in 2^{<\mathbb{N}}$, $j(\sigma) \geq y(\sigma)$.*

Proof of Lemma 4.2.8. The proof proceeds by induction. For ε , we have that $j(\varepsilon) = 0 = y(\varepsilon)$. For $\sigma \neq \varepsilon$, suppose that $j(\sigma^-) \geq y(\sigma^-)$. If $j(\sigma) > j(\sigma^-)$, then

$$\begin{aligned} j(\sigma) &\geq j(\sigma^-) + 1 \\ &\geq y(\sigma^-) + 1 \\ &\geq y(\sigma). \end{aligned}$$

If, on the other hand, $j(\sigma) = j(\sigma^-)$, then there are two cases to consider.

- (1) If $c(\sigma^-) \leq 0$, then $\ell(\sigma) = \text{no}$ and $y(\sigma) = y(\sigma^-)$. Since $j(\sigma^-) \geq y(\sigma^-)$, $j(\sigma) \geq y(\sigma)$.
- (2) If $c(\sigma^-) > 0$, then let τ be the longest string with $\tau \sqsubset \sigma$ such that $j(\tau) > j(\tau^-)$. Such a string must exist, for otherwise $c(\sigma^-)$ could not be positive, by the definition of c . Then, for all $\tau \sqsubset \rho \sqsubseteq \sigma$, $j(\rho) = j(\rho^-)$, which means that, for all such ρ , $j(\rho) = j(\tau)$. In particular, $j(\sigma) = j(\tau)$. Since $c(\sigma^-) > 0$, $c(\rho) > 0$ for all $\tau \sqsubset \rho \sqsubset \sigma^-$, as well. For suppose there were some ρ' with $\tau \sqsubset \rho' \sqsubset \sigma^-$ and $c(\rho') = 0$. Then, since $j(\rho) = j(\rho^-)$ for all $\tau \sqsubset \rho \sqsubseteq \sigma^-$, by definition, the value of c would remain the same on all initial segments of σ^- extending ρ' , including σ^- itself. Thus, contrary to our assumption, we would have that $c(\sigma^-) = 0$.

Since $j(\tau) > j(\tau^-)$, $j(\tau) = y(\tau) + c(\tau)$. Moreover, for each $\tau \sqsubset \rho \sqsubseteq \sigma$, $c(\rho) = c(\rho^-) - 1$ and $y(\rho) = y(\rho^-) + 1$. Hence, for all such ρ , $y(\rho) + c(\rho) = y(\tau) + c(\tau)$. This implies that

$$\begin{aligned} j(\sigma) &= j(\tau) \\ &= y(\tau) + c(\tau) \\ &= y(\sigma) + c(\sigma) \\ &\geq y(\sigma), \end{aligned}$$

where the last inequality follows from the fact that $c(\sigma) = c(\sigma^-) - 1 \geq 0$, as $c(\sigma^-) > 0$. \dashv

Lemma 4.2.9. *For all $n \in \mathbb{N}^+$,*

$$\{\alpha \in 2^{\mathbb{N}} : \#\{i \in \mathbb{N} : \ell(\alpha \upharpoonright i) = \text{yes}\} \geq n\} = \{\alpha \in 2^{\mathbb{N}} : f(\alpha) \geq s_n\}$$

up to a set of measure zero.

Proof of Lemma 4.2.9. (\subseteq) Let $n \geq 1$ and $\eta \in 2^{\mathbb{N}}$ with $\#\{i \in \mathbb{N} : \ell(\eta \upharpoonright i) = \text{yes}\} \geq n$. Then, there are at least n initial segments of η on which ℓ outputs **yes**. Let $\eta \upharpoonright m_1, \dots, \eta \upharpoonright m_n$ be the first n such initial segments. Suppose that $g_{m_i}(\eta \upharpoonright m_i) < s_n$ for all $1 \leq i \leq n$. Then, $j(\eta \upharpoonright m_n) \leq n - 1$, while $y(\eta \upharpoonright m_n) = n$, so $j(\eta \upharpoonright m_n) < y(\eta \upharpoonright m_n)$. This, however, contradicts Lemma 4.2.8. So, there must be some $1 \leq i \leq n$ with $g_{m_i}(\eta \upharpoonright m_i) \geq s_n$. Hence, $f(\eta) \geq s_n$. Therefore, $\{\alpha \in 2^{\mathbb{N}} : \#\{i \in \mathbb{N} : \ell(\alpha \upharpoonright i) = \text{yes}\} \geq n\} \subseteq \{\alpha \in 2^{\mathbb{N}} : f(\alpha) \geq s_n\}$.

(\supseteq) Let $n \geq 1$ and suppose that $f(\eta) \geq s_n$. If $f(\eta) = \infty$, then $\#\{i \in \mathbb{N} : \ell(\eta \upharpoonright i) = \text{yes}\} = \infty$. On the other hand, if $f(\eta) < \infty$, let $k \geq n$ be maximal with $f(\eta) \geq s_k$. If $f(\eta) > s_k$, then there is some $m \geq 1$ such that $g_m(\eta \upharpoonright m) \geq s_k > g_{m-1}(\eta \upharpoonright m - 1)$. By the maximality of k , $j(\eta \upharpoonright m) = k > j(\eta \upharpoonright m - 1)$. By the definition of ℓ , $\ell(\eta \upharpoonright m + i) = \text{yes}$ for all $0 \leq i \leq c(\eta \upharpoonright m)$, and $\ell(\eta \upharpoonright m + i) = \text{no}$ for all $i > c(\eta \upharpoonright m)$. By the definition of c , $j(\eta \upharpoonright m) = y(\eta \upharpoonright m) + c(\eta \upharpoonright m)$. Hence, $y(\eta \upharpoonright m) + c(\eta \upharpoonright m) = k$, which entails that $\#\{i \in \mathbb{N} : \ell(\eta \upharpoonright i) = \text{yes}\} = k \geq n$. If $f(\eta) = s_k$, on the other hand, then η belongs to a null set: by choice of the sequence $\{s_n\}_{n \in \mathbb{N}}$ (and the sequence $\{r_n\}_{n \in \mathbb{N}}$), $\lambda(\{\alpha \in 2^{\mathbb{N}} : f(\alpha) = s_k\}) = 0$. It follows that $\bigcup_{m \geq n} \{\alpha \in 2^{\mathbb{N}} : f(\alpha) = s_m\}$ is a null set, which, in turn, implies that $\{\alpha \in 2^{\mathbb{N}} : f(\alpha) \geq s_n\} \subseteq \{\alpha \in 2^{\mathbb{N}} : \#\{i \in \mathbb{N} : \ell(\alpha \upharpoonright i) = \text{yes}\} \geq n\}$ up to a set

of measure zero. ⊣

Now, $\int_{2^{\mathbb{N}}} f d\lambda = 1$ by assumption. Hence, Markov's inequality entails that, for every n ,

$$\lambda(\{\alpha \in 2^{\mathbb{N}} : f(\alpha) \geq s_n\}) \leq \frac{1}{s_n} \int_{2^{\mathbb{N}}} f d\lambda = \frac{1}{s_n}.$$

So, for every n , $\lambda(\{\alpha \in 2^{\mathbb{N}} : f(\alpha) \geq s_n\}) \leq 2^{-n}$. By Lemma 4.2.9 (and the obvious fact that $\lambda(\{\alpha \in 2^{\mathbb{N}} : \#\{i \in \mathbb{N} : \ell(\alpha \upharpoonright i) = \text{yes}\} \geq 0\}) = 1$), for all n , $\lambda(\{\alpha \in 2^{\mathbb{N}} : \#\{i \in \mathbb{N} : \ell(\alpha \upharpoonright i) = \text{yes}\} \geq n\}) \leq 2^{-n}$, which establishes (b).

For (c), recall that the measures $\lambda(\{\alpha \in 2^{\mathbb{N}} : f(\alpha) \geq s_n\})$ are computable reals, uniformly in n , since $\{s_n\}_{n \in \mathbb{N}}$ is a computable subsequence of $\{r_n\}_{n \in \mathbb{N}}$. This, together with Lemma 4.2.9, allows to conclude that the measures $\lambda(\{\alpha \in 2^{\mathbb{N}} : \#\{i \in \mathbb{N} : \ell(\alpha \upharpoonright i) = \text{yes}\} \geq n\})$ are computable reals uniformly in n , too. Hence, ℓ is a computable learning function that computably uniformly weakly detects that ω is patterned. □

Theorem 4.2.7 establishes that, by restricting attention to a certain collection of effective measure-zero patterns (as specified by Definition 4.2.6), Schnorr randomness too can be given a learning-theoretic characterisation. This result concludes our present investigation into the expressive power of this learning-theoretic framework.

4.3 Conclusion

In this chapter, we further explored the learning-theoretic approach to algorithmic randomness first introduced by Osherson and Weinstein [2008]. Our main results are novel learning-theoretic characterisations of both Martin-Löf randomness and Schnorr randomness. To recapitulate: weak 2-randomness, Martin-Löf randomness, Schnorr randomness, and weak 1-randomness can all be characterised within this framework.

Our characterisations of Martin-Löf randomness and Schnorr randomness rely on imposing further measure-theoretic constraints on the success sets of learning functions. It is thus natural to wonder whether these additional constraints can be weakened, or altogether eliminated, by suitably strengthening the definition of the corresponding success

criteria. In other words, is it possible to provide characterisations of Martin-Löf randomness and Schnorr randomness that do not depend on any measure-theoretic conditions other than requiring that the target patterns be measure-zero sets? Are there any success criteria more stringent than merely having to answer yes infinitely often (but weaker than having to answer yes cofinitely often) that can make up for the elimination of Condition (2) from Definition 4.2.3 or Condition (2) and Condition (3) from Definition 4.2.6, in favour of the weaker “ $\lambda(\mathcal{P}_\ell) = 0$ ”—where \mathcal{P}_ℓ denotes either success set defined via these alternative success criteria? Answering this question in the affirmative would allow to provide characterisations of Martin-Löf randomness and Schnorr randomness that share with the measure-theoretic typicality paradigm only very minimal assumptions. Answering this question in the negative, on the other hand, would highlight the tight dependence of Martin-Löf randomness and Schnorr randomness on measure-theoretic intuitions, thereby setting them apart from other algorithmic randomness notions that may be given a “purely learning-theoretic” characterisation, such as weak 1-randomness and weak 2-randomness.

On closer inspection, however, it is not obvious what providing a negative answer would exactly amount to. One would first have to specify a collection of allowable success criteria. Since all of the criteria discussed in this chapter are expressible by arithmetical sentences of varying complexity, an immediate issue is whether the above may be done by only appealing to arithmetical sentences. More generally, a natural question is whether there are any other algorithmic randomness notions (besides weak 1-randomness and weak 2-randomness) that can be characterised via success criteria of the following form:

Let ℓ be a computable learning function and φ an arithmetical sentence that features ℓ . Then, ℓ is said to detect that a sequence $\omega \in 2^{\mathbb{N}}$ is patterned relative to the success criterion determined by φ if and only if

- (1) $\omega \in \{\alpha \in 2^{\mathbb{N}} : \alpha \text{ satisfies } \varphi\}$, and
- (2) $\lambda(\{\alpha \in 2^{\mathbb{N}} : \alpha \text{ satisfies } \varphi\}) = 0$.

For instance, besides Martin-Löf randomness and Schnorr randomness, do computable randomness or some of the notions in the algorithmic randomness hierarchy that fall between 2-randomness (i.e., Martin-Löf randomness relative to the halting problem \emptyset') and Martin-Löf randomness—such as difference randomness [Franklin and Ng, 2011], Demuth

randomness [Demuth, 1982], and density randomness—admit reasonable learning-theoretic characterisations that comply with the above specifications? Addressing these questions would constitute a first step towards a systematic investigation of the expressivity of this learning-theoretic framework from the perspective of logical definability. This would allow to classify algorithmic randomness notions in terms of the definitional complexity of the learning-theoretic success criteria used for their characterisations.

One success criterion that might be worth studying with this goal in mind is the one below, which requires that the asymptotic density of the positive guesses of a learning function be 1:

Definition 4.3.1 (Limit detectability). *A learning function ℓ is said to detect in the limit that a sequence $\omega \in 2^{\mathbb{N}}$ is patterned if and only if*

- (1) $\omega \in \left\{ \alpha \in 2^{\mathbb{N}} : \lim_{n \rightarrow \infty} \frac{\#\{m \leq n : \ell(\alpha \upharpoonright m) = \text{yes}\}}{n+1} = 1 \right\}$, and
- (2) $\lambda\left(\left\{ \alpha \in 2^{\mathbb{N}} : \lim_{n \rightarrow \infty} \frac{\#\{m \leq n : \ell(\alpha \upharpoonright m) = \text{yes}\}}{n+1} = 1 \right\}\right) = 0$.

Intuitively, limit detectability targets measure-zero patterns whose presence becomes more and more evident to the learning function as the number of observations increases—in the sense that the relative frequency of affirmative answers goes to 1 in the limit. In terms of logical strength, limit detectability sits in between strong detectability and weak detectability. We leave it as an open question whether it yields a learning-theoretic characterisation of weak 2-randomness, weak 1-randomness, or any of the core algorithmic randomness notions that lie in between.⁸

Another related question worth investigating is whether there are any reasonable success criteria that, in spite of not giving rise to any standard algorithmic randomness notion, can nonetheless be used to define new learning-theoretic randomness concepts. Just as one may define new randomness notions in terms of effective versions of almost-everywhere theorems from analysis (as in the case of density randomness), there might in fact be natural randomness concepts that directly emerge from the learning-theoretic perspective on randomness discussed here.

⁸This question was answered by Steifer [2021], who showed that limit detectability yields a characterisation of weak 2-randomness.

The possible connections between algorithmic randomness and formal learning theory are not exhausted by the framework investigated here. In this chapter, we explored the use of formal learning-theoretic tools and concepts in the study of algorithmic randomness; however, it is also natural to ask what algorithmic randomness can do for formal learning theory. Some promising points of contact between the two fields focusing on this direction of the relationship were mentioned in the conclusion of Chapter 3 (in the context of convergence-to-the-truth results), and, as we hope our initial explorations suggest, many more remain to be uncovered.

More generally, our hope is that this dissertation has succeeded in bringing to the fore the many diverse ways in which the rich and conceptually illuminating framework offered by algorithmic randomness can fruitfully interact with the foundations of inductive learning—especially in the context of computationally limited, less-than-ideal learners.

Bibliography

- K. Ambos-Spies, K. Weihrauch, and X. Zheng. Weakly Computable Real Numbers. *Journal of Complexity*, 16: 676–690, 2000.
- A. Baltag and S. Smets. Keep changing your beliefs and aiming for the truth. *Erkenntnis*, 75(2): 255–270, 2011.
- A. Baltag, N. Gierasimczuk, and S. Smets. Truth-Tracking by Belief Revision. *Studia Logica*, 107: 917–947, 2019.
- G. Barmpalias, N. Fang, and F. Stephan. Equivalences between learning of data and probability distributions, and their applications. *Information and Computation*, 262: 123–140, 2018.
- V. Becher. Turing’s Normal Numbers: Towards Randomness. In S. B. Cooper, A. Dawar, and B. Löwe, editors, *How the World Computes*, pages 35–45, Springer, Berlin-Heidelberg, 2012.
- V. Becher and S. Figueira. An example of a computable absolutely normal number. *Theoretical Computer Science*, 279(1-2): 947–958, 2002.
- G. Belot. Bayesian orgulity. *Philosophy of Science*, 80(4):483–503, 2013.
- L. Bienvenu and W. Merkle. Constructive equivalence relations on computable probability measures. *Annals of Pure and Applied Logic*, 160: 238–254, 2009.
- L. Bienvenu, A. R. Day, M. Hoyrup, I. Mezhirova, and A. Shen. A constructive version of Birkhoff’s ergodic theorem for Martin-Löf random points. *Information and Computation*, 210: 21–30, 2012a.
- L. Bienvenu, N. Greenberg, A. Kučera, A. Nies, and D. Turetsky. K-Triviality, Oberwolfach randomness, and differentiability. *Mathematisches Forschungsinstitut Oberwolfach Preprint Series*, 2012b.

- L. Bienvenu, R. Hölzl, J. S. Miller, and A. Nies. Denjoy, Demuth, and Density. *Journal of Mathematical Logic*, 14(1): 1–35, 2014.
- L. Bienvenu, S. Figueira, B. Monin, and A. Shen. Algorithmic identification of probabilities is hard. *Journal of Computer and System Sciences*, 95: 98–108, 2018.
- D. Blackwell and L. Dubins. Merging of opinions with increasing information. *The Annals of Mathematical Statistics*, 33: 882–886, 1962.
- Boëthius and V. E. Watts. *The consolation of philosophy*. Penguin Books, London, 1999.
- V. Brattka, J. S. Miller, and A. Nies. Randomness and differentiability. *Transactions of the American Mathematical Society*, 368(1): 581–605, 2016.
- C. D. Broad. *Ethics and the History of Philosophy*. Routledge and Kegan Paul, London, 1952.
- C. Calude. Borel normality and algorithmic randomness. In G. Rozenberg and A. Salomaa, editors, *Developments in Language Theory*, pages 113–129, World Scientific, Singapore, 1994.
- R. Carnap. *The Continuum of Inductive Methods*. University of Chicago Press, Chicago, 1952.
- G. J. Chaitin. On the length of programs for computing finite binary sequences. *Journal of the Association for Computing Machinery*, 13: 547–569, 1966.
- G. J. Chaitin. A theory of program size formally identical to information theory. *Journal of the Association for Computing Machinery*, 22: 329–340, 1975.
- A Church. On the Concept of a Random Sequence. *Bulletin of the American Mathematical Society*, 46: 130–135, 1940.
- J. Cisewski, J. Kadane, M. J. Schervish, T. Seidenfeld, and R. Stern. Standards for Modest Bayesian Credences. *Philosophy of Science*, 85:53–78, 2018.
- A. D’Aristotile, P. Diaconis, and D. Freedman. On merging of probabilities. *Sankhyā: The Indian Journal of Statistics*, (50): 363–380, 1988.
- A. R. Day and J. S. Miller. Density, forcing, and the covering problem. *Mathematical Research Letters*, 3(22): 719–727, 2015.
- B. de Finetti. Funzione caratteristica di un fenomeno aleatorio. *Atti del Congresso Internazionale dei Matematici (Bologna, 3-10 settembre 1928)*, pages 179–190, 1929.

- B. de Finetti. La Pr evision: Ses lois logiques, se sources subjectives. *Annales de l'Institut Henri Poincar e*, 7: 1–68, 1937.
- O. Demuth. Borel types of some classes of arithmetical real numbers. *Commentationes Mathematicae Universitatis Carolinae*, 23(3):593–606, 1982.
- P. Diaconis and D. Freedman. On the consistency of Bayes estimates. *The Annals of Statistics*, 14(1): 1–26, 1986.
- P. Diaconis and D. Freedman. On the uniform consistency of Bayes estimates for multinomial probabilities. *The Annals of Statistics*, 18(3): 1317–1327, 1990.
- J. L. Doob. *Stochastic Processes*. Wiley, New York, 1953.
- R. G. Downey and E. J. Griffiths. Schnorr randomness. *The Journal of Symbolic Logic*, 69: 533–55, 2004.
- R. G. Downey and D. R. Hirschfeldt. *Algorithmic Randomness and Complexity*. Springer, New York, 2010.
- R. G. Downey, E. J. Griffiths, and G. Laforte. On Schnorr and computable randomness, martingales, and machines. *Mathematical Logic Quarterly*, 50(6): 613–627, 2004.
- R. Durrett. *Probability: theory and examples*, volume 31 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, fourth edition, 2010.
- J. Earman. *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. MIT Press, Cambridge, 1992.
- W. Edwards, H. Lindman, and L. J. Savage. Bayesian statistical inference for psychological research. *Psychological Review*, 3(70): 193–242, 1963.
- A. Elga. Bayesian Humility. *Philosophy of Science*, 83: 305–323, 2016.
- J. N. Y. Franklin and K. M. Ng. Difference randomness. *Proceedings of the American Mathematical Society*, 139(1): 345–360, 2011.
- J. N. Y. Franklin and H. Towsner. Randomness and non-ergodic systems. *Moscow Mathematical Journal*, 14: 711–714, 2014.
- J. N. Y. Franklin, N. Greenberg, J. S. Miller, and K. M. Ng. Martin-L of random points satisfy Birkhoff’s ergodic theorem for effectively closed sets. *Proceedings of the American Mathematical Society*, 140(10): 3623–3628, 2012.

- C. E. Freer, B. Kjos-Hanssen, A. Nies, and F. Stephan. Algorithmic Aspects of Lipschitz Functions. *Computability*, 3(1): 45–61, 2014.
- P. Gács. Uniform test of algorithmic randomness over a general space. *Theoretical Computer Science*, 341(1-3): 91–137, 2005.
- H. Gaifman and M. Snir. Probabilities over rich languages, testing and randomness. *The Journal of Symbolic Logic*, 47: 495–548, 1982.
- C. Glymour. *Theory and Evidence*. Princeton University Press, Princeton, 1980.
- E. M. Gold. Limiting Recursion. *Journal of Symbolic Logic*, 30: 27–48, 1965.
- E. M. Gold. Language Identification in the Limit. *Information and Control*, 10: 447–474, 1967.
- N. Goodman. *Fact, Fiction and Forecast*. Harvard University Press, Cambridge, MA, 1983.
- P. Hertling and K. Weihrauch. Random elements in effective topological spaces with measure. *Information and Computation*, 181(1): 32–56, 2003.
- C. Howson. *Hume’s Problem: Induction and the Justification of Belief*. Oxford University Press, New York, 2000.
- M. Hoyrup and C. Rojas. Computability of probability measures and Martin-Löf randomness over metric spaces. *Information and Computation*, 207(7): 830–847, 2009.
- M. Hoyrup and C. Rojas. Absolute continuity of measures and preservation of randomness. Manuscript, pages 1–12, 2011.
- D. Hume. *A Treatise of Human Nature*. L. A. Selby-Bigge, editor, Clarendon Press, Oxford, reprinted from the 1739/40 edition, 1978.
- S. Huttegger. Merging of Opinions and Probability Kinematics. *The Review of Symbolic Logic*, 8: 611–648, 2015a.
- S. Huttegger. Bayesian Convergence to the Truth and the Metaphysics of Possible Worlds. *Philosophy of Science*, 82: 587–601, 2015b.
- S. Huttegger. *The Probabilistic Foundations of Rational Learning*. Cambridge University Press, Cambridge, 2017.
- S. Huttegger, S. Walsh, and F. Zaffora Blando. Algorithmic randomness and Lévy’s Upward Theorem. Manuscript, 2021.
- M. Hutter and A. Muchnik. Universal Convergence of Semimeasures on Individual Random

- Sequences. In S. Ben-David, J. Case, and A. Maruoka, editors, *Algorithmic Learning Theory*, pages 234–248, Springer, Berlin-Heidelberg, 2004.
- E. T. Jaynes. Prior probabilities. *Institute of Electrical and Electronic Engineers Transactions on Systems Science and Cybernetics*, pages 227–241, 1968.
- R. C. Jeffrey. *Contributions to the theory of inductive probability*. Ph.D. Dissertation, Princeton University, 1957.
- R. C. Jeffrey. *The Logic of Decision*. McGraw-Hill, New York, 1965.
- R. C. Jeffrey. Probable knowledge. In I. Lakatos, editor, *The Problem of Inductive Logic*, pages 166–180. North-Holland, Amsterdam, 1968.
- R. C. Jeffrey. Mises redux. In *Basic problems in methodology and linguistics (Proc. Fifth Internat. Congr. Logic, Methodology and Philos. of Sci., Part III, Univ. Western Ontario, London, Ont., 1975)*, volume 11 of *Univ. Western Ontario Ser. Philos. Sci.*, pages 213–222. D. Reidel, Dordrecht, 1977.
- E. Kalai and E. Lehrer. Weak and strong merging of opinions. *Journal of Mathematical Economics*, 23: 73–86, 1994.
- S. M. Kautz. *Degrees of Random Sets*. Ph.D. Dissertation, Cornell University, 1991.
- K. T. Kelly. *The Logic of Reliable Inquiry*. Oxford University Press, Oxford, 1996.
- K. T. Kelly. The Learning Power of Belief Revision. *Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge (TARK '98)*, pages 111–124, 1998.
- A. N. Kolmogorov. Three approaches to the quantitative definition of information. *Problems of Information Transmission*, 1: 1–7, 1965.
- S. A. Kurtz. *Randomness and Genericity in the Degrees of Unsolvability*. Ph.D. Dissertation, University of Illinois at Urbana-Champaign, 1981.
- H. E. Kyburg, Jr. *The Logical Foundations of Statistical Inference*. D. Reidel, Dordrecht, 1974.
- M. van Lambalgen. *Random Sequences*. Ph.D. Dissertation, Universiteit van Amsterdam, 1987a.
- M. van Lambalgen. Von Mises' definition of random sequences reconsidered. *Journal of Symbolic Logic*, 52: 725–755, 1987b.
- 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, editors,

- Statistics, probability and game theory: Papers in honor of David Blackwell*, pages 347–367, 1996.
- T. Lattimore and M. Hutter. On Martin-Löf (non-)convergence of Solomonoff’s universal mixture. *Theoretical Computer Science*, 588:2–15, 2015.
- E. Lehrer and R. Smorodinsky. Merging and Learning. *Lecture Notes-Monograph Series 30, Statistics, Probability and Game Theory: Papers in Honor of David Blackwell*, 30: 147–168, 1996.
- L. Levin and A. K. Zvonkin. The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Uspekhi Matematicheskikh Nauk*, 25: 85–127, 1970.
- L. A. Levin. On the notion of a random sequence. *Soviet Mathematics Doklady*, 14: 1413–1416, 1973.
- L. A. Levin. Uniform tests of randomness. *Doklady Akademii Nauk SSSR*, 227(1): 33–35, 1976.
- P. Lévy. *Théorie de l’addition des variables aléatoires*. Gauthier-Villars, Paris, 1937.
- C. I. Lewis. *An Analysis of Knowledge and Valuation*. Open Court, LaSalle, Illinois, 1946.
- M. Li and P. M. B. Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications*. Texts in Computer Science. Springer, New York, fourth edition, 2019.
- J. L. Lockett. *Convergence in Total Variation of Predictive Distributions: Finite Horizon*. Ph.D. Dissertation, Stanford University, 1971.
- P. Martin-Löf. The Definition of a Random Sequence. *Information and Control*, 9: 602–619, 1966.
- W. Merkle, N. Mihailović, and T. A. Slaman. Some results on effective randomness. *Theory of Computing Systems*, 39(5): 707–721, 2006.
- R. I. Miller and C. W. Sanchirico. The role of absolute continuity in “merging of opinions” and “rational learning”. *Games and Economic Behavior*, 1-2(29): 170–190, 1999.
- R. von Mises. Grundlagen der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 5: 52–99, 1919.
- R. von Mises. *Probability, Statistics and Truth*. Dover, New York, second revised English edition, 1981.

- K. Miyabe. L^1 -Computability, Layerwise Computability and Solovay Reducibility. *Computability*, 2(1): 15–29, 2013a.
- K. Miyabe. Characterization of Kurtz Randomness by a Differentiation Theorem. *Theory of Computing Systems*, 52(1): 113–132, 2013b.
- K. Miyabe. Algorithmic randomness over general spaces. *Mathematical Logic Quarterly*, 60(3): 184–204, 2014.
- K. Miyabe, A. Nies, and J. Zhang. Using almost-everywhere theorems from analysis to study randomness. *Bulletin of Symbolic Logic*, 22(3): 305–331, 2016.
- A. A. Muchnik, A. L. Semenov, and V. A. Uspensky. Mathematical metaphysics of randomness. *Theoretical Computer Science*, 207(2): 263–317, 1998.
- M. Nielsen and R. T. Stewart. Persistent disagreement and polarization in a Bayesian setting. *The British Journal for the Philosophy of Science*, pages 1–31, 2019.
- A. Nies. *Computability and Randomness*. Oxford University Press, 2009.
- D. Osherson and S. Weinstein. Recognizing strong random reals. *The Review of Symbolic Logic*, 1(1): 56–63, 2008.
- D. Osherson, M. Stob, and S. Weinstein. *Systems that Learn*. MIT Press, First Edition, Cambridge, MA, 1986.
- J. C. Oxtoby. *Measure and Category: A Survey of the Analogies between Topological and Measure Spaces*. Springer-Verlag, New York, 1980.
- N. Pathak, C. Rojas, and S. G. Simpson. Schnorr randomness and the Lebesgue differentiation theorem. *Proceedings of the American Mathematical Society*, 142(1): 335–349, 2014.
- C. S. Peirce. Fraser’s “The Works of George Berkeley”. *North American Review*, 113: 449–72, 1871.
- C. S. Peirce. *The Collected Papers of Charles Sanders Peirce*, volume 5. Belknap Press, Cambridge, MA, 1965.
- M. B. Pour-El and J. I. Richards. *Computability in analysis and physics*. Springer, 1989.
- H. Putnam. ‘Degree of Confirmation’ and Inductive Logic. In P. A. Schilpp, editor, *The Philosophy of Rudolf Carnap*, pages 761–783. Open Court, LaSalle, Illinois, 1963.
- H. Putnam. Trial and Error Predicates and a Solution to a Problem of Mostowski. *Journal of Symbolic Logic*, 30: 49–57, 1965.

- F. P. Ramsey. Truth and Probability. In Richard B. Braithwaite, editor, *Foundations of Mathematics and Other Logical Essays*, pages 156–198. Routledge and Kegan Paul, London, 1931.
- R. D. Rosenkrantz. *Foundations and Applications of Inductive Probability*. Ridgeview Publishing, Atascadero, CA, 1981.
- J. Rute. Algorithmic randomness, martingales, and differentiability I. Preprint, 2012.
- J. Rute. *Topics in algorithmic randomness and computable analysis*. Ph.D. Dissertation, Carnegie Mellon University, 2013.
- J. Rute. Computable randomness and betting for computable probability spaces. *Mathematical Logic Quarterly*, 62(4-5): 335–366, 2016.
- J. Rute. Schnorr randomness for noncomputable measures. *Information and Computation*, 258:50–78, 2018.
- L. J. Savage. *The Foundations of Statistics*. John Wiley & Sons, New York, 1954.
- L. J. Savage. Implications of Personal Probability for Induction. *The Journal of Philosophy*, (64):593–607, 1967.
- M. Schervish and T. Seidenfeld. An approach to consensus and certainty with increasing evidence. *Journal of Statistical Planning and Inference*, 3(25): 401–414, 1990.
- C. P. Schnorr. Zufälligkeit und Wahrscheinlichkeit. In *Lecture Notes in Mathematics*, 218, Springer-Verlag, Berlin-Heidelberg-New York, 1971a.
- C. P. Schnorr. A unified approach to the definition of a random sequence. *Mathematical Systems Theory*, 5: 246–258, 1971b.
- B. Skyrms. *Pragmatics and Empiricism*. Yale University Press, New Haven and London, 1984.
- B. Skyrms. *From Zeno to Arbitrage*. Oxford University Press, Oxford, 2012.
- R. J. Solomonoff. A formal theory of inductive inference, I and II. *Information and Control*, 7: 1–22 and 224–254, 1964.
- R. Solovay. Draft of a paper (or a series of papers) on Chaitin’s work. Unpublished notes: 1–215, 1975.
- T. Steifer. A note on the learning-theoretic characterizations of randomness and convergence. *The Review of Symbolic Logic*, pages 1–16, 2021.

- T. F. Sterkenburg. Solomonoff prediction and Occam's razor. *Philosophy of Science*, 83 (4): 459–479, 2016.
- T. F. Sterkenburg. A generalized characterization of algorithmic probability. *Theory of Computing Systems*, 61(4): 1337–1352, 2017.
- T. F. Sterkenburg. *Universal prediction: A philosophical investigation*. Ph.D. Dissertation, Rijksuniversiteit Groningen, 2018.
- P. Suppes. A Bayesian approach to the paradoxes of confirmation. In *Studies in Logic and the Foundations of Mathematics*, volume 43, pages 198–207. Elsevier, 1966.
- J. Ville. Étude critique de la notion de collectif. In *Monographies des probabilités*, Gauthiers-Villars, Paris, 1939.
- P. M. B. Vitányi and N. Chater. Identification of probabilities is hard. *Journal of Mathematical Psychology*, 76, Part A: 13–24, 2017.
- V. V'yugin. Ergodic theorems for individual random sequences. *Theoretical Computer Science*, 207(2): 343–361, 1998.
- Y. Wang. *Randomness and Complexity*. Ph.D. Dissertation, University of Heidelberg, 1996.
- B. Weatherson. For Bayesians, Rational Modesty Requires Imprecision. *Ergo*, 2: 529–545, 2015.
- K. Weihrauch. *Computable Analysis. An Introduction*. Texts in Theoretical Computer Science. Springer-Verlag, Berlin, Heidelberg, 2000.
- D. Williams. *Probability with martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991.
- F. Zaffora Blando. A learning-theoretic characterisation of Martin-Löf randomness and Schnorr randomness. *The Review of Symbolic Logic*, 14(2): 531–549, 2021.