Philosophy of Probability

Foundations, Epistemology, and Computation

Sylvia Wenmackers

Philosophy of Probability

Foundations, Epistemology, and Computation

Sylvia Wenmackers

The work described in this thesis was carried out in the Formal Epistemology Project (FEP) of the Theoretical Philosophy (TP) group, which belongs to the Faculty of Philosophy of the University of Groningen.

Cover: *'Zero probability reveals its infinite depths.'* The cover image shows '$P = 0$', in a composition made of dice—a material representation of the abstract concept of chance. The rippled edge of the right-hand picture refers to hydromancy: a method of divination by means of water. In modern times, we still feel a need for predicting the future, but we rely on science to inform of us of the relevant probabilities. This method is riddled with mysteries of its own: the interpretation of events with zero probability is notoriously complicated. To resolve these difficulties, we need a concept of probability that is more fine-grained than the real numbers.

RIJKSUNIVERSITEIT GRONINGEN

Philosophy of Probability

Foundations, Epistemology, and Computation

Proefschrift

ter verkrijging van het doctoraat in de
Wijsbegeerte
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. E. Sterken,
in het openbaar te verdedigen op
maandag 2 mei 2011
om 16.15 uur

door

Sylvia Wenmackers

geboren op 12 mei 1980
te Maasmechelen, België

Promotor: Prof. dr. mr. I. E. J. Douven

Beoordelingscommissie: Prof. dr. L. Horsten
Prof. dr. J. Peijnenburg
Prof. dr. T. Williamson

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

"Shape Of My Heart"—Sting

Contents

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 con-*
> *sidération 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
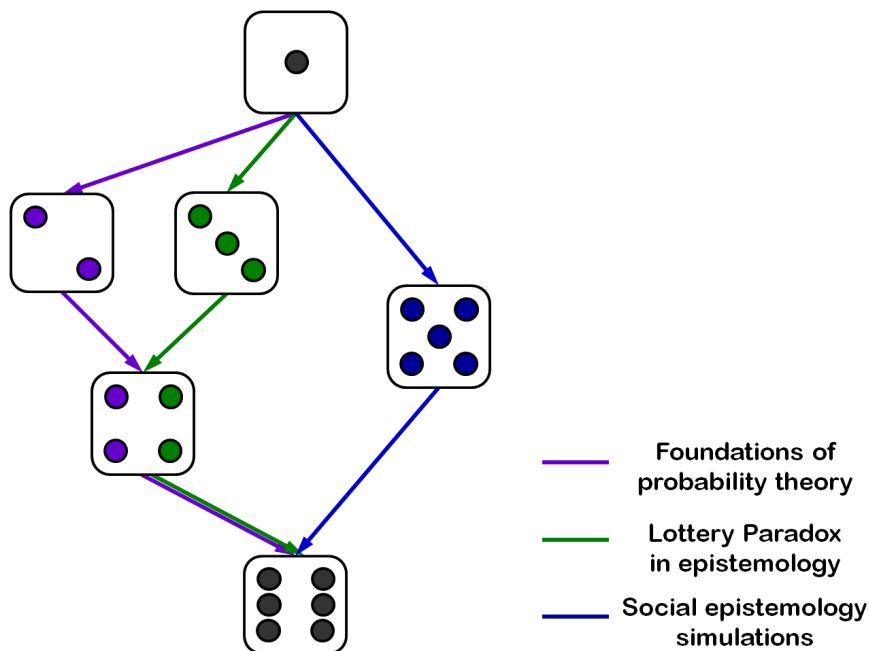
Figure 1.1: *Three reading tracks for this thesis. (The number of eyes on each die indicates the chapter number.)*

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.

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.

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.[1]

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 com-
putational 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 program-
mers 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 the-
ory, 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 interac-
tions (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 founda-
tions 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

---

[2]For 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!'.[3]

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'[4] (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

---

[3]The acronym is attributed to an engineer, Kelly Johnson.
[4]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[6]: 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

---

[6]This is similar to the definition which Stone (2008b, p. 7) gives for a lottery.

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 alternative interpretations of the theory in which no collapse occurs, such as the very elegant multiverse-interpretation of Everett (1957). According to the latter interpretation, 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 interpretation 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 multiverse. 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 mechanistic interpretation of probabilities that does not rely on the use of counterfactuals.

Taking an epistemic approach to possible outcomes—*i.e.* relating possible out-

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.[8] 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 $C_{42}^6 = \frac{42!}{(42-6)! \times 6!} = 5\ 245\ 786$. 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.[9] 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 évènements 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^{ω}, 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 Niés (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.
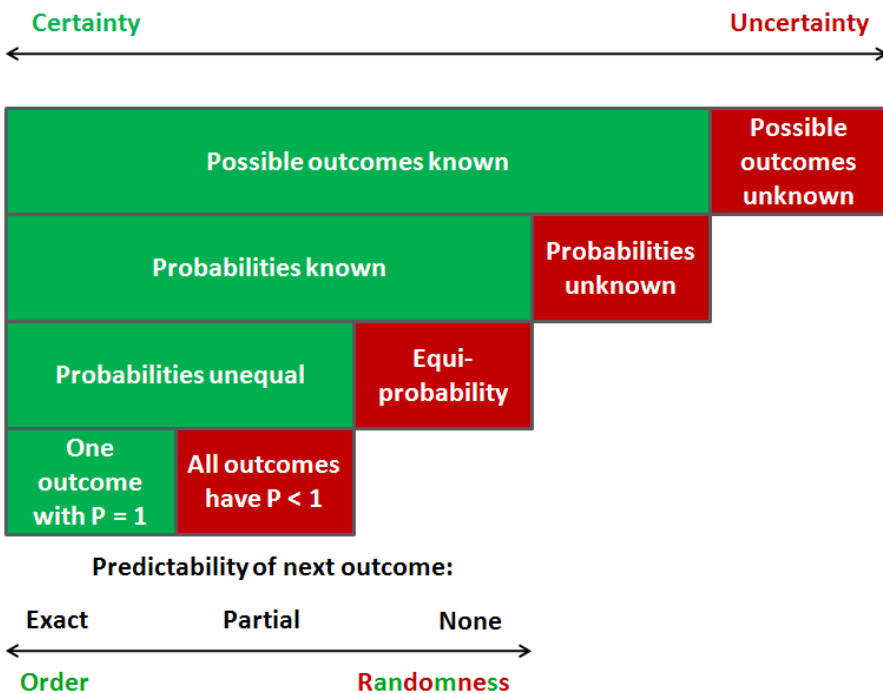
Figure 1.2: *Schematic representation of different situations involving more or less uncertainty. The situation in which all possible outcomes have an equal probability is an intermediate case in terms of (un)certainty, although it is maximal in terms of randomness.*

Chance processes with unequal probabilities, such as a weighted lottery, are po-

sitioned 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).

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

> *Das Unendliche hat wie keine andere Frage von jeher so tief das* Gemüt *der Menschen bewegt; das Unendliche hat wie kaum eine andere* Idee *auf den Verstand so anregend und fruchtbar gewirkt; das Unendliche ist aber auch wie kein anderer* Begriff *so der* Aufklärung *bedürftig.*
>
> David Hilbert (1926, p. 163)

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 zuvörderst 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'.[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.[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."[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,[13] as will Bernhard Bolzano in his 'Paradoxes of the Infinite' (1851).[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'

---

[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'.

[11]We will not deal with asymptotic density here, as it will be discussed in relation to probability, in Chapter 2.

[12]For further references to this section, please consult the bibliography included in Mancosu (2009).

[13]"His notion of equality for infinite collections is stronger than mere one-to-one correspondence" according to Mancosu (2009, p. 623).

[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'.[15] 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 $\mathbb{N}$ (Benci and Di Nasso, 2003a, p. 357): we will take this concept as the starting point for a uniform probability measure on $\mathbb{N}$ (Chapter 2). Mancosu (2009) places

---

[15]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,[16] 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 α (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.

---

[16]Descartes (1644, Pars prima:XXVI): "Non igitur respondere curabimus iis, qui quaerunt, 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.*

Cantor's cardinalities	Benci's numerosities
Satisfy Hume's principle: One-to-one correspondence $\Leftrightarrow$ same cardinality	**Fail Hume's principle:** One-to-one correspondence $\not\Rightarrow$ same numerosity
Fail Euclid's principle: Proper subset $\not\Rightarrow$ strictly smaller cardinality than whole set	**Satisfy Euclid's principle:** Proper subset $\Leftrightarrow$ strictly smaller numerosity than whole set
Example: Even numbers have same cardinality as $\mathbb{N}$	**Example:** Even numbers have smaller numerosity than $\mathbb{N}$
Correspond with counting measure for finite sets	Correspond with counting measure for finite sets
Poor algebra; in particular, do not have an inverse: Normalization not possible	Good algebra; in particular, do have an inverse: Normalization possible
$\Rightarrow$ No basis for a probability measure	$\Rightarrow$ Basis for probability measure with infinitesimals

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 σ-algebra $\mathfrak{A} \subseteq \mathcal{P}(\Omega)$ and the probability function takes the following form:

$$P : \mathfrak{A} \to \mathbb{R}$$

(K1) **Positivity.** $\forall A \in \mathfrak{A}$,

$$P(A) \geq 0$$

(K2) **Normalization.**

$$P(\Omega) = 1$$

(K3) **Finite additivity.** $\forall A, B \in \mathfrak{A}$,

$$A \cap B = \varnothing \implies P(A \cup B) = P(A) + P(B)$$

(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 σ-algebra $\mathfrak{A} \subseteq \mathcal{P}(\Omega)$.

(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:

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

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[17] 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

---

[17]The 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."[18] 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

---

[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 🎲

Fair Infinite Lotteries

> *Any finite number divided by infinity is as near to nothing as makes
> no odds, so the average population of all the planets in the Universe
> can be said to be zero. From this it follows that the population of
> the whole Universe is also zero, and that any people you meet from
> time to time are merely the products of a deranged imagination.*
>
> Douglas Adams (1980, Ch. 19)

> *Only God can make random selections.*
>
> Marion J. Levy (1981, Ninth Law)

2.1 Introduction: from the finite to the transfinite

In this chapter, we deal with the foundations of probability theory inspired by
the case of a fair lottery on the natural numbers. Although such a lottery may not
objectively exist, we have strong intuitions about it. We may even give quantitative
answers to questions such as: "What is the probability that the winning number is
odd?" We may answer "Fifty percent", by considering the probability of even and
odd numbers in finite lotteries.

When forming an image of infinite mathematical objects, we rely on our experience
with finite objects (Lavine, 1995). More often than not it is impossible to construct
or discover an infinite counterpart of a finite concept that fulfils *all* our intuitions
concerning the former. In such cases we have to choose which of those intuitions
is most dear to us and weaken or give up at least one other. An example from
classical mathematics would be the assignment of cardinalities to infinite sets by
Cantor: he took the existence of a one-to-one correspondence between sets as the
guiding principle for assigning equal sizes to them, but had to give up the intuition
that the whole is always larger than the parts. If the infinite was like the finite in

every relevant respect, then it would not be so interesting. Giving up some of our intuitions and tacit assumptions is just the price we have to pay if we want to study a new object. So it is with infinite lotteries. We will have to give up some of our intuitions governing finite lotteries. The question is: which ones, and to what extent.

In this chapter, we consider a fair lottery in which exactly one winner is randomly selected from a countably (or denumerably) infinite set of tickets. We intend to find a description of such a lottery that is mathematically sound and philosophically adequate, by examining our intuitions governing finite lotteries. It may be objected that since there are no infinite lotteries in reality, it is not clear how we can have any intuitions about the concept. Real world lotteries are always finite, but —no matter how the drawing is realized— the outcome can never be guaranteed to be random. Therefore even the idea of a truly fair n-ticket lottery, where n is some finite natural number, is a highly idealized concept, but one that is useful in analyzing a broad range of practical situations. Allowing the lottery to have an infinite number of possible outcomes is an additional idealization. The idea of a fair lottery on the natural numbers occurs in probabilistic number theory (Tenenbaum, 1995) and may be a useful approximation for large lottery-like phenomena. This problem also goes by the name of 'de Finetti's lottery' (Bartha, 2004) and 'God's lottery' (McCall and Armstrong, 1989). Although we will never be confronted with a lottery consisting of an infinite set of tickets in reality, it is valid to ask what probability can rationally be assigned to a ticket.

In the subjectivistic approach to probability, it has been argued (for instance by Ramsey, 1931, de Finetti, 1974) that our subjective probability assignments can only be rational if they agree with Kolmogorov's laws of probability (Kolmogorov, 1933). Within Kolmogorov's axiomatization however, there simply is no description available for a fair countably infinite lottery. To describe this case, we have to formulate new axioms or at least change one of the assumptions or axioms of Kolmogorov's system. One solution, advocated by de Finetti (1974), is to relax the requirement of countable additivity to finite additivity. In this chapter, we will develop a different approach. We propose to replace the co-domain (or range) of the probability measure by a non-standard set: this allows us to assign a non-zero, infinitesimal probability to single tickets. In contrast to de Finetti's solution, there will be a sense in which the probability of a countably infinite union of events supervenes on the probabilities of the individual events. This sense is captured by an additivity principle that is a close analogue of the usual assumption of countable additivity.

The chapter is structured as follows. In section 2.2, we examine our intuitions concerning finite and infinite lotteries. Inn section 2.3 we review asymptotic density and a generalization thereof by means of which a finitely additive, real-valued probability measure can be obtained on the full power set of the natural numbers. In section 2.4, we introduce some central concepts of non-standard analysis. In section 2.5, we construct a hyperrational-valued probability measure (based on the concept of numerosity) and show that it is hypercountably additive rather than countably additive. In section 2.6, we compare our hyperrational approach with the real-valued solution based on (generalized) asymptotic density, and with a hyperfinite lottery.

In the 2.7th and last section, we review the most salient features of the proposed hyperrational description of a fair, countably infinite lottery.

Regarding notation, throughout this chapter we take $\mathbb{N}$ to be the set of strictly positive integers, $\{1, 2, 3, \ldots\}$. *Even* will denote the set of even natural numbers and *Odd* that of the odd natural numbers. We use $\langle \ \rangle$-brackets to indicate an ω-sequence; if only one element is given between the brackets, it will be the general element at a position $n \in \mathbb{N}$. We abbreviate 'non-standard analysis' as NSA, 'finite additivity' and 'finitely additive' with FA, and 'countable additivity' and 'countably additive' with CA. The 'H' in HFA and HCA adds the prefix 'hyper-' to the former abbreviations.

2.2 Intuitions concerning lotteries

First we need to be precise about what we mean exactly with a finite lottery, give a mathematical description of it, and make our intuitions about it explicit. Subsequently we investigate to what extent these intuitions carry over to the infinite case, and what has to be changed in the mathematical description to maximize the intuitive appeal of it.

2.2.1 Finite lotteries

2.2.1.1 Probability measure

By a finite lottery we mean a process that assigns exactly one winner among a discrete set of tickets in a fair way. By fair we mean that each ticket initially has the same probability of winning. So this process can be modeled by a uniform, discrete function (given below in eq. 2.1), which fulfills all of Kolmogorov's axioms for probability measures (Kolmogorov, 1933).

The sample space is the set of tickets. The tickets may be numbered, but they need not be: they may be characterized by other symbols with no apparent order. Because the tickets are finite in number, say $n \in \mathbb{N}$, they can be labeled with an initial segment of the natural numbers, and this set of numbers $\{1, \ldots, n\}$ may be used as the sample space instead.

The event space is a σ-algebra[1] which contains all combinations of tickets to which we can assign a probability. Since we may do so for any possible combination of tickets, the event space is the powerset $\mathcal{P}(\{1, \ldots, n\})$ of the sample space. The probability values for an n-ticket lottery form the set $\{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1\}$. The co-domain of a probability measure on a finite lottery with an unspecified number of tickets is therefore the set of all rational numbers in the $[0, 1]$ interval: $[0, 1]_{\mathbb{Q}} = [0, 1] \cap \mathbb{Q}$. The probability measure is given by the function:

---

[1] An algebra is a family of subsets of the sample space that contains the sample space itself and is closed under complementation and finite unions; a σ-algebra is closed under denumerable unions on top of that.

$$
\begin{aligned}
P_n: \quad \mathcal{P}(\{1,\ldots,n\}) \quad &\rightarrow \quad [0,1]_{\mathbb{Q}} \\
A \quad &\mapsto \quad \tfrac{\#(A)}{n}
\end{aligned}
\tag{2.1}
$$

where $\#$ is the counting function, that maps a finite set to its number of elements (finite cardinality). Thus, P_n is a counting measure normalized by the total number of elements in the sample space. $\#$ is the prototype of a CA measure, a property that we will employ in the proof of the additivity of the non-standard measure that we will construct.

Thus, P_n is CA too, but only in a trivial sense: for each countable family of disjoint subsets $\{A_1, A_2, A_3, \ldots\}$ of the sample space, there will be a finite value $k \in \mathbb{N}$ such that for all $m \geq k$, A_m is an empty set. Hence, in the countable sum $\sum_{m \in \mathbb{N}} P_n(A_m)$, all terms with $m \geq k$ will be zero and CA reduces to FA in this case.[2]

2.2.1.2 Intuitions

Now we list our intuitions governing a finite lottery. Some of these may seem highly related, and they are —at least in the finite case—, but they need not be in the infinite case, so we name them separately:

FAIR The lottery is fair.

ALL Every ticket has a probability of winning.

SUM The probability of a combination of tickets can be found by summing the individual probabilities.

LABEL The labelling of the tickets is neutral with respect to the outcome.

The assumption FAIR embodies the thought that one ticket does not have a higher probability than any other one: a fair lottery is governed by equiprobability. This can only be implemented by the formal requirement that the associated probability function is uniform.

The assumption ALL can only be implemented by the requirement that the probability of any possible combination of tickets is defined. In other words, the probability function must be defined over the whole power set of the event space.

The assumption LABEL is motivated by the intuition that labelling is no more than a convention inspired by the need for referring to specific tickets. It is implemented by requiring of the associated probability function that it is invariant under permutations of the domain.

The assumption SUM is motivated by the intuition that the probability of a set containing the winning number supervenes on the chances of winning that accrue to the individual tickets. The usual assumption of countable additivity (CA, sometimes also called σ-additivity) is one attempt of making the intuition that is encapsulated by SUM precise. We will argue, however, that this is not the right way to do it in

---

[2]k cannot be fixed in general, but depends on the family. Although at most n sets can be non-empty, k may need to be larger than n, since a family of sets can consist of many empty sets 'in between' the non-empty ones.

this case. In other words, we will argue that the implementation of SUM is not as straightforward an affair as is commonly thought.

The constraints FAIR, ALL, and SUM jointly entail that every point event must be assigned a non-zero probability. Thus in the context of infinite lotteries the so-called principle of Regularity holds. This is not to say, however, that this principle must hold in all probabilistic scenarios.[3]

These assumptions motivate the standard account of finite lotteries, and they are jointly satisfied by the standard description given by P_n. Let us now briefly survey how these assumptions fare in the context of infinite lotteries.

2.2.2 Infinite lotteries

The infinite counterpart of a finite lottery that we are interested in here is an infinite, denumerable lottery, in particular a lottery that has $\mathbb{N}$ as its sample space.

For a start, suppose that we are very keen on the intuitions ALL and SUM, and that SUM is formally cashed out as CA. There are (uncountably) many probability distributions that satisfy these two constraints. However, it is easy to see that all of them violate FAIR. But the assumption FAIR simply is non-negotiable. The intuition of fairness is absolutely central to our concept of a lottery. Whereas real world lotteries may never be completely fair, we are considering ideal lotteries. Indeed, when one considers infinite lotteries at all, one is leaving the real world behind anyway.

The assumption LABEL might seem reasonable at first blush, but as a consequence of Cantor's theory of infinite cardinalities, it will have to be abandoned. Every infinite subset of $\mathbb{N}$ is in one-to-one onto correspondence with every other infinite subset of $\mathbb{N}$. So if we insist on invariance under permutation, then every infinite subset of $\mathbb{N}$ will receive the same probability. This immediately leads to a contradiction with the laws of probability.[4] Thus, whereas for finite sample spaces the labelling of the point events is immaterial, in the infinite case it is of the essence. Giving up LABEL admittedly gives rise to a feeling of discomfit. But if we have a naked choice between giving up some of the laws of probability on the one hand, and giving up LABEL on the other hand, then we should surely take the second option.[5]

The assumption ALL seems negotiable to some extent. For one thing, it is a well-known consequence of the Axiom of Choice that there is no probability measure on the whole of $\mathcal{P}(\mathbb{R})$ (Truss, 1997, Chapter 11). If there are no probability measures on $\mathcal{P}(\mathbb{R})$, then it should perhaps come as no surprise that there are no *satisfactory* probability measures on $\mathcal{P}(\mathbb{N})$ in the context of infinite lotteries. In any event, we will require of any solution to the infinite lottery problem that to the extent that we have strong intuitions about probabilities of a subset $A \in \mathbb{N}$, the solution takes

---

[3]Skyrms (1980) and Lewis defend Regularity in general, whereas Williamson (2007), Hájek (2010) and Easwaran (2010) argue that it cannot hold in all cases (even when infinitesimal probability values are allowed).

[4]By considering *Even* and *Odd* on the one one hand, and their union $\mathbb{N}$ on the other hand, we see that all three sets have the same measure, so FA fails.

[5]For a discussion of the role of LABEL in assigning probabilities in the context of the realism debate, see (Douven et al., 2010).

the probability of A to be defined and in agreement with our intuitions —or else the solution will have to contain a winning story about why our intuitions are mistaken.

Like ALL, the assumption SUM is not so easy to assess. It is well-known that there is no uniform probability function on $\mathcal{P}(\mathbb{N})$ that is CA (de Finetti, 1974). So if we want to preserve FAIR and ALL, and insist on SUM, then we must make it precise in a way that is different from CA.

2.3 Asymptotic density: real-valued probability with finite additivity

2.3.1 Limiting relative frequency

Although all infinite subsets of the natural numbers have the same cardinality, there are ways to discriminate the 'size' of *Even* from that of $\mathbb{N}$ for instance. In number theory, the asymptotic density ad (or natural density) of a subset A of $\mathbb{N}$ is defined as follows (*e.g.* in Tenenbaum, 1995, p. 270):

$$ad(A) = \lim_{n \to \infty} \frac{\#(A \cap \{1, \ldots, n\})}{n} \quad \text{if the limit exists} \tag{2.2}$$

Asymptotic density captures the idea that a lottery over $\mathbb{N}$ is obtained from a finite lottery (eq. 2.1) in the limit of the number of tickets, n, going to infinity. Thus, for a set A that has a defined asymptotic density $ad(A) = \lim_{n \to \infty} P_n(A)$. Because $\mathbb{Q}$ is not closed under the limit-operation, the co-domain of ad is the $[0, 1]$-interval of the real numbers, rather than $[0, 1]_{\mathbb{Q}}$ as in the finite case.

Asymptotic density gives rise to a FAIR probability assignment. However, it fails ALL: since asymptotic density is not defined for all subsets of $\mathbb{N}$[6], it cannot be introduced as a measure with $\mathcal{P}(\mathbb{N})$ as its domain (Tenenbaum, 1995). The collection of all subsets of $\mathbb{N}$ that do posses asymptotic density is not closed under intersection and union, so it does not form an algebra.

2.3.2 A generalization

It is possible to extend asymptotic density to a measure that assigns a value to all subsets of $\mathbb{N}$ (Schurz and Leitgeb, 2008). This requires a generalization of the limit-concept, that assigns a value to all bounded —convergent and non-convergent— sequences: the Hahn-Banach limit (HB-lim) is a real-valued generalization, that equals the value of the classical limit for convergent sequences.[7] For any $A \in \mathcal{P}(\mathbb{N})$ the sequence $\langle \frac{\#(A \cap \{1, \ldots, n\})}{n} \rangle$ is bounded (by 1). Therefore, the Hahn-Banach limit of this sequence is defined on all of $\mathcal{P}(\mathbb{N})$, giving rise to the probability measure P_{ad}:

---

[6]ad is defined for all finite subsets of $\mathbb{N}$, but not for all infinite subsets. An example of a set for which ad is undefined is the set of numbers whose binary notation contains an even number of digits. A second example comes for free: consider its complement, the set of numbers whose binary representation contains an odd number of digits.

[7]We will give a definition of HB-lim in the non-standard framework in paragraph 2.4.5.

$$P_{ad}: \quad \mathcal{P}(\{1,\ldots,n\}) \quad \rightarrow \quad [0,1]_{\mathbb{R}}$$
$$A \quad \mapsto \quad \text{HB-lim} \frac{\#(A\cap\{1,\ldots,n\})}{n} \quad (2.3)$$

The great selling point of this construction is that probabilities are aligned with limiting relative frequencies whenever these are defined. But again, the resulting measure is only FA (Schurz and Leitgeb, 2008). So even though ALL can be obtained, SUM does not hold.

P_{ad} can be thought of as giving precise content to a suggestion by de Finetti: by weakening the requirement for CA to FA, all other axioms of classical probability theory can be saved. This solution to the problem of the fair infinite lottery was advocated by de Finetti (1974). It allows us to assign probability values to infinite subsets of $\mathbb{N}$ that correspond well to our intuitions (such as a probability 1/2 for *Even* as well as for *Odd*). The solution does come with a major drawback, however. It amounts to giving up on the intuition that the chance of a ticket from an infinite set winning is an infinite sum over the chances of individual tickets from the set winning (SUM). As a result, regularity fails, so we also have to give up on the intuition that each ticket has a non-zero chance of winning.

2.4 Infinitesimals

With FA, we can save FAIR, ALL and a finite version of SUM. But it seems odd that the measure of singletons can be exactly 0 while the measure of their union is 1. Now the question is: can we do better? So far, we have only looked at real-valued probability functions, and we have seen that in that framework the answer is 'no'. Now we will reconsider the question. It seems that we can do better indeed, by assigning an infinitesimal probability to the singletons, rather than 0. In the finite case, the probability of a singleton is 1 divided by the number of elements in the sample space. Since our sample space is $\mathbb{N}$, with an infinite number of elements, we should assign the inverse of an infinite number to the singletons. Cantor's cardinalities are not suitable for this, since they have no inverse. Non-standard analysis provides a consistent way of working with unbounded or infinite numbers, which do have an inverse: infinitesimals. So, non-standard analysis allows us to have a different co-domain for the probability measure: $^*[0,1]_{^*\mathbb{Q}}$, which is the unit interval within $^*\mathbb{Q}$, a non-standard extension of $\mathbb{Q}$.

We will give a short overview of some essential concepts in non-standard analysis (NSA) (Robinson, 1966). We do not aim at completeness here, but restrict our attention to those ingredients that we will need in the course of this chapter. The information of this section is mainly based on (Cutland, 1983) and Benci et al. (2006a).

2.4.1 The star-map and Transfer

All approaches to NSA need a tool that maps any standard object, A (which can be a number, set, function, . . .), to its unique non-standard counterpart or

hyperextension, $^\ast A$. This function, called the star-map, should preserve a large class of properties, which is ensured by the Transfer principle. In axiomatic approaches to NSA, the Transfer principle is stipulated as an axiom; in other approaches such as the ultrafilter-construction that we will employ here, it is a theorem.

An important example is the star-map of an ω-sequence. This is a hypersequence, *i.e.* a sequence that takes values on all of $^\ast\mathbb{N}$, which is a nonstandard extension of $\mathbb{N}$. We introduce the notation $\langle\!\langle \ \rangle\!\rangle$ to distinguish a hypersequence from a standard sequence; if only one element is given between the brackets, it will be the general element at position N, with $N \in {}^\ast\mathbb{N}$. (We will encounter the star-map of a sequence of sets in eq. (2.23).)

The star-map can be obtained using free ultrafilters; this requires the introduction of equivalence classes determined by a free ultrafilter.

2.4.2   Equivalence classes based on a free ultrafilter

Consider the set of ω-sequences on a general set X (or the set $X^\mathbb{N}$ of functions $\mathbb{N} \to X$). The idea is to interpret a whole sequence as one entity, be it a non-standard one. Even if X is a set of numbers, when $X^\mathbb{N}$ fails to form a field, it does not provide a useful number system. To this end, we need to make a choice of 'which positions in the sequence matter'. For instance, a difference in only finitely many positions should not matter. Fixing a free ultrafilter on the label set, $\mathbb{N}$, is a way to settle all these choices at once.[8] A nice introduction to ultrafilters can be found in (Komjáth and Totik, 2008).

A free (or non-principal) ultrafilter, $\mathcal{U}$, on $\mathbb{N}$ is a collection of subsets of $\mathbb{N}$ ($\mathcal{U} \subset \mathcal{P}(\mathbb{N})$), which fulfills four requirements:

1. $\varnothing \notin \mathcal{U}$

2. $(\forall A, B \in \mathcal{U}) \quad A \cap B \in \mathcal{U}$

3. $(\forall A \subset \mathbb{N}) \quad A \notin \mathcal{U} \Rightarrow \mathbb{N}\backslash A \in \mathcal{U}$ (ultra)

4. $(\forall A \subset \mathbb{N}) \quad A$ is finite $\Rightarrow \mathbb{N}\backslash A \in \mathcal{U}$ (free)

Two sequences are equivalent (or equal 'almost everywhere') with respect to a free ultrafilter just if the set of labels where their terms are exactly equal is an element of the filter:

$$(\forall \langle x_n \rangle, \langle y_n \rangle \in X^\mathbb{N}) \quad \langle x_n \rangle \approx_\mathcal{U} \langle y_n \rangle \Leftrightarrow \{n \mid x_n = y_n\} \in \mathcal{U} \tag{2.4}$$

We may define the equivalence class of a sequence $\langle x_n \rangle$ modulo the just defined equivalence relation, $[\langle x_n \rangle]_\mathcal{U}$, as follows:

---

[8]In general, non-standard analysis may be developed from considering a free ultrafilter on any infinite set, but in this chapter we will always use $\mathbb{N}$ as the index set.

$$(\forall \langle x_n \rangle \in X^{\mathbb{N}}) \quad [\langle x_n \rangle]_{\mathcal{U}} = \{\langle y_n \rangle \in X^{\mathbb{N}}) \mid \langle y_n \rangle \approx_{\mathcal{U}} \langle x_n \rangle\} \tag{2.5}$$

The set of equivalence classes of sequences does provide a good basis for a number system and may be interpreted as ${}^{\star}X$, the hyperextension of X.

2.4.3 ${}^{\star}\mathbb{N}$ and ${}^{\star}\mathbb{Q}$

To illustrate how the equivalence class of sequences can provide the star-map of a set X, we consider the hyperextensions of $X = \mathbb{N}$ and $X = \mathbb{Q}$.

${}^{\star}\mathbb{N}$ is defined as the set of equivalence classes (determined by the choice of a free ultrafilter on $\mathbb{N}$) of elements of $\mathbb{N}^{\mathbb{N}}$, $[\langle m_n \rangle]_{\mathcal{U}}$. The elements of ${}^{\star}\mathbb{N}$ are called hypernatural numbers. The equivalence class of a constant sequence $[\langle m_n \rangle]$, with $m_n = m \in \mathbb{N}$ for all $n \in \mathbb{N}$, can be written in short as ${}^{\star}m$ and may be identified with the standard natural number m: this embeds $\mathbb{N}$ in ${}^{\star}\mathbb{N}$. Because of the construction, ${}^{\star}\mathbb{N}$ is called a sequential extension of $\mathbb{N}$: it consists of (equivalence classes of) ω-sequences of standard natural numbers.

For probability values, the sequential extension of $\mathbb{Q}$ seems more appropriate than ${}^{\star}\mathbb{N}$. The set of hyperrational numbers ${}^{\star}\mathbb{Q}$ can be obtained in several ways. Because the set of integers, $\mathbb{Z}$, can be introduced as the closure of $\mathbb{N}$ under substraction, and the set of rational numbers, $\mathbb{Q}$, can be introduced as the fraction field of $\mathbb{Z}$, ${}^{\star}\mathbb{N}$ can be extended similarly to ${}^{\star}\mathbb{Z}$ and ${}^{\star}\mathbb{Q}$ subsequently. Alternatively, by considering sequences of integer or rational numbers, ${}^{\star}\mathbb{Z}$ and ${}^{\star}\mathbb{Q}$ can be constructed using ultrafilters in a similar fashion as ${}^{\star}\mathbb{N}$. (As was already mentioned, the star-map returns the hyperextension of any standard object, so also ${}^{\star}\mathbb{R}$ and ${}^{\star}\mathbb{C}$ can be obtained.)

2.4.4 $\mathbb{R}$ as an approximation to ${}^{\star}\mathbb{Q}$

To relate results of NSA to standard analysis, the standard part (or shadow) function, *st*, is a useful concept: *st* maps any non-standard number to the closets real value, which is uniquely determined. Clearly, 'taking the standard part' comes down to rounding up to infinitesimals.

For fair lotteries on finite sample spaces, the probabilities are fractions (rational numbers). For infinite sample spaces, probabilities can be associated with ω-sequences of rational values. Observe that both the construction of $\mathbb{R}$ from Cauchy sequences and the above construction of ${}^{\star}\mathbb{Q}$ start from ω-sequences of rational numbers. Only the rule by which the whole sequence is associated with a new —real or hyperrational— number differs. Thus, for the range of a fair infinite lottery we seem to have a choice between $\mathbb{R}$ and ${}^{\star}\mathbb{Q}$ (rather than between $\mathbb{R}$ and ${}^{\star}\mathbb{R}$).

If we use the standard part function on ${}^{\star}\mathbb{Q}$, we can obtain any value of $\mathbb{R}$. Thus, from the non-standard viewpoint, real-valued probabilities can be seen as an approximation of hyperrational numbers. If the real values give satisfactory answers, they are all we need. If not, we may need to look at a more precise description in terms of hyperrationals.

2.4.5 Limits

2.4.5.1 Alpha-limit

We may call $^{*}a_{\alpha}$ the 'ideal value' or the 'alpha-limit' of the sequence $\langle a_n \rangle$. It is equal to the value of the hypersequence $\langle\!\langle ^{*}a_N \rangle\!\rangle$ at position $N = \alpha$ and also to the ultrafilter equivalence class of the sequence $\langle a_n \rangle$ (Benci and Di Nasso, 2003a, p. 367):

$$^{*}a_{\alpha} = [\langle a_n \rangle]_{\mathcal{U}} \tag{2.6}$$

2.4.5.2 Classical limit

The definition of the classical limit of a real-valued sequence, $\lim_{n \to \infty} a_n$, can be reformulated in NSA as follows (Väth, 2007, p. 88): $\langle a_n \rangle$ converges with limit $L \in \mathbb{R}$ if and only if

$$(\forall N \in {}^{*}\mathbb{N} \backslash \mathbb{N}) \quad {}^{*}a_N - L \text{ is infinitesimal} \tag{2.7}$$

In particular for $N = \alpha$, if $\lim_{n \to \infty} a_n = L$, then $st(^{*}a_{\alpha}) = L$. By equation (2.6), we find that:

$$\lim_{n \to \infty} a_n = st\!\left([\langle a_n \rangle]_{\mathcal{U}} \right) \quad \text{if the limit exists} \tag{2.8}$$

2.4.5.3 Hahn-Banach limit

Hahn-Banach limits (HB-lim) are a real-valued generalization of the limit that is defined for all bounded sequences and is equal to the value of the classical limit if the sequence converges. In NSA, the Hahn-Banach limit of a real-valued sequence $\langle a_n \rangle$ is defined as follows (Väth, 2007, p. 133): $\langle a_n \rangle$ is bounded with HB-lim$(a_n) = L \in \mathbb{R}$ if and only if

$$L = st\!\left(\sum_{N=H_0}^{H_1} h_N \, {}^{*}a_N \right) \tag{2.9}$$

for some $H_0, H_1 \in {}^{*}\mathbb{N} \backslash \mathbb{N}$ with $H_0 < H_1$ and for some internal sequence of hyperreals $h_{H_0}, \ldots, h_{H_1}$ such that $\sum_{N=H_0}^{H_1} h_N = 1$.

2.4.6 Internal and external objects

Not all non-standard objects are the image of some standard object by the star-map; those that are, are called 'internal', the others 'external'. This distinction is important but not easy to understand immediately. As a first example, $\mathbb{N}$ is external; in fact, every infinite countable set is external. It is also important to note that whereas the star-map preserves most set-operations, it does not do so for the powerset: the operation $^{*}\mathcal{P}$ returns only the internal subsets of a given set, and so, for any infinite standard set A: $^{*}\!\big(\mathcal{P}(A)\big) \subsetneq \mathcal{P}(^{*}A)$. The probability function we will construct will also turn out to be external.

2.5 Hyperrational valued probability

Let us now construct a non-standard valued probability measure to describe a lottery on $\mathbb{N}$. If we want to find a function that is formally similar to the probability measure of a finite lottery, then we must find a way to 'count' finite as well as infinite subsets of $\mathbb{N}$ and divide by the size it assigns to the whole sample space to normalize this function.

The construction proceeds in four steps. (1) Every subset of $\mathbb{N}$ can be represented as an infinitely long bit string by considering its characteristic function. (2) Then we consider the sequence of partial sums of these bits. (3) By introducing a free ultrafilter on $\mathbb{N}$, we can interpret the whole partial sum sequence as one nonstandard (or hypernatural) number: its numerosity. (4) By a suitable normalization, we finally obtain a hyperrational-valued probability measