Chapter

Introduction

Earth was always strange and new to herself. . . . She loved the risk, the randomness, the lottery probability of a winner. . . . This planet that seems so obvious and inevitable is the jackpot. Earth is the blue ball with the winning number on it.

Jeanette Winterson—‘Weight’ (2005)

The topics discussed in this dissertation are all related to the philosophy of probability: we discuss the foundations of probability theory—a special topic in the philosophy of mathematics—and the epistemology of probabilistic information. Whereas epistemology traditionally deals with beliefs of only a single agent, we will also apply probability theory to describe the beliefs of groups of interacting agents (social epistemology).

Section 1.1 gives the rationale for this dissertation and makes its structure explicit. In section 1.2, we will define the branches of philosophy to which this dissertation belongs: formal philosophy and computational philosophy. In section 1.3, we give an overview of the foundations of probability and randomness. This section also describes the author’s position on the interpretation of probability, which is an epistemic one. Section 1.4 gives an overview of the mathematics and philosophy of the concept of infinity.

1.1 Motivation and structure

Il est remarquable qu’une science qui a commencé par la considération des jeux, ce soit élevée aux plus importants objects des connaissances humaines.

Pierre-Simon de Laplace (1814, p. 220)
Human knowledge is no solid rock of certainties. Scientific knowledge is often limited to the knowledge of probabilities, for example in quantum mechanics. In other fields, the relevant probabilities are not known exactly or it may not even be clear what the relevant possible outcomes are.

Because of the great importance of probability theory as a mathematical tool in all of the sciences, a first task for the philosophy of science is to critically investigate its foundations. In recent years, analyses in terms of probabilities have become more common in philosophy, too, in particular in (Bayesian) epistemology. This shows that philosophy has to be open-minded about new methodologies, but self-critical as well.

The first goal of this dissertation is related to the foundations of probability theory: to develop a mathematical basis for probability theory that allows dealing with infinite sample spaces in a way that is epistemologically satisfactory. We start from the application of non-standard analysis and non-standard measure theory to problems with a countably infinite number of possible outcomes. In the next step, we evaluate whether this solution also points towards:

- a solution for problems related to beliefs concerning finite sample spaces, in particular the Lottery Paradox;
- a more general approach that is also suitable for higher cardinalities.

The second goal is related to the application of probability theory to problems in formal epistemology. We will study an artificial group of interacting agents. The agents are modeled to have an opinion about a limited number of aspects of the world; their opinion is modeled as a theory about the world. We investigate the probability that an agent arrives at an inconsistent theory by updating his or her opinion based on the opinions of the other agents (in a specific way). We write a computer program to simulate groups of agents and analyze the data with special attention to the philosophical implication of the results.

Figure 1.1 makes the structure of this dissertation explicit. Although this work can be read in a linear way, from Chapter 1 to Chapter 6, its contents is only partially ordered. There are at least three distinct reading paths: one track focuses on the foundations of probability theory, a second one highlights the Lottery paradox in epistemology, and a third one focuses on social epistemology. Here is an overview of the contents of the following chapters:

Chapter 2 focuses on situations with infinitely many possible outcomes and investigates whether there is a sum-rule for probabilities that holds for such cases. In the mathematical treatment of probabilities, one has to make a distinction between cases in which the number of possible outcomes is finite, countably infinite (like the natural numbers), or non-countably infinite (like the real numbers). The countably infinite case is of particular interest to the philosophy of probability: it is the simplest case where problems with the sum-rule appear. Although countable additivity is one of the basic properties of Kolmogorov’s classical approach to probability theory (1933), de Finetti (1974) argued that only finite additivity is acceptable.

Although Chapter 2 is a technical study of a rather specific problem, it is related to other problems in the philosophy of probability, as well as to the philosophy of
1.1. Motivation and structure

mathematics in general, where problems related to infinity have always played an important role.

The general epistemological question “What do we know?” is closely related to the concept of probability. Many authors regard it as a necessary condition that we assign unit probability to a statement before its content can be part of our knowledge; for these authors, knowledge requires certainty. Others claim that knowledge is never absolute, but rather context-dependent; for them, knowledge requires “almost certainty”. They have to explain what it means that we can know something with “almost certainty”. The statement seems to indicate a very high probability, but how can we establish precisely how high the probability should be? We come back to this issue in Chapter 3. To give an adequate answer to this question, we have no choice but to investigate the nature of probability itself.

Chapter 4 focuses on the analogies between the topics discussed in Chapter 2 and Chapter 3.

Chapter 5 serves as an example of a study in computational philosophy: it gives a quantitative answer to the question of how likely it is that an agent arrives at an inconsistent theory by starting out with a consistent theory and updating over other agents who all hold a consistent theory. As such, this is also the only chapter in which groups of agents are studied instead of a single agent.

Figure 1.1: Three reading tracks for this thesis. (The number of eyes on each die indicates the chapter number.)
In Chapter 6, we evaluate the contents of this dissertation and sketch a plan for future work.

Multiple examples in this introduction come from the philosophy of physics, because the history of the theory of probability is closely linked to that of physics (in particular, statistical physics; see e.g. Uffink, 2007) and because the author is most familiar with this science.

1.2 From formal to computational philosophy

What do philosophers do? Twenty years ago, one might have heard such answers to this question as “analyze concepts” or “evaluate arguments”. The answer “write computer programs” would have inspired a blank stare and even a decade ago I wrote that computational philosophy of science might sound like the most self-contradictory enterprise in philosophy since business ethics.

Paul R. Thagard (1998, p. 48)

1.2.1 Formal philosophy

Formal philosophy uses techniques from mathematics (including logic) to analyze philosophical problems. Making mathematical analyses and modeling problems requires a way of thinking that is economical and constructive. Mathematics is a typically human activity, which certainly has its limitations and may require considerable effort, but nevertheless constitutes something we are good at (at least as a collective). In the sciences, the value of this way of thinking has already proven to be successful. Also in contemporary analytical philosophy, there is much interest in the formal approach. The use of formal methods in philosophy is far from a new development: logic always played a major role in Western philosophy, right from its beginning in ancient Greece. The new element in the current approach to formal philosophy is the broader interest for mathematical methods in general and probability theory in particular.

Famous examples of ancient Greek logics are those of Plato, Aristotle, and the Stoics. The word ‘logic’ stems from the Greek word ‘logos’ which can be translated as ‘word’, but also as ‘argument’, ‘logic’, or ‘reason’. Later philosophers such as Descartes, Leibniz, and Spinoza thought that we can learn about the world through reasoning and logic—a position called ‘rationalism’ (in contrast to the later empiricism, which gives more weight to sensory experience).

At the start of the twentieth century, there was a renewed interest in logic in philosophy: Frege and Russell worked on axiomatic formal logic, and Carnap and his Wiener Kreis started the tradition of logical empiricism, further developed by Reichenbach and his Berlin School. These were the seeds from which analytical philosophy blossomed. The position of an analytical philosopher is that philosophical problems can be (dis-)solved using logical methods, in particular first-order logic. However, logical empiricism was rejected by most philosophers of science in the 1960s. As a reaction, philosophers started to focus on the socio-historical dimension in the
development of science. A famous and extreme example of the social approach to the philosophy of science and technology is the work by Latour and Woolgar (1979), who observed scientists in their labs, much like anthropologists observed tribes in New Guinea.

In our time, the socio-historical approach is still important in the philosophy of science, but there is also a renewed interest in formal methods. Whereas the formalization of a problem originally meant ‘to translate the problem into the language of first-order logic’, now the term is used in a broader sense: ‘to rephrase (an essential part of) the problem in mathematical terms’. Although logic belongs to mathematics, so do probability theory, game theory, and graph theory—all of which are being applied in the philosophy of science. Examples of topics in the philosophy of science and epistemology to which mathematical methods have been applied include, but are not limited to: scientific laws and theories, scientific discovery and explanation, causality, confirmation, reduction, common knowledge, conditional reasoning, coherence, and judgment aggregation. Moreover, logic itself has also become a much broader field: originally, logic was just first-order logic, but now it also includes model theory (Tarski et al.), proof theory, set theory, recursion and complexity theory (Gödel, Church, Turing, and Kleene), and intensional logic (Kripke).

Domotor (2001) distinguishes between three main directions in present-day philosophy of science: the set-theoretical predicate approach initiated by Suppes (1957), the topological state space view originally proposed by Beth (1961), and the structuralist program initiated by Stegmüller (1976). However, the number of mathematical methods available to philosophers is not limited to these three: compared to the situation of almost one century ago, formally oriented philosophers now have a wide variety of tools to choose from, and their first task is to make this choice wisely. As Horsten and Douven (2008, p. 158) put it: “Finding the right formal framework for a problem is a highly nontrivial task. There is no general recipe for it.”

The latter observation alleviates at least one worry of those philosophers who fear that formalism will push away philosophical considerations: the choice and design of an appropriate formal method is itself not a formal affair, but a philosophical one. It requires critical reflection and a choice in the criteria that the method should meet. The same holds for evaluating the results of a formal analysis; it is not because the methodology is mathematical in nature that the results achieved by it are unrefutable: the method may not apply, or a more appropriate formalization may be found.

Like rephrasing a problem in the language of logic, using the language of any mathematical framework does have the benefit of being precise and explicit. Although the choice of mathematical formalisms has become much broader, the choice of doing a mathematical analysis may still seem like a form of narrow-mindedness. This is not entirely true: doing mathematics is a creative activity. The activity is not as rigid as the rigidity of the product it tries to achieve; quite the opposite, philosophers of science have reported a flourishing pluralism. Pedeferri and Friend (2010), for instance, argue in favor of methodological pluralism in mathematics.

We will come back to the use of probability theory as a tool in philosophy in subsection 1.2.3.
Chapter 1. Introduction

1.2.2 Computational philosophy

The increased computing power opens avenues for new research in formal philosophy. Of course, the increased power and availability of computers influence all researchers, including philosophers, whether they are formally oriented or not: the information revolution brought about by the emergence of the internet influences how scholars search for information and communicate with each other. Although this is an interesting topic in its own right, this is not what we mean by ‘computational philosophy’ here.

Computational philosophy can be understood in two ways: (1) as the philosophy concerning computer science or (2) as a way of doing philosophy, a way in which computation is an important tool. To give an example of the former, Thagard (1993) discusses epistemological issues that arise in the context of computer research. Here, however, we will be interested in the second meaning of computational philosophy as a part of formal philosophy, in which computation plays the role of preferred formal technique.

Computers can be used to run simulations, which allow philosophers to study phenomena of interest in isolation, disentangling them from other effects. Like the strictness of mathematics, this number-crunching is regarded with suspicion: these methods produce ‘results’, but do they provide insight into the underlying philosophical questions? Do they explain anything? The answer is: not on their own, of course. Simulations are merely research tools, which on their own do not solve any questions—philosophical or otherwise. They still require a researcher to select and interpret the data, to think about them, and forward conclusions. They can be a primary or secondary source of information, next to observations and intuitions.

Programming courses do not appear in a typical philosophy curriculum. So, if philosophers want to start performing simulations, they have to learn programming first (or at least learn how to rephrase their problems in such a way that a programmer can start working on them). According to Thagard (1998, p. 55), this is precisely the reason why computational philosophy has not seen wider acceptance so far:

Almost twenty years ago, Aaron Sloman (1978) published an audacious book, *The Computer Revolution in Philosophy*, which predicted that within a few years any philosopher not familiar with the main developments of artificial intelligence could fairly be accused of professional incompetence. Since then, computational ideas have had a substantial impact on the philosophy of mind, but a much smaller impact on epistemology and philosophy of science. Why? One reason, I conjecture, is the kind of training that most philosophers have, which includes little preparation for actually doing computational work. Philosophers of mind have often been able to learn enough about artificial intelligence to discuss it,

---

1 Of course, the two senses of ‘computational philosophy’ have a non-empty intersection: reflecting on computers and artificial intelligence may also spark new ideas in the philosophy of mind, the philosophy of science, and epistemology. For instance, Thagard (1988) presents a computational model of problem solving and discovery in science based on research in artificial intelligence.
but for epistemology and philosophy of science it is much more useful to perform computations rather than just to talk about them.

It is not my intention to argue here for the introduction of programming courses in the philosophy curriculum. I would rather suggest another way to forward computational philosophy: starting interdisciplinary research projects, such as the one in which I had the opportunity to participate (i.e. the Formal Epistemology Project). This strategy requires flexibility from its partners: philosophers should learn about the possibilities and limitations of computer-aided research, whereas the programmers should learn about the research interests of the philosophers. It is a matter of learning each other’s language. This theme recurs in section 1.2.3.

1.2.3 Chances for philosophy

As mentioned at the beginning of this section, mathematics is a key ingredient of many branches of science, and in recent years, the use of formal methods has become more popular in philosophy, too. Here, we will focus on the prominent position of probability theory in the methodology of the sciences and of formal philosophy.

Not surprisingly, philosophers have drawn attention to the analogy of probability and their other tool of preference: logic. Carnap (1950), for instance, regarded probability theory an extension of first-order logic, in particular as a logic of partial entailment, and tried to base a theory of confirmation on probability.

One branch of contemporary philosophy that heavily relies on probability theory, is Bayesianism (Hartmann, 2008, Hartmann and Sprenger, 2010): the Bayesian school not only provides an interpretation of probability (see section 1.3.1), but also advocates the application of Bayesian analysis to particular problems in philosophy.

Another example in which probability plays an important role is the study of the opinions of groups of people, modeled as idealized agents. For physicists, this is a new application of their methods for describing many-particle systems (Lorenz, 2005); economists hope to model the complex, dynamical pattern of social interactions (Phan and Varenne, 2010); and philosophers apply it to study how humans share knowledge and how they could improve the process to come closer to the truth (Hegselmann and Krause, 2002).² Precisely as in classical physics, combinatorics and probability theory may be applied here to summarize the torrent of information.

Although mathematical models, including those based on probability, are powerful tools, the foundations of those models are not free of problems, as will be discussed in subsections 1.3 and 1.4. Should we regard the use of probability theory—with all its problems and paradoxes—as the introduction of a Trojan Horse into the bastion of philosophy?

As philosophers, we cannot ignore the fact that there are problems in the foundations of the methods we use. On the other hand, it would not be wise to abstain from a powerful mathematical technique just because there are problems associated with it. Observe that there are problems with informal modes of reasoning, too, so there

²For an explanation of ‘closer to the truth’, see Kuipers (1987).
is no problem free alternative available. Any methodological choice should strike a balance between advantages and weaknesses.

In the application of probability theory lies a great opportunity for philosophy—we should take that chance! New branches of mathematics have been developed because mathematicians responded to interesting problems in the sciences, in particular physics. As a result, these domains can now profit from a tailor-made mathematical set of instruments. Philosophy may also benefit from this effect if philosophers learn to pose their questions in a formal language. Once philosophy has such a mathematical tool-set available, the effectiveness of this tool may appear unreasonable—just as happened previously, in the natural sciences (Wigner, 1960).

This dissertation is a contribution to the contemporary philosophy of probability. Epistemology requires a thorough study of the concept of probability, both in the objective, or physical, and in the subjective, or epistemic, sense of the word.

We develop a framework for probability theory that is capable of describing a fair lottery with a countably infinite number of tickets. With classical probability theory alone (in the sense of Kolmogorov), it is impossible to assign equal probabilities to all the tickets in such a lottery. We investigate a meaningful extension of the concept of probability to include infinitesimal probabilities. A suitable framework that allows the introduction of infinitesimal probability values is non-standard analysis. Once we have a system that is able to deal with a fair infinite lottery in an adequate way—both mathematically and philosophically—we can investigate whether lotteries on sets of larger cardinalities can be dealt with in a similar fashion.

The topic of this dissertation is not limited to a study of the concept of objective or physical probability in itself, but also in its relation to the formation of rational beliefs. The Lottery Paradox illustrates that the relation between probabilistic information and beliefs based on such information may be problematic, even for the simple case of a finite lottery.

1.2.4 Use with caution

The author has a very positive attitude towards the application in philosophy of mathematical methods in general and probability theory in particular. But despite this optimism, some warnings are at place here.

Philosophers who are interested in formal methods do not only need to learn how to apply them (or how to properly instruct someone else to do it for them), but also need to learn how to select a tool that is suitable for their case, how to model the problem in an efficient fashion, how to compare simulated data to other sources of information, and how to draw conclusions from it. Just as in the case of the natural scientists, who started using advanced mathematical methods and simulated data years ago, it is to be expected that philosophers will make beginners’ mistakes; a new way of looking at things always requires some time to adapt.

Formal or computational philosophy is a good place to try to apply what we have learned from the philosophy of science to philosophy itself: when we build models, they have to be simple enough, to keep their application manageable, and yet they have to be realistic enough to bear any significance towards the problem of interest.
This idea is sometimes referred to as the ‘KISS principle’; the letters stand for ‘Keep It Simple, Stupid!’.

After the analysis has been carried out, still another pitfall awaits us: that of confusing the model with reality. As in the sciences, we should be aware that the fundamental concepts or the structure of the mathematical model that we use to analyze a problem of interest do not necessarily correspond to the fundamental building blocks or the structure of the world. The interpretation of data based on a model always has to be done in such a way as to carefully distinguish the model from reality!

This warning may remind us of the words of Whitehead (1920, p. 163): “The aim of science is to seek the simplest explanations of complex facts. We are apt to fall into the error of thinking that the facts are simple because simplicity is the goal of our quest. The guiding motto in the life of every natural philosopher should be, Seek simplicity and distrust it.”

1.3 Foundations of probability and randomness

Comment oser parler des lois du hasard? Le hasard n’est-il pas l’antithèse de toute loi?

Joseph L. F. Bertrand (1888)

This large section gives an overview of two concepts that are intimately related: probability and randomness. Subsection 1.3.1 focuses on the various philosophical interpretations of probability. We will put forth an epistemic approach to objective probabilities in subsection 1.3.2 and propose a matching definition of a chance process in subsection 1.3.3. In subsection 1.3.4, we review the intuitive and the mathematical approaches to randomness. The generation of (pseudo-)random numbers is discussed in subsection 1.3.5. We will visualize different grades of certainty and the relation between probability and randomness in subsection 1.3.6. Subsection 1.3.7 closes this section with some thoughts on the relation between probability and luck.

1.3.1 Interpretations of probability

This subsection provides a brief, non-exhaustive overview of different interpretations of the concept of probability. For a more extensive treatment, consult Hájek (2007, 2008) for instance. The different interpretations are usually also related to a specific view of other concepts—such as certainty, possibility, and randomness—and may be derived from, or contribute to, a full philosophy of science.

The classical interpretation of probability is typical for the work of Laplace, but also for that of Bernouilli, Huygens, Leibniz and Pascal. An important element of the classical probability theory of Laplace (1814), is the ‘Principle of Insufficient Reason’ or the ‘Principle of Indifference’ (PI). This principle states that whenever no information is available to choose one possibility over another (e.g. due to symmetry), an equal probability should be assigned to those possibilities. However, because

\footnote{The acronym is attributed to an engineer, Kelly Johnson.}

\footnote{This name for the principle was introduced by Keynes (1921).}
possible outcomes can be labeled differently by different agents, the application of this simple idea is not without its problems, as has been illustrated with the paradoxes of Bertrand (1888) and later reactions to them (Jaynes, 1973, 1957, Seidenfeld, 1979). (We will come back to PI in subsections 1.3.3.2 and 1.4.2.2.)

A second interpretation of probability is frequentism: it assumes that probabilities are relative frequencies. Proponents of this idea were Reichenbach, Venn, and—most notably—von Mises. Just as with PI, the probabilities for finite references classes can only be rational numbers. As a reaction to this, the possibility of infinite references classes has been investigated, with the probability set equal to the limit of relative frequencies. De Finetti (1974) pointed out that limits of relative frequencies are not countably additive (cf. subsection 1.4.2.2).

Let us now distinguish between objective and subjective approaches. An objective or physical interpretation tries to identify probability as an intrinsic property of the physical system; this is also called the propensity-interpretation and has been advocated by Popper. This interpretation may seem appealing in relation to scientific theories which involve probabilities, but does not match well with many other situations in which the use of probability seems equally warranted. For instance, if we are faced with a process that has already taken place, such as a coin toss, but the result of it remains unknown to us until now, it seems natural to assign probabilities to the possible outcomes, although the actual outcome is already fixed. In this case, the probabilities assigned by us do not correspond directly to the propensity of the system, but rather to our limited knowledge.

The subjective interpretation does justice to this idea: it regards probability as a way to represent the information that a subject has about a system. In this view, probabilities may be used to describe a system, irrespective of whether that system is chance-like in nature or not. The subjective interpretation is also called the Bayesian interpretation or personalism. Some important subjectivists were Arnauld, de Finetti, Good, Jeffreys, Koopman, Lindley, Morgenstern, Ramsey, Savage, and von Neumann. Subjectivism is the dominant interpretation in the philosophy of probability theory today. It interprets probabilities as ‘degrees of belief’ of an individual ‘agent’: the probability an agent assigns to a statement is supposed to represent his or her confidence in the truth of that proposition. Subjective probabilities are often discussed in relation to wagers and betting strategies. Usually, money is at stake in these situations. Because sums of money are quantized, it is sometimes necessary to regard what is at stake as something more abstract: a continuous quantity called ‘utility’.

However, the purely subjective approach of Savage and other Bayesians has the drawback that a specific agent can never be considered to be wrong in the way he or she chooses his or her probabilities based on the available information. This drawback is alleviated by the intersubjective approach of Keynes, which introduces an appeal to what the agent should be able to conclude from the available information, in other words: norms of rationality which go beyond mere probabilistic coherence. This is also called the credence-based approach, for instance in contributions by Carnap and Lewis. Lewis (1986b) also proposed a way to relate subjective probabilities to objective probabilities by means of the so-called Principal Principle: if the objective
probabilities are known, our subjective probability estimations (degrees of belief) should be set equal to them.

Whereas both versions of the subjective interpretation can be called ‘epistemic’, whenever this term is used in this thesis, it will be reserved for the intersubjective version.

### 1.3.2 An epistemic approach to objective probability

When considering the foundations of probability theory, it is important to distinguish between properties of a mathematical model of the world and properties of the world itself. As an example, which I attribute to Vieri Benci, consider the question of determinism, which is considered an important issue in the philosophy of science. We may establish whether or not a certain model is deterministic, but this is not sufficient to infer whether or not the world is deterministic, too.

For the same phenomenon, there may very well exist a deterministic model next to an indeterministic (stochastic) one. As an example, consider a chaotic system: such a system can be described by deterministic equations, but it is highly sensitive to boundary conditions. Because the values of the starting conditions can never be measured with sufficient precision to warrant long-term predictions, a stochastic model of the same system may be more useful.

A choice between a deterministic and an indeterministic model is available not only in case of chaotic systems. Werndl (2009) demonstrates “that every stochastic process is observationally equivalent to a deterministic system, and that many deterministic systems are observationally equivalent to stochastic processes”; models that are ‘observationally equivalent’ warrant the same predictions. Within this light, it is clear that the status of the model does not reveal the nature of reality. It does say something about us, however, that we often prefer the deterministic model—or at least I do.

The probabilities that occur in physical theories are called ‘objective’ probabilities. Yet we have just seen that this need not imply that the world is chance-like in nature, which makes it less appealing to interpret probabilities as an intrinsic physical property (propensity). Moreover, it is part of the very nature of science that our current theories may be refuted at some future point in time. This motivates an epistemic view of science in general—a view in which the current content of science is considered to be the best sense we could make of all the experiments conducted so far, but nothing final or absolute. Similar reasons are motives for the adoption of an epistemic approach to probability, too.

Despite my conviction that probability is best interpreted as an epistemic matter, a large part of this thesis is devoted to what is called ‘objective probability’. This may seem contradictory, and hence deserves further explanation. My interest in so-called objective probabilities is motivated by the role they play in our models of the world. After all, we do often reason under the assumption that the (objective) probabilities are such and such. Since this assumption concerns a model, it does not contradict my general, epistemic attitude towards probability. This is my (crude) summary of the view: “We use probabilities to try and handle uncertain outcomes. However, no
matter how sophisticated the models that we employ are, the bottom line is that we can never predict anything with certainty. We can be relatively certain, but no matter how high our confidence level, we may always end up being completely wrong.”

Let us look at this topic from a slightly different angle. There are two distinct ways in which one may gain information about the probabilities of a certain process.

(1) **From evidence to probability** The first way is the most natural one: one may get some information on the process (which devices are used, how they are used), together with a body of ‘evidence’, by which we mean past results of the process. Based on this information, one may model the outcomes of the process by a probability distribution, but one can never be certain whether future trials will match with the current model.\(^5\)

(2) **From probability to belief** The second way to receive information about probabilities is to simply assume a certain probability measure. This may happen in textbook examples, which usually do not offer a body of evidence to examine. Also when one buys a game with one or more dice, one does not sit down first to throw the dice a large number of times to verify whether they are fair. One just assumes that all faces have an equal probability of \(\frac{1}{6}\) to come up on top. This is reasonable: if the die is, in fact, heavily loaded, one will notice soon enough, and if it is just slightly off, it will probably not matter for the game. As long as the (small) bias is unknown, it seems irrelevant. Very often, we do not even try to identify the possible flaw of a coin, die, or other chance device, which allows us to assume fair odds. These are only games, of course, but assumption of a probability measure also happens when we learn science: when one learns a physical theory, such as quantum mechanics, one may accept certain probability measures without verifying the experiments oneself.

The two ways to come to accepting a probability measure pose different philosophical questions. Case (1) is deeply connected to the core of epistemological questions, such as the problem of induction. Many philosophers of science have worked on this problem, including but not limited to: Hume (1739–1740) (hence the name Hume’s problem), Popper (1959), Hempel (1981), Kelly (1996), Williamson (2002), and Taleb (2007). It is in this context that different interpretations of probability have been proposed (see also subsection 1.3.1). As indicated before, the epistemic or intersubjective account of probabilities sounds the most convincing to me. Because “science relies on intersubjectively available evidence” (p. 215), also Williamson (2002, Ch. 10) deals with what he calls ‘evidential probability’ in terms of “a form of objective Bayesianism” (p. 212) and credences which should be distinguished from outright belief. Williamson remarks that evidence itself, or at least the propositions we associate with a certain body of evidence, may be uncertain. He proposes a theory of higher-order probabilities and combines it with an account of margins of error.

\(^5\)To model also the uncertainty in the probability assignment, one may introduce interval-valued probabilities (cf. Dempster-Schafer theory Dempster, 1967), higher-order probabilities (Williamson, 2002), ranking functions (Spohn, 2009), or some other, more advanced system.
Interesting as the issue may be, it will not be treated in any depth in the current thesis.

Case (2) is somewhat removed from issues related to learning from evidence. It may be considered to be the realm ‘objective probability’, but this is misleading. From reflecting on the first case, we conclude that we can never know whether any process with such perfect odds as a fair lottery objectively exists. However, it is not necessary that a perfectly fair lottery exists—and can be identified as such, on top of that—in order that we be interested in this case. All that is needed to motivate further study of this category is the observation that we often assume that a chance process is characterized by such and such probabilities. While being aware that this assumption may never be perfectly applicable, we may still be interested to investigate the properties of our model of a chance process, rather than any real-world process: this approach is called . . . mathematics. It is precisely in this context that questions arise related to (countably) infinite lotteries—a case of which we can never have any direct evidence—as we will discuss in Chapters 2 and 4. This example poses an interesting topic in the philosophy of mathematics.

Case (2) can also be related to epistemology. Suppose that somebody tells you the probability measure he would use to describe a chance process, rather than giving you access to the evidence on which he bases this model. What should you believe in that case? The question of rational beliefs based on knowledge of a precise probability distribution is taken up in Chapter 3.

Because Bayesianism is an important school that holds an epistemic view of probability, we should establish whether the current view belongs to it or not. I am not convinced by the Bayesian discourse in general, neither as a methodology nor as a philosophy, for reasons similar to those forwarded by Cousins (1995) and Gelman and Shalizi (2010). An essential ingredient to many branches of Bayesianism is the view that conditional probabilities are more fundamental quantities than unconditional probabilities. This comes close to the epistemic approach that I endorse, but the approaches are not completely identical. I do agree that probabilities always come with assumptions: in order to specify a probability value, one has to assume a certain set of possible outcomes, a certain set of variables, a certain form of the probability function, and so on. However, this is not what is expressed by conditional probabilities: even in conditional probabilities, many of these model-assumptions are tacit. So, whether considering conditional or unconditional probability values, one should always be aware that the model as a whole may be inaccurate, inadequate, or completely inappropriate in the given application.

A view that is much closer related to my own, is that of van Fraassen (1989). On the one hand, van Fraassen denies the existence of objective probabilities; on the other hand, he writes (p. 199): “when physics says that a radium atom has a 50 per cent probability of decaying within 1600 years, it says something about what the world is like, and nothing about opinion”. How do these two positions rhyme? It can be understood like this: the world does put constraints on the probability values that we can put in our models—if we replace the value of 50% in the example by 0.5% or by 5%, it will be easily refuted—but how this works precisely is beyond our grasp; in any case, it does not require a direct correspondence between the numbers.
in our model and a feature of the world. Van Fraassen (1989, p. 199) also offers a way to marry the use of objective chances to an epistemic, anti-realist interpretation of probability: accepting a theory means that the involved probabilities are taken to be the best available estimation (an expert function). In this view, the wisest thing to do is to align one’s own probability estimations and beliefs accordingly (cf. Lewis’ Principal Principle). Yet nothing in this view requires the objective probability to have any metaphysical status.

1.3.3 Definition of a chance process

First, we should indicate what we mean by a ‘chance process’. Various definitions can be found in the literature, but no formulation is completely neutral with respect to the interpretation of probability. Hence, we should select a definition that is compatible with the view of this thesis, which is the position that probabilities reflect the knowledge an agent has of a system, rather than an intrinsic property of the system itself. Moreover, we will phrase the definition in terms of a rational agent, rather than some actual person. This leaves open the possibility that an actual agent may be mistaken in his or her judgment about whether or not a given process is a chance process.

We will use the following definition for a chance process: if all the knowledge that is available about a certain process at a given time suffices for a rational agent to specify at least two possible outcomes for the next occurrence of the process, but does not suffice for the agent to predict the specific outcome that will be realized with certainty, the process is a chance process.

The weak spot in this definition is that it does not specify what it means to be rational. We cannot hope to give an ultimate answer to this issue here, for it requires a complete philosophical discussion of its own. It seems as though we should carefully phrase the rationality-constraint in such a way as to allow it to be verified. If not, the problem that a definition of chance in terms of an intrinsic property of the system requires a god’s-eye view, unattainable by any human being, will reappear in the context of rationality. We would indeed reintroduce the problem if we were to grant the rational agent an unlimited amount of time in which to formulate his or her conclusion on whether or not a process is chance-like. In contrast, one could demand that the agent should be able to formulate his or her conclusion in a finite time, or—more stringently—that the agent should be able to do so before the next occurrence of the process (or at least before he or she gets the knowledge of this outcome). A time constraint is a necessary but not a sufficient condition on the verifiability of rationality. Should we also specify which external means the rational agent is allowed (or expected) to use? An agent who has access to the internet, has access to a gigantic pool of information, provided that he or she masters the use of a search engine. But it seems strange to hard-code ‘googling skills’ into any definition of rationality.

Probably we should follow Williamson’s advice to “resist demands for an operational definition. . . . Sometimes the best policy is to go ahead and theorize with a

---

6This is similar to the definition which Stone (2008b, p. 7) gives for a lottery.
1.3. Foundations of probability and randomness

vague but powerful notion. One’s original intuitive understanding becomes refined as a result, although rarely to the point of a definition in precise pretheoretic terms” (Williamson, 2002, p. 211). So, for now, we just advise readers to insert their preferred account of rationality into the above definition. Do not worry if you are not fully satisfied at this point: just as you may update your opinion on whether or not a given process is chance-like, you may update your opinion on rationality at any time.

In the following two subsections, we will show why we also take an epistemic view on possible outcomes and review three different cases of odds: fair ones, weighted ones, and unknown odds.

1.3.3.1 Possible outcomes: from possible worlds to multiverse

Before we can assign any probabilities, we have to address a more fundamental question: what are possible outcomes?

The definition of a chance process that we have adopted speaks of ‘at least two possible outcomes’. These outcomes have to be specified as such in advance, for there are usually multiple ways to do this. For instance, the toss of a coin will always result in ‘some side up’, but this is only a single possible outcome; hence, with respect to this view, the process is not a chance process. In general, different descriptions of the same process may be regarded as different chance processes. There may also be cases in which the possible outcomes are unknown: Goodin (1978) speaks in this context of ‘profound uncertainty’ (see also subsection 1.3.6).

In the rest of this section, we focus on cases where we can identify multiple possible outcomes. Just as we have taken an epistemic attitude regarding the concept of probability, we will do so regarding possible outcomes. To explain why, let us start from the more naive position that identifies possible outcomes with different worlds.

In classical physics, probability theory is used to do calculations of situations in which we do not possess knowledge of all information, even though it is in principle available, or is used (in statistical physics) to compress complex information regarding large numbers of particles. The situation seems to be different in quantum mechanics, where probabilities are considered to say something fundamental about Nature itself: namely that She is indeterministic at the micro-scale.

To illustrate this, let’s consider two scenarios:

Scenario 1 - Head or tails. Before the throw of a coin, there are two possible outcomes: the coin may land with head or tails facing upwards (according to a certain plane of reference, such as a table). As soon as the piece has landed, it may show “heads”. What does it mean to say at this moment that it could have been “tails” as well? Has it not been proven, meanwhile, that this outcome was not possible: did it not just seem to be so?

Scenario 2 - Spin up or spin down. Replace the coin in the previous scenario with an electron that happens to be in a superposition of spin up and spin down (according to a certain axis). We can measure the spin along this reference axis and may find spin up or spin down—again, two possible outcomes. Suppose
the experiment is performed and results in spin up. After the experiment, what
does it mean to say that it also could have been spin down? Where did this
possibility go?

In both scenarios we are faced with the same question: “Where have the non-
actualized possibilities gone at the moment that one outcome becomes realized?”
However, there seems to be a distinction as well: in scenario 1, the process is said to
be deterministic, but we lack information about the exact circumstances of the throw
to be able to predict the outcome, whereas the process in scenario 2 is considered to
be indeterministic and thus intrinsically unpredictable.

What we actually demonstrated in scenario 2 is the collapse of the wave function:
according to the orthodox interpretation of quantum mechanics—the Copenhagen
interpretation—the wave function of a system that initially exists in a superposition
reduces irreversibly to a specific component during measurement. There are alter-
native interpretations of the theory in which no collapse occurs, such as the very
elegant multiverse-interpretation of Everett (1957). According to the latter interpre-
tation, every time that a particle or a system which is in a superposition is probed
for the relevant quantity, multiple worlds branch off, one for every possible outcome.
In scenario 2, this would imply that when we measure spin up, a parallel world has
branched off in which another version of us has found spin down and may wonder
whether it could have been just as well spin up. (Although we cannot phone or
otherwise communicate with our parallel counterpart, there is a sort of interaction
possible with close branches of the multiverse: via interference, another quantum
phenomenon with counterintuitive consequences.)

When I was still a student of physics, I was attracted to the many-worlds interpre-
tation of quantum mechanics: it gives an elegant explanation of a typical quantum
phenomenon. However, as illustrated by scenario 1, not all situations that confront
us with different possibilities can be reduced to a process that appears as a collapse
of the wave function; in such a case, no additional universes branch off in the multi-
verse. Yet, here too, there is an interpretation available that strongly resembles the
multiverse-story: with the help of modal logic, scenario 1 can be analyzed in terms
of possible worlds. David Lewis concluded that many problems with counterfactuals
have a simple solution: we just have to assume that the possible worlds really exist
(Lewis, 1986a).

Although the details are different, both Lewis (1986a) and Everett (1957) take
the step of setting possible worlds equal to actual worlds. If this is justified in both
scenarios, this leads us to the conclusion that there must be an enormous number of
worlds! It is advisable to proceed with caution: both interpretations are devised by
humans and thus an important question is whether we should believe that all these
worlds exist or rather that the multiverse-concept is a natural reaction of humans
when confronted with descriptions in terms of probabilities. For both scenarios, there
are alternative interpretations available for the involved probabilities. For situations
as in scenario 1 (coins, wheel of fortune, roulette, . . .), Abrams developed a mecha-
nistic interpretation of probabilities that does not rely on the use of counterfactuals.

Taking an epistemic approach to possible outcomes—i.e. relating possible out-
1.3. Foundations of probability and randomness

comes to our knowledge of the world rather than to some intrinsic property of the world(s)—solves many problems associated with counterfactuals, collapsing wave functions, and exploding numbers of worlds.

1.3.3.2 Odds: fair, weighted, or unknown

A chance process is called **fair** if its possible outcomes have equal probability. On the other hand, a chance process is **weighted** if there is at least one possible outcome that is more probable than some other possible outcome. Note that this does not exhaust all the possibilities: one may envisage a chance process about which nothing is known, except the set of possible outcomes. In such a case, the probabilities are **unknown**.

We will now look into the details of fair odds or equiprobability. Chapter 2 concerns the discussion of a fair infinite lottery. It may be objected that this is a non-problem, since infinite lotteries do not exist. However, it should also be noted that the idea of a fair finite lottery is a highly idealized concept! There is no way in which perfectly equal odds can be attained within a finite system. Coin tosses are a popular example in philosophical discussions of probability. Yet, the toss of a coin is a deterministic process, which can be described by classical mechanics. Moreover, for a real-world throw it is never true that heads and tails have equal probability; this is only obtained in the zero-friction limit, where the coin keeps bouncing back on the table for an infinite number of times. (See Diaconis et al. (2007) who conclude that coin tossing physics is not random.) Thus, the assumption of equiprobability is no less idealized than an infinite lottery. We may even conjecture that only an infinite process can produce perfectly fair odds.

Using the second Borel–Cantelli lemma (Milbrodt, 2010, II.4.D, p. 177–181), it can be proven that within a given, infinitely long string of characters (such as letters and punctuation), in which each character is chosen at random, any finite sequence of characters occurs almost surely (and actually, infinitely many times). A famous variant of this result is called the ‘infinite monkey theorem’, attributed to Émile Borel (Milbrodt, 2010, p. 179): one monkey hitting keys on a typewriter keyboard at random for an infinite amount of time will almost surely type the complete works of William Shakespeare (infinitely many times). As a child, I was fascinated by the suggestion that if space is infinite and contains an infinite amount of matter, any possible configuration of matter should exist somewhere—including planets that look just like ours except maybe for some small details.\(^7\) I could not believe that infinity entails this, and I think that I now understand why: for the argument to hold, it would also require randomness, which was not included in the story, and the qualification ‘almost surely’ also plays an important role. This boils down to the difficulty of interpreting unit probability, which does not entail logical necessity.

There is a curious relation between equal and unknown probabilities. In cases

\(^7\)What the source of this story was, I do not recall, but a contemporary variant of this idea can be found in Vilenkin (2006), who considers the possibility of eternal cosmic inflation and concludes (p. 112): “A striking consequence of the new picture of the world is that there could be an infinity of regions with histories absolutely identical to ours”.

with some freedom in the formulation of the problem, one can try to describe the system in such a way that all possibilities have an equal (but unknown) weight. With the help of combinatorics, the possibilities can be counted (call this number \( N \)) and subsequently the weight of a single possibility can be set equal to its inverse (\( \frac{1}{N} \)). This procedure may provoke the following question: how can we establish that the possibilities have the same probability if we do not yet know this probability? As we saw in subsection 1.3.1, in the classical probability theory of Laplace (1814), this is achieved by the ‘Principle of Indifference’ (PI), which states that whenever there is no information available for choosing one possibility over another, an equal probability should be assigned to those possibilities. However, because possible outcomes can be labeled differently by different agents, the application of this simple idea is not without its problems, as has been illustrated with the paradoxes of Bertrand (1888) and later reactions to them (Jaynes, 1973, Seidenfeld, 1979). (We will come back to this in subsection 1.4.2.2.) The main question seems to be: how can we model ignorance in a mathematically correct way, without adding information in the process? Stone (2008b, p. 25) states that in applying PI, we are “flouting Aristotle’s memorable advice, and imposing greater precision than the circumstances allow.” We should rather accept that there are cases with unknown probability, which simply cannot be modeled with fair or weighted odds.

We will come back to the cases of weighted and fair odds in subsection 1.3.6 and Figure 1.2, after reviewing the concept of randomness.

### 1.3.4 Measuring randomness

#### 1.3.4.1 Looking for patterns

Let us first consider a realistic example. In Belgium, the national lottery is performed by selecting six balls out of forty-two balls (numbered from 1 to 42). (Actually, there is also a seventh ball drawn, but this plays no role in assigning the first prize winner.) People who participate in the lottery have to indicate six numbers in advance of the lottery. If all six numbers correspond to a ball that actually gets drawn, they win the first prize (or have to share it with others who selected the same numbers). The lottery machine is supposed to have the effect that, at each draw, every ball contained in it has the same probability of being drawn. Combinatorics tells us that the number of ways to select 6 balls from a set of 42 (disregarding the order) is \( C_{42}^6 = \frac{42!}{(42-6)!6!} = 5 \, 245 \, 786 \). Hence, the probability of any particular outcome is \( \frac{1}{5 \, 245 \, 786} \) or about 0.000 000 19. Even without calculating this value, it is clear that any particular outcome has an equal probability. Yet, if we were to learn that this week’s lottery outcome happens to be the numbers 1, 2, 3, 4, 5, and 6, we would feel like something strange has happened. It may lead us to doubt that the

---

8 In the context of scenario 1 of the previous subsection, it can be argued that if a coin is a thin cylinder with a homogenous distribution of mass, there is no reason to assume that the coin will land on one side more often than on the other: heads or tails both get probability \( \frac{1}{2} \).

9 According to the list of past results on the website of the national lottery, found at http://www.lotto.be/NL/Spelen_en_Winnen/Trekkingsspelen/Lotto/Statistics, this outcome has never been realized yet.
lottery machine works properly or that the result of the process is communicated in all honesty.

Considering a similar example, Laplace (1814, p. 14) writes: “C’est ici le lieu de définir le mot extraordinaire. Nous rangeons par la pensée, tous les événements possibles, en diverses classes; et nous regardons comme extraordinaires, ceux des classes qui en comprennent un très petit nombre.” Some sequences appear to have clear patterns, others have more intricate patterns or do not seem to have any. We regard the first class to be a small one and thus to have a lower probability. If a result of this class shows up, we feel that it is something out of the ordinary.

Humans are good at recognizing patterns. Our potential for pattern recognition may be used as a quick and dirty way of determining whether a bit string is random or not: the bits can be transformed into a black and white image in which a periodical pattern appears if the numbers are produced by a pseudo-random generator (see subsection 1.3.5.1) rather than a true random process (Allen, 2010). On the other hand, when the data does consist of random noise, humans may still be under the misapprehension that there is a pattern in the data stream (Hake and Hyman, 1953). The human tendency to look for patterns even when this is not warranted has been dubbed ‘patternicity’ by Shermer (2008). Of course, the sequence ‘1, 2, 3, 4, 5, and 6’ of our lottery example does show a clear pattern. However, the balls are all similar and the numbers that are attributed to them function only as labels to distinguish them, not to order them. Instead of numbers, we might just as well have used some other list of symbols which does not suggest an ordering. This way, we may understand that this particular sequence is indeed no more special than any other outcome.

At first sight, arguing that ‘it’s just a label’ does not seem to work for the similar example of a long sequence of tosses with a fair coin, which all result in heads (as considered, for instance by Gács, 1978): no matter how you refer to heads or tails, this result means that the—supposedly fair—coin lands with the same face up each time. However, the coin tosses are supposed to be independent. If this is so, you should have the freedom to decide which face you call ‘heads’ in between each toss, and this should not matter. In that case too, the ‘always heads’ result may not seem so special anymore. However, our brain does signal the pattern, no matter how it is produced—with relabeling or not. So, the presence or absence of a pattern is still an interesting feature.

Intuitively, we call ‘random’ those processes whose results seem to show no patterns. The absence of a pattern in a list of past results implies that we have no basis for predicting a specific outcome for future occurrences of the same process. Since we lack certainty, we may characterize the process in terms of probabilities. Sometimes a partial prediction is possible: if in the past a specific outcome has occurred more often than any other, we have a good reason to bet on this outcome for the next manifestation of the process, all the more so, if the body of evidence based on former results of the evidence is large and the preference for the specific outcome is well-pronounced. In such a case, we cannot make a prediction with certainty, but the randomness is not maximal. Thus, we see that maximal randomness coincides with equiprobability.
The criterion of ‘absence of patterns in the results’ has been turned into a mathematical definition of randomness: this is the topic of the next subsection.

1.3.4.2 Mathematical approach to randomness

Here we give a very brief chronology of the developments in the mathematical study of randomness, based on the information in Bienvenu et al. (2009). All the approaches mentioned here focus on the randomness of individual infinite sequences of zeros and ones (bits). The infinite binary sequences live on the Cantor space, written as \( \{0, 1\}^\mathbb{N} \), \( 2^\mathbb{N} \), or \( 2^\omega \), and can be interpreted as representing real numbers or sets of natural numbers.

The history of randomness in a mathematical context begins in 1919 with von Mises’ study of the collective (‘Kollektiv’), by which he means a random sequence, defined in terms of limiting frequency and selection rules (von Mises, 1919, 1928).

In the 1930s, the topic is taken up by Wald, Ville, and Church. Ville (1939) replaces the selection rules by martingales, which “can be seen as describing the capital of a player who is trying to guess the bits of an infinite binary sequence, betting money (never more than his current capital) on their values, and is rewarded in a fair way” (Bienvenu and Merkle, 2007, p. 119–120). Church is the first to give a definition for the term ‘random sequence’.

In the 1940s and 1950s, the emphasis shifts to measure theory. In this context, randomness is defined in terms of a computable measure; sets that have weight one with respect to the chosen measure are called random. “Misbehaving frequencies and unbounded martingales are merely examples of sets of measure zero” (Bienvenu et al., 2009, p. 2).

In the 1960s, the relation of randomness to complexity is investigated. Researchers such as Solomonov, Kolmogorov, and later Chaitin propose to define random objects as objects of maximal complexity or minimal compressibility. Chaitin (1975, p. 4–5 of 1987-reprint) states his version of the definition as follows: “A series of numbers is random if the smallest algorithm capable of specifying it to a computer has about the same number of bits of information as the series itself.” In his book, Chaitin rephrases this as follows: “something is random if it is algorithmically incompressible or irreducible” (Chaitin, 2001, p. 111). Clearly, these definitions connect the topic of randomness to computability and information theory. An important development is that of Martin-Löf randomness, which implies that the notion of measure zero can also be made algorithmic (Martin-Löf, 1966).

In the 1970s, further work was done by Schnorr, Levin, and others. These developments led to the current algorithmic randomness theory.

Bienvenu and Merkle (2007) distinguish between two approaches to randomness. On the one hand, randomness may be studied in relation to the Law of Large Numbers, which deals with the convergence of frequencies. In this case, randomness is defined in terms of selection rules and the used measure is typically the uniform measure on a Cantor space. On the other hand, randomness may be defined in terms of betting strategies (martingales) and an arbitrary computable probability measure. This category encompasses different notions of randomness, including Martin-Löf
randomness, computable or recursive randomness, Schnorr randomness, and weak or Kurtz randomness.

Additional historical information on the development of randomness and probability theory can also be found in Vovk and Shafer (2003). For a more detailed treatment of the mathematics of randomness, the reader is referred to textbooks such as Nié’s (2008).

1.3.5 Producing (pseudo-)random numbers

Despite the absence of patterns in them, random numbers, processes and structures are very useful in data simulations: one may be interested in studying how a system evolves over time by considering many random start configurations (Monte Carlo method) or one may model the movement of particles at the micro-scale by so-called random walks. Instead of doing many calculations that involve a random start condition or random events in the system’s time evolution, one may also be interested in studying a system whose configuration is completely random: this would make it possible to do only one calculation that results in a very good estimate of the average of many similar, non-random systems. This idea may be applied in computational materials research, when investigating the ‘typical’ properties of a material consisting of a fixed proportion of atoms of different elements, but in an unspecified configuration. Representing a random configuration turns out to be very heavy computationally, because a random structure is non-periodic.

Because ‘random’ is defined as an absence of patterns, one would expect that a random system has no typical properties at all. One would also expect that it can never be approximated by a periodical system. It turns out that both assumptions are wrong: a random system can be characterized by specific numbers (statistical parameters) and it can be approximated by a periodical system, precisely by selecting a system that has a value close to that of the ideally random system for the relevant parameters. Relative to a particular application, these non-random systems may be more practical to use (easier to obtain), than their ideally random counterpart, without much loss in the quality of the results. Periodical configurations that are employed to resemble a random configuration in certain aspects are called pseudo- or quasi-random systems (Szemerédi, 1975). As Nagle et al. (2006) put it: “Roughly speaking, a quasi-random structure is one which, while deterministic, mimics the behavior of random structures from certain important points of view.”

1.3.5.1 Pseudo-random numbers

As a child, I was fascinated by our first home computer, a Commodore 64, and in particular its option to produce a random number. The command for this function was RND and it made me wonder how a computer could choose a number freely. It made me wonder whether the machine had a soul or at least a will of its own. Only later, much later, I learned that a computer does not produce random numbers at all, but performs a calculation that deterministically results in a number.

John von Neumann (1951) famously wrote:
Anyone who considers arithmetical methods of producing random digits is, of course, in a state of sin. For, as has been pointed out several times, there is no such thing as a random number—there are only methods to produce random numbers and a strict arithmetical procedure is not such a method.

In other words, an algorithm can at best produce a pseudo-random number.

The basic idea behind the pseudo-random function of a computer is to find an algorithm that starts from a number—called the seed—and then produces, by means of deterministic calculation, a long sequence of other numbers before returning to the seed state. The calculation can be based on an equation that describes a chaotic system. Assigning a good seed state is just as important as finding a useful equation: for some seeds, the cycle of numbers may be so short that the result will not look random at all. The problem with BASIC’s $RND$-function was precisely that it started from the same seed at every run of a program.

The first algorithm for pseudo-random number generation is called the middle-square method and was designed by von Neumann (Ulam et al., 1947, Metropolis, 1987). Another well-known example of an algorithm for pseudo-random numbers is the Mersenne twister algorithm (Matsumoto and Nishimura, 1998). Whereas it may suffice to use the internal clock to produce a random number for some applications, in many applications, the pseudo-random generators have to be cryptographically secure; this category includes examples such as stream and block ciphers, the Yarrow algorithm, the Micali-Schnorr algorithm, and the Blum Blum Shub algorithm (see e.g. Krhovják, 2006, for an overview). In this dissertation, we have used pseudo-random numbers in the simulated populations of Chapter 5.

Pseudo-random numbers are easy to produce, but since all lists of pseudo-random numbers are cyclic in nature, in effect, these numbers are proven to be non-random. Can we do better than this?

1.3.5.2 True random numbers

As Hayes (2001) remarks, we cannot really produce random numbers. (If we were to know how to achieve this, we would need to have a recipe. But if we have a short description or an algorithm, the numbers produced by it are not random, by definition.) What we actually do is more akin to mining of a natural resource.

Although it can never be proven that a certain process or sequence of numbers is really random (cf. subsection 1.3.6), there are some sources that are considered to be true random number generators. (Instead of writing ‘true’ random numbers, we just drop the quotes.) They do not necessarily rely on complicated machinery, but rather on the recording of natural sources of noise. The website random.org (Haahr, 1998–2010), for instance, provides true random numbers based on atmospheric noise.

Meanwhile, there are computers that are able to produce true random numbers. This requires a special piece of hardware, a built-in apparatus that employs a certain physical process, such as thermal noise or the photoelectric effect, for the sole purpose of generating random numbers.
There exist tabulated data of true random numbers. For instance, a book with a million random digits was published by the RAND corporation (1955), based on the results of their electronic true random number generator. What is strange about this approach is that the numbers can now be referred to as being in this specific book: this is a very concise description, and thus shows that the numbers are no longer random. However, they may still be useful for many applications, just like pseudo-random numbers.

If we need only a few random bits, how about tossing a coin? Does this produce truly random numbers? The movement of a coin that is tossed is deterministic; it can be described by classical mechanics—in principle at least, for in practice many small effects, such as air resistance, are often neglected in such a calculation. As we mentioned before, Diaconis et al. (2007) have shown that real-world coin flips do not produce perfectly fair odds. A large part of the chance-like nature of flipping a coin stems from the fact that we want to use it as a procedure to assign fair chances. Hence, we do not look at which face is up before we throw it up in the air, we do not try to control the toss, and at a soccer game we let the referee make the toss. All these factors help to keep coin tosses as a sufficiently good approximation to a source of random bits.

### 1.3.5.3 Making up random numbers

People are notoriously bad at choosing numbers at random. One example is ‘first digit bias’: when data are forged, the numbers usually fail Benford’s law (Durtshi et al., 2004), which states that in many applications (in particular, when the values range over multiple orders of magnitude) the leading digit of natural data is most often equal to 1. Benford’s law may seem a strange law, but it is easy to see that it should hold provided that the logarithm of the numbers, rather than the numbers themselves, are distributed in a uniform way. Note that pseudo-random numbers of computers are usually uniformly distributed numbers between 0 and 1, and hence do not (and should not) follow Benford’s law.

A second example of bias is that when making up numbers, people avoid equal or subsequent digits because they appear non-random, whereas this does occur to a certain extent in true random data, of course. As a third example, most people have a preference for even numbers, which may result in a too high percentage of even digits in made-up numbers. As a fourth example, results of a real experiment usually show a statistical spread, which is often much lower in forged results.

Even when the forger is aware of an existing bias, he may still reveal his fraud by producing data that are ‘too good to be true’: he will overcompensate for his natural tendencies, resulting in a too high percentage of odd digits (especially 7’s and 3’s), or a too high spread on the data, and so on (Buyse et al., 1999).

People may be tempted to falsify existing data (by selecting the favorable results or by changing some of the numbers) or to ‘produce’ the data by making up the numbers themselves. This may happen in different contexts, such as accountancy, elections, or scientific studies. The aforementioned failures can be used in forensic accounting (Nigrini, 1996, Durtshi et al., 2004), or to detect fraud in other data
(see e.g. Buyse et al., 1999, for a study on fraud in clinical trials). You can test your own (in-)ability to produce numbers that look random at this website: http://faculty.rhodes.edu/wetzel/random/intro.html.

1.3.6 Amount of certainty

Figure 1.2 offers a visual representation of the different positions on the certainty–uncertainty axis. This illustration was inspired by remarks of various authors, in particular by Stone (2008b).

Goodin (1978, p. 35) invites us to “distinguish two different levels of uncertainty. With the modest form, the uncertainties essentially surround our probability estimates. With the profound form, we are instead uncertain of the completeness of our list of alternative possibilities.” Thus, profound uncertainty refers to situations in which even the possible outcomes cannot be identified; this option is located on the righthand side of the (un)certainty-axis in Figure 1.2.

![Figure 1.2: Schematic representation of different situations involving more or less uncertainty. The situation in which all possible outcomes have an equal probability is an intermediate case in terms of (un)certainty, although it is maximal in terms of randomness. Chance processes with unequal probabilities, such as a weighted lottery, are po-](image-url)
tioned more towards the left side of the (un)certainty-axis. The lower axis in the figure suggests that **weighted chance** processes result in sequences of outcomes that are less random than those of fair chance processes. In order to see this, Chaitin (1970) considers a coin that is biased towards heads: in a long sequence of length $n$, it produces heads in approximately 75% of the tosses and tails in 25% of the tosses. The result can be represented by ones (for heads) and zeros (for tails). The computer program to compute the corresponding binary sequence only needs to be about 80% of $n$, the length of the sequence it computes Chaitin (1970, p. 7). This shows that introducing bias in a chance process indeed lowers the randomness of and increases the predictability of the produced results.

According to Edward Gibbon (1805, p. 122), the laws of probability are “so true in general, so fallacious in particular.” This quote seems particularly well-suited to describe a **fair chance** process: there is no good strategy for predicting a specific outcome of such a process, but the average of a long sequence of outcomes can be predicted very well. (Notice that for a process with unknown odds, the former also holds, but not the latter.) Chance processes with fair odds take up a special position in Figure 1.2: they are completely balanced between certainty and uncertainty, but they are maximal in terms of randomness.

Of course, assigning an amount of certainty to a process may be subject to change over time. From the epistemic interpretation of probability and the mathematical definition of randomness, the picture emerges of chance-like or random ‘until proven otherwise’. According to the epistemic definition of a chance process (subsection 1.3.2), it can never be definitively established that something is a chance process, since, as more outcomes become known, a pattern may emerge that allows exact prediction of the subsequent results. In the abstract of his popular article, Chaitin (1975) writes: “Although randomness can be precisely defined and can even be measured, a given number cannot be proved to be random.” Hence, the properties of being truly random, patternless, and chance-like are unprovable. In other words, this definition of randomness matches well with the epistemic approach to probability advocated in subsection 1.3.2.

Hayes (2001) paints the picture of randomness as a resource—something we cannot produce, but that has to be mined—that some day may run out. He focuses on the fact that a process that is currently regarded as random or chance-like may be found to be partially or completely predictable later on. This would imply that all changes of certainty-assessments can be represented as movements from right to left on the axis of Figure 1.2. However, movements in the opposite direction may occur, too: a sequence of outcomes or numbers with a clear pattern may be part of a longer sequence which turns out to be random after all. Hence, also certainty only has a temporary status.

### 1.3.7 Relation of probability to luck and justice

We use the phrases ‘fortunate’, ‘lucky’, and ‘good luck’ for cases in which a chance process happens to have a positive consequence for us; if the consequence is considered to be negative, we refer to it as ‘bad luck’ or use the term ‘accidental’. But how do
these concepts relate to randomness?

According to Barry (1989, p. 219): “To say that something is accidental or fortunate is normally to suggest that almost exactly the same causal sequence might have produced a much better or much worse outcome.” He also writes: “[a] close shave is lucky; the less close the shave the less we are inclined to talk of luck.” Stone (2008b, p. 36) interprets the phrase ‘almost exactly the same’ in epistemic terms: “A small change is a change that is difficult to notice, easy to miss, or at the limit unnoticeable.” The notion of indistinguishability can be modeled with relative analysis: see Chapter 3.

Apart from luck, probability is also related to justice. So far, we have used the word ‘fair’ only in the context of equiprobability. However, the word definitively has an ethical ring to it: ‘fair’ also means ‘just’. Indeed, fair chance processes are relevant for justice, as has been argued by political philosopher Peter Stone: he investigates the ‘problem of allocative justice’ (Stone, 2008a), which is a special case of what Rawls (1999) calls “distributive justice”. Stone (2008b) claims that a fair lottery is a just way—in fact, the only just way—to allocate goods when multiple individuals have equally good claims to the goods (in cases in which the goods cannot be shared or divided).

In the example considered in Stone (2008b) is that of a hospital director, faced with two equally needy and equally appropriate candidates for an organ transplant, but only one organ is available. The director has to decide who will get the organ and he has to do so soon. Tossing a coin to decide the matter is supposed to provide equal probabilities and is easy and quick. Stone argues that it is also a just way to allocate the organ. Arguments to use fair chance processes in such cases have been provided earlier by Katz (1973) and Kilner (1981).

In both examples, the underlying problem is the same: ranking of the possible recipients based on relevant criteria results in a partial order at best, not a total order. Even after giving relative weights to the criteria—to summarize them into one number (representing the strength of the claim of each possible recipient)—ex aequo’s are possible. (See also Brüggeman et al. (2005) for the problem of ranking substances based on their physico-chemical properties.)

In such a case, which Stone calls a case with ‘indeterminacy’, allocative justice demands a fair chance process. Why? Because, in cases in which it is not possible for the possible recipients to receive an equal amount of the good, they should at least get an equal chance to receive it. But there are further restrictions: it must be possible for all the people involved (appointer and candidates) to know which process will be used to appoint the person who will receive the goods, and they should all agree that it is a fair chance process: a demand of ‘public reason’.

In order to use something as a fair chance process, Stone (2008b) claims that all the relevant information should be common knowledge among all the agents involved. Here, he uses ‘common knowledge’ in the sense of Lewis (1969, Part II.1) and Aumann (1976), which implies not only that the agents have all of the relevant information, but also that they know this of each other, that they know of each other that they know this, and so on. In practice, this means that the drawing should take place at a public meeting or in the presence of a witness who is considered to be reliable by
all parties (such as a referee or process-server).

1.4 Infinity and probability

According to Aristotle, only potential infinities are an acceptable topic of study; under his influence, the study of infinity as something actual and completed has long been taboo in Western mathematics (Rucker, 1982, Chapter 1). Aristotle’s position still resonated in the 19th century, when Gauss wrote in a letter to Schumacher on July 12, 1831: “[S]o protestire ich zuvorderst gegen den Gebrauch einer unendlichen Grösse als einer Vollendeten, welche in der Mathematik niemals erlaubt ist. Das Unendliche ist nur eine Façon de parler, indem man eigentlich von Grenzen spricht, denen gewisse Verhältnisse so nahe kommen als man will, während anderen ohne Einschränkung zu wachsen gestattet ist” (Gauss and Schumacher, 1860, p. 269), and in the same letter: “In der Bildersprache des Unendlichen . . . ist aber nichts Widersprechendes, wenn der endliche Mensch sich nicht vermisst, etwas Unendliches als etwas Gegebenes und von ihm mit seiner gewohnten Anschauung zu Umspannendes betrachten zu wollen” (Gauss and Schumacher, 1860, p. 271). Yet, the topic of actual infinities proved to be a resilient one, and is important in almost all branches of contemporary mathematics.

It seems obvious that there has to be some relation between finite concepts and their infinite counterparts. Humans can only experience finite stimuli and their brains and associated mental capacities are finite too, so our concept of infinity has to be derived, or idealized somehow from finite concepts. Lavine (1995) argues that ‘infinity’ is our idealization of the (finitistic) concept of ‘indefinitely large’ (related to availability); in particular, it is an idealization that removes the context-dependence of the latter. In chapter 4, we will come back to this relation between the finite and the infinite realm.

1.4.1 Measuring infinite sets

In mathematics and in the philosophy of mathematics, infinity is a central concept. Friend (2007), for instance, takes the problem of infinity as the guiding example in the philosophy of mathematics in her introduction to that field. The concept of infinitely large and infinitely small quantities has always been riddled with paradoxes. A famous example is Zeno’s paradox of (the impossibility of) motion: in order to move from one place to another, it seems like infinitely many smaller movements have to
be made in a finite time, something which we would now call a ‘supertask’.\textsuperscript{10} Not surprisingly, problems related to infinity—in particular, infinite outcome spaces—also appear in the foundations of probability theory.

Within the scope of this introduction, it is not possible to give a full overview of all mathematical and philosophical problems related to infinity, nor would that be necessary in order to prepare for the specific case of infinite outcome spaces in probability theory. With that application in mind, we give an overview of how to measure infinite sets, in particular sets of natural numbers. We start with a brief historic overview; the emphasis is on Cantor’s cardinal numbers and Benci’s numerosities, and a comparison of the two. We will see that the concept of numerosity is a more natural choice than cardinality for applications in probability theory.\textsuperscript{11}

Mancosu (2009) deals with a question—sometimes called ‘Galileo’s paradox’—concerning infinity that has been posed time and time again: how to compare the size of the whole set of natural numbers to that of an infinite yet proper subset, such as the even numbers, the square numbers, or the primes? Of course, this question will also take a central position in our discussion of a fair lottery on the natural numbers (see subsection 1.4.2.3 and Chapter 2).

1.4.1.1 Historic dispute

According to Mancosu (2009, p. 614), the Islamic mathematician Thabit ibn Qurra (ninth century A.D.) “defends an infinitistic position according to which there are infinite numbers and that an infinite can be larger than another infinite.”\textsuperscript{12} In the Greek tradition however, the existence of different sizes of infinity was found to be paradoxical; as examples, Mancosu quotes Proclus (fifth century) and Philoponus (sixth century). Mancosu refers to ‘De Luce’ written by Robert Grosseteste (at approximately 1220) as the first text in the Latin West which argues that the collection of all natural numbers is greater than the collection of the even numbers (although both are infinite). Later on, Emmanuel Maignan (1673) will argue in favor of the same position,\textsuperscript{13} as will Bernhard Bolzano in his ‘Paradoxes of the Infinite’ (1851).\textsuperscript{14} Galileo (1638) and Leibniz (1875-1890), however, side with the ancient Greeks and deny the existence of different sizes of infinite collections. Their positions are subtly different: whereas Galileo only denies the applicability of ‘equal to’, ‘greater than’

\textsuperscript{10}Supertasks—tasks consisting of infinitely many sub-tasks—are considered in the context of philosophy and computation theory (Hamkins, 2002). The word ‘supertask’ was coined by Thomson (1954–55), who provided the example now known as ‘Thomson’s lamp’.

\textsuperscript{11}We will not deal with asymptotic density here, as it will be discussed in relation to probability, in Chapter 2.

\textsuperscript{12}For further references to this section, please consult the bibliography included in Mancosu (2009).

\textsuperscript{13}“His notion of equality for infinite collections is stronger than mere one-to-one correspondence” according to Mancosu (2009, p. 623).

\textsuperscript{14}Bolzano does know that an infinite set stands in a one-to-one correspondence with proper subsets of itself, but denies that this suffices to justify the conclusion that the set and its proper subsets have an equal size (which he calls the ‘multiplicity of their members’). However, Bolzano later on regards this as a mistake, which he explains as an “unjustified inference from a finite set of numbers” (in a letter written in 1848).
or ‘less than’ to infinite quantities, Leibniz denies that a size can be attributed to an infinite collection altogether.

### 1.4.1.2 Cantor’s cardinal numbers

It was Cantor who gave the first mathematically rigorous proof that there do exist different kinds of infinity, by showing that the real numbers do not form a countable set and are thus of a larger kind of infinity than the set of natural numbers (Cantor, 1874, 1891). The observation that an infinite set can be put into one-to-one correspondence with a proper subsets of itself was turned into a definition of infinity by Dedekind, who wrote: “Ein System $S$ heißt unendlich, wenn es einem echten Teile seiner selbst ähnlich ist; im entgegengesetzten Falle heißt $S$ ein endliches System.” Dedekind (1888, § 5, item 64). Dedekind calls two (simply ordered) sets ‘ähnlich’ (similar) if there exists a one-to-one correspondence between them (that preserves the order). Thus, the above definition says that a set is infinite only if their exists a one-to-one correspondence between the set and one of its proper subsets.

Even if you cannot count to a sufficiently high number to count the objects in a given (finite) collection of objects, you can establish whether this number is smaller, equal to, or larger than the numbers in another given set, by trying to put the objects of one set in a one-to-one correspondence with those of the other set. If you succeed in making the one-to-one correspondence you still don’t know the number of objects, but you have established that both are equal! This is what Gazalé (2000, p. 9) calls ‘matching’, an activity that does not require names for numbers as does proper counting.

In Cantor’s theory of cardinality, that what Dedekind calls ‘similarity’ or that what Gazalé refers to as ‘matching’ is related to the size of infinite sets: whenever two sets can be put into one-to-one correspondence with each other, they have the same size, expressed as a cardinal number. When a finite set is a proper subset of another, the former has a smaller (and finite) cardinality (a natural number that counts its elements). However, when an infinite set is a proper subset of another, the cardinality of the former is less or equal to that of its superset. In particular, all infinite subsets of the natural numbers have the same cardinality as the full set of the natural numbers. Moreover, this is also equal to the cardinality of the set of the rational numbers. This least infinite cardinal, that expresses the cardinality of all countable sets, is written as $\aleph_0$.

Power sets introduce infinitely many infinite cardinalities: by the diagonal argument of Cantor (1891), one can show that the cardinality of the power set of a set $X$ with cardinality $x$ is equal to $2^x$, which is strictly larger than $x$. In particular, the cardinality $c$ of the continuum (i.e. the set of real numbers) is larger than the cardinality of the set of the natural (or rational) numbers, $\aleph_0$: $c = 2^{\aleph_0} > \aleph_0$. The continuum-problem is the question as to whether there exists a cardinality in between $\aleph_0$ and $c$. Cantor assumed that the answer is ‘no’ (called the ‘continuum hypothesis’) and hence denoted the cardinality of the continuum by $c = \aleph_1$. Despite considerable effort, he was not able to prove his hypothesis. Later, Gödel showed that the continuum hypothesis cannot be disproved within Zermelo-Frankel set theory with the
Axiom of Choice (ZFC), whereas Cohen showed that it cannot be proved in ZFC either.

One of the notorious paradoxes associated with the concept of cardinality is ‘Hilbert’s hotel’ (attributed to David Hilbert by Gamow, 1947, p. 17). In this hypothetical hotel, there are a denumerably infinite number of rooms, numbered by the natural numbers on the doors. It seems as if any finite or denumerably infinite number of additional guests can be accommodated at all times—even when the hotel is fully booked—by cleverly instructing the guests who had already checked-in to move to a room with a higher room number. (Although one may doubt whether many guests would come to a hotel with such a bad service!) There exist many similar paradoxes, such as Craig’s library, the Al-Ghazali’s problem, Shandy’s autobiography, and counting from infinity (to zero) (see e.g. Oppy, 2006, p. 8–10).

In my opinion, these puzzles do not show anything paradoxical about cardinals at all: they simply show that ‘countably infinite’ is a property that does not behave like a number. What the puzzles suggest is that there may be further distinctions to be made among countably infinite sets, a distinction that cardinals simply do not make. Despite Gödel’s claim that Cantor’s way of assigning sizes to infinite sets was inevitable (Mancosu, 2009), there is a way to make these distinctions: with numerosities. They are the topic of the next subsection.

1.4.1.3 Numerosities

For finite sets, there are two properties that hold for their size (number of elements): (a) if a set is a subset of another set, the former has a smaller size if and only if it is a proper subset (referred to as ‘Hume’s principle’), and (b) two sets have an equal size if and only if one-to-one correspondence exists between them (referred to as ‘Euclid’s principle’) (Mancosu, 2009).

To determine the size of infinite sets, we cannot use the usual counting function for finite sets. We have to extend the notion of size in some way. It seems natural to attempt to do this in such a way as to respect principles (a) and (b). However, it turns out that the combination of the principles is inconsistent in the case of infinite sets. Hence, one has to chose between them. Clearly, Cantor’s cardinality approach is based on principle (b), which expresses the intuition that the size of a set should not depend on the labeling of its elements. However, it violates another intuition, namely that the whole is always larger than the part.

Only recently, Benci and Di Nasso (2003b) have developed a way of measuring infinite sets such that principle (a) holds, but (b) is necessarily violated; they call their measure of the size of finite and infinite sets ‘numerosity’. The numerosity-approach is closely related to non-standard analysis (NSA). In alpha-theory (Benci and Di Nasso, 2003a), NSA is developed from the idea of adding a new ideal number, α, to the set of natural numbers. This α can be interpreted as the numerosity of the set N (Benci and Di Nasso, 2003a, p. 357): we will take this concept as the starting point for a uniform probability measure on N (Chapter 2). Mancosu (2009) places

\footnote{According to (Mancosu, 2009, p. 628-630), a similar idea was developed by Fred M. Katz in his 1981 dissertation “Sets and Sizes” written at MIT.}
the concept of numerosity in a long tradition of “thinkers who argued in favor of the assignment of different sizes to infinite collections of natural numbers”.

Although Descartes (1644) would not bother to reply to those who ask if the infinite number is even or odd, a question which can indeed not be answered in terms of cardinality, the question is relevant in the context of numerosities. As has been pointed out by Benci and Di Nasso (2003b) and Mancosu (2009), the values of the numerosity of the subset of even natural numbers and that of the odd natural numbers depend on the choice of the value of $\alpha$ (which depends on the model, which can be stated, for instance, in terms of a free ultrafilter). For a probability function based on numerosities, considered in Chapter 2, it will turn out that this issue makes a difference by an infinitesimal amount.

Whereas Benci and Di Nasso (2003b) only considered the numerosity of denumerable sets, the numerosity of non-denumerable sets is discussed in (Benci et al., 2006b, Di Nasso and Forti, 2010).

### 1.4.1.4 Cardinality versus numerosity

At this point, we have available two ways of measuring infinite sets: with Cantor’s cardinalities and Benci’s numerosities. These methods are related, but do not always provide the same answer to the question ‘Are these two sets equal in size?’

Table 1.1 gives an overview of the properties of the two approaches to measure infinite sets. For instance, if the numerosity of two sets is the same, this guarantees that their cardinality is equal, too, but the converse does not hold: if two sets have the same cardinality, they do not necessarily have the same numerosity. Because numerosities are a particular count of hyperreal numbers, they inherit the rich algebra of non-standard analysis. In particular, the reciprocal value (inverse) of an infinite numerosity is an infinitesimal number: we will employ this property in our probability function for an infinite lottery in Chapter 2.

Unlike cardinality, numerosity does not allow relabeling. Hence, there are no counterintuitive conclusions to be drawn from Hilbert’s hotel or similar puzzles: if you express the number of rooms with the appropriate numerosity, it is clear that there is no way to accommodate any additional guests once the hotel is full.

### 1.4.2 Infinite sample spaces

Classical probability theory is based on the axioms of Kolmogorov (1933) and is considered to be a special case of measure theory. First we will review the axioms and rules of the orthodox axiomatization. Then we will comment on the restrictions it poses in cases with infinite sample spaces.

---

16Descartes (1644, Pars prima:XXVI): “Non igitur respondere curabimus iis, qui quærunt, an si daretur linea infinita, ejus media pars esset etiam infinita; vel an numerus infinitus sit par anve impar, & talia; quia de iis nulli videntur debere cogitare, nisi qui mentem suam infinitam esse arbitrantur.”
Table 1.1: Overview of two mathematical approaches to measure infinite sets.

<table>
<thead>
<tr>
<th>Cantor’s cardinalities</th>
<th>Benci’s numerosities</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Satisfy Hume’s principle:</strong></td>
<td><strong>Fail Hume’s principle:</strong></td>
</tr>
<tr>
<td>One-to-one correspondence</td>
<td>One-to-one correspondence</td>
</tr>
<tr>
<td>⇔ same cardinality</td>
<td>⇐ same numerosity</td>
</tr>
<tr>
<td><strong>Fail Euclid’s principle:</strong></td>
<td><strong>Satisfy Euclid’s principle:</strong></td>
</tr>
<tr>
<td>Proper subset</td>
<td>Proper subset</td>
</tr>
<tr>
<td>⇒ strictly smaller cardinality</td>
<td>⇐ strictly smaller numerosity</td>
</tr>
<tr>
<td>than whole set</td>
<td>than whole set</td>
</tr>
<tr>
<td><strong>Example:</strong> Even numbers have</td>
<td><strong>Example:</strong> Even numbers have</td>
</tr>
<tr>
<td>same cardinality as ( \mathbb{N} )</td>
<td>smaller numerosity than ( \mathbb{N} )</td>
</tr>
<tr>
<td>Correspond with counting</td>
<td>Correspond with counting</td>
</tr>
<tr>
<td>measure for finite sets</td>
<td>measure for finite sets</td>
</tr>
<tr>
<td>Poor algebra; in particular, do</td>
<td>Good algebra; in particular, do</td>
</tr>
<tr>
<td>not have an inverse:</td>
<td>have an inverse:</td>
</tr>
<tr>
<td>Normalization not possible</td>
<td>Normalization possible</td>
</tr>
<tr>
<td>⇒ No basis for a probability</td>
<td>⇒ Basis for probability measure</td>
</tr>
<tr>
<td>measure</td>
<td>with infinitesimals</td>
</tr>
</tbody>
</table>

1.4.2.1 Kolmogorov’s axioms

Here, we present axioms that are equivalent to the original axiomatization of Kolmogorov (1933). In particular, K4 is not Kolmogorov’s Continuity Axiom, but rather (an equivalent formulation of) the property of Countable Additivity, which follows from the Continuity Axiom and Finite Additivity.

(K0) **Domain and range.** The events are the elements of a \( \sigma \)-algebra \( \mathcal{A} \subseteq \mathcal{P}(\Omega) \) and the probability function takes the following form:

\[
P : \mathcal{A} \to \mathbb{R}
\]

(K1) **Positivity.** \( \forall A \in \mathcal{A}, \]

\[
P(A) \geq 0
\]

(K2) **Normalization.**

\[
P(\Omega) = 1
\]

(K3) **Finite additivity.** \( \forall A, B \in \mathcal{A}, \]

\[
A \cap B = \emptyset \implies P(A \cup B) = P(A) + P(B)
\]
1.4. Infinity and probability

(K4) **Countable additivity.** Let

\[ A = \bigcup_{j=0}^{\infty} A_j \]

with \((\forall j \in \mathbb{N}) A_j \subseteq A_{j+1}\); then

\[ P(A) = \sup_{j \in \mathbb{N}} P(A_j) \]

Furthermore, we may split the axiom (K0) into two further parts:

(K0A) **Domain.** The domain of \(P\) is a \(\sigma\)-algebra \(\mathcal{A} \subseteq \mathcal{P}(\Omega)\).

(K0B) **Range.** The range of \(P\) is (a subset of) \(\mathbb{R}\).

From the combination of axioms (K0B) and (K1), we see that the range of \(P\) is \(\mathbb{R}^+\). This set provides a structure that allows for addition and multiplication of probability values. When we also take into account the Normalization axiom (K2), we obtain that:

\[ P : \mathcal{A} \to [0, 1]_{\mathbb{R}} \]

where \([0, 1]_{\mathbb{R}}\) is the unit interval in \(\mathbb{R}\).

We also mention two important definitions which are not axioms. The first one is the product formula for independent events. The second one is the definition of conditional probability, where \(P(A|B)\) is read as ‘The probability of \(A\) under the condition that \(B\)’.

(D1) **Independent events.** If \(A\) and \(B\) are events, we say that \(A\) and \(B\) are independent if and only if

\[ P(A \cap B) = P(A) \times P(B) \]

(D2) **Conditional probability.** If \(A\) and \(B\) are events such that \(P(B) \neq 0\),

\[ P(A|B) \equiv \frac{P(A \cap B)}{P(B)} \]

1.4.2.2 Various approaches to probability give rise to problems related to infinite sample spaces

In the classical interpretation of probability, as well as in the later frequentistic interpretation, and even in Kolmogorov’s axiomatization, problems occur related to infinite sample spaces.

In the classical approach to probability of Laplace (1814) and others, the problem related to infinite sample spaces occurs in the context of the ‘Principle of Indifference’ (PI) (see subsections 1.3.1 and 1.3.3.2). Recall that PI states that whenever there is no information available to choose one possibility over another, an equal probability should be assigned to those possibilities. The principle can only be applied when there
is a finite number of possible outcomes. As a consequence, all probabilities based on it are rational numbers, no irrational numbers. However, \( \pi \) can be adapted as to be applicable to situations with countably infinitely many possible outcomes: this is the principle of maximal entropy, known from information theory (Jaynes, 1957).

In the frequentistic approach to probability (cf. subsection 1.3.1), probabilities are treated as relative frequencies. Instead of considering actually observed frequencies, which necessarily consider a finite set of outcomes, the approach was generalized: probabilities were regarded as limiting relative frequencies. However, limits of relative frequencies are not countably additive, as de Finetti (1974) noticed, and thus do not conform to Kolmogorov’s axiom (K4).

Of Kolmogorov’s axioms, (K4) is the only one that is specific for infinite sample spaces. Strangely enough, this axiom is not neutral with respect to the kind of situations it can describe: some problems cannot be described within Kolmogorov’s system. Thus, the classical theory may be very well suited to study certain problems, but may be too restrictive or too tolerant to be useful for others. We are free to apply different mathematical structures depending on the problem we are interested in. As de Finetti (1974) remarked, Kolmogorov’s theory may assign probability zero to possible outcomes and this framework does not allow an adequate description of a fair, countably infinite lottery, such as a lottery on the natural numbers: see also subsection 1.4.2.3.

The solution forwarded by de Finetti (1974) himself was to adapt one of the axioms of Kolmogorov: instead of the sigma-additivity or countable additivity (CA) of (K4), he advocated the weaker restriction of finite additivity (FA). However, Kadane et al. (1986) showed that the introduction of FA implies some unexpected statistical consequences.

Classical measure theory is built on classical analysis (calculus) of the real numbers. It is in axiom (K4) that the classical limit is explicitly incorporated in probability theory. A different type of analysis has been developed by Robinson (1966): his non-standard analysis makes use of the standard real numbers, as well as new, infinitely large and infinitely small (or infinitesimal) numbers. When we have a measure available that allows us to assign infinitesimal probabilities to an infinite number of possibilities, then they may add up to a non-infinitesimal value. Thus, non-standard measure theory may be a useful framework to solve the problem of the infinite lottery. This idea does require a precise approach: we should check whether the original problem has been solved and whether no other—possibly worse—problems have been introduced. We will follow this approach in Chapter 2.

1.4.2.3 Countably infinite sample spaces

De Finetti (1974) remarked that a fair lottery on a countably infinite sample space, such as the natural numbers, cannot be described within Kolmogorov’s axiomatization of probability theory. Here, we introduce the problem. Although the problem of the ‘infinite lottery of de Finetti’ is now known for more than forty-five years and appears to be quite straightforward, it is still a topic of discussion (Kelly, 1996, Williamson, 1999, Bartha and Johns, 2001, Bartha, 2004, Burock, 2006).
The problem of the infinite lottery arises due to the fact that the classical axiomatization of probability theory (including countable additivity) does not allow assignment of a homogeneous probability distribution on the natural numbers, or any other countably infinite outcome space. Suppose that one wants to model a process in which a random number is drawn from the natural numbers. If one assigns the same non-zero weight to every possible outcome, these weights add up to infinity and cannot be normalized, as is required by the normalization-axiom. The only option that avoids the divergence of the sum is to assign zero to each outcome. However, this implies that the total sum is zero as well, although we know that the probability of the full set is unity: the probability of infinite sets cannot be found by taking an infinite sum over the probability of finite sets. In other words, countable additivity fails.

The only way to satisfy normalization and countable additivity simultaneously is to assign unequal probabilities (in such a way that smaller numbers get a larger probability), but this is not the problem we set out to model: a lottery in which different tickets have different probabilities is not fair. So it seems that we have three options: drop the requirement of normalization, drop the requirement of countable additivity, or deny that an infinite lottery can be fair. The option of non-normalizing probabilities has been investigated by Rényi (1955), whereas the solution of dropping countable additivity was advocated by de Finetti (1974), who claims that the sum-rule only holds for finite sums (finite additivity), not countably infinite ones. The option to deny the existence of a fair infinite lottery has very strange consequences. As remarked by Kelly (1996), this would imply that when one wants to test a universal hypothesis by repeated experiments, one would—in the case in which the hypothesis is false—encounter a counterexample sooner rather than later. In Chapter 2, we will deal with the infinite lottery problem using infinitesimal probabilities.

De Finetti’s infinite lottery is not the only example of a problem related to a countably infinite sample space. Leslie (1998) describes the Doomsday argument and relates it to a new problem: the shooting-room (Eckhardt, 1997). The shooting-room is a thought experiment in which a group of people is summoned, after which two dice are rolled and the people are killed if it is a double-six. Their chance of surviving appears to be equal to $\frac{35}{36}$. Yet, a different analysis shows that 90% of the people who are summoned will die, because at each call ten times more people are summoned compared to the previous call (just until the first occurrence of a double-six). Bartha and Hitchcock (1999) analyzed this new paradox using non-standard analysis.

Elsewhere, Bartha (2004) also discusses the relabeling-paradox, attributed to Norton. Also this paradox gives us more insight about countably infinite sets and the associated probabilities. The conclusion is that the relabeling of possible outcomes, which is unproblematic in the finite case, is not permissible in case of countably infinite sets. In Chapter 2, we shall see that relabeling is indeed impermissible for

---

17The Doomsday argument—in particular in the version due to Carter (1983)—claims that it is very likely (95% certain) that the extinction of the human race is near, and that we are among the last 95% of all individuals ever to be born. One of the key assumptions is that the individual you happen to be in the human population can be regarded as a fair lottery.
probability functions based on infinitesimals.

1.4.2.4 Infinite sample spaces and the additivity of probability values

De Finetti (1974, p. 116–128) formulates a number of ‘critical questions’ concerning zero probabilities; question III (p. 117) reads: “can a union of events with zero probabilities have a positive probability (in particular, can it be the certain event)?” His own response is this: “Question (III), which evidently requires to be put in the context of infinite partitions, might lead one to think and state that one can only have possible events with zero probability if they belong to infinite partitions (!). This is monstrous.” (Italics as in original, p. 117). Further on, de Finetti (1974, p. 118) states: “we can pose question (III) once again by asking whether in an infinite partition one can attribute zero probability to all the events. In this form, the question becomes essentially equivalent to that concerning the different types of additivity: finite, only for a finite sum; countable, for the denumerable case; perfect, if the additivity always holds.” De Finetti (1974) elaborates: if one answers his question III with ‘yes’, ‘no’, or ‘it depends’, this corresponds respectively to the assumption that probability is finitely, perfectly, or countably additive. The answer of de Finetti to his own question III is ‘yes’, which means that he opts for finite additivity.

Intuitively, one could expect probabilities to exhibit perfect rather than countable additivity. However, this is clearly not possible with real-valued probability functions. Even the weaker requirement of countable additivity may be problematic, as we have seen in the example of the infinite lottery. Yet, the property of perfect additivity may be attainable by non-Archimedean probabilities.

1.4.3 Implications for cases with finite sample spaces

It should be noted that even finite lotteries are not without pitfalls. In this case, the mathematical part of assigning probabilities is a trivial task, but the description in terms of rational belief is not (yet) well established. In Chapter 4, it will be argued that there is a close relationship between probabilistic problems with infinite sample spaces, and cases involving large but finite sample spaces.

Chapter 3 deals with the Lottery Paradox, originally discussed by Kyburg (1961). When one ticket will be drawn from a large but finite number of tickets, it may initially seem reasonable to believe of any given ticket that it will not win. Because this reasoning can be made for all tickets, it seems to lead to the conclusion that it is also reasonable to believe that none of the tickets will win. However, this is in clear contradiction with the fact that one ticket will win. Like the infinite lottery puzzle, this Lottery Paradox is also still debated in the philosophical literature. Douven and Williamson (2006), for instance, remarked that a formal analysis of the Lottery Paradox goes hand in hand with a formal analysis of what is ‘reasonable to believe’. Douven (2008) used the problem in relation to the even more general

\textsuperscript{18} Oppy (2006, Ch. 6) calls the latter option ‘uncountable additivity’.
question regarding our epistemic goal and notices that a solution to the paradox at least provides a first step towards a theory of justification of knowledge.

A second related problem is the Preface Paradox, originally published by Makinson (1965). If one assigns a high probability to every statement in a book, it may nevertheless seem reasonable to assume that their conjunction (i.e. the book as a whole) has a very low probability. Indeed, it is not uncommon to find a statement in the preface to a non-fiction book which indicates that the author finds it highly unlikely that there are no errors in the book, because there are so many individual statements. The beliefs in the individual statement and the disbelief in their conjunction are “logically incompatible beliefs”. Makinson argues that it is rational to believe the individual statements as well as the negation of their conjunction, even though they form an inconsistent set. The Preface Paradox seems to be closely related to the Lottery Paradox: it deals with a lottery of sorts on the individual statements (tickets) which all have a small but non-zero probability of being wrong (‘winning’). Unlike the Lottery Paradox, the Preface Paradox does not deal with clear objective probabilities, but only with (rational) beliefs. A common reaction to this paradox is to dismiss the Conjunction Principle, or at least to adapt it (see for instance Douven and Uffink, 2003).
Chapter 1. Introduction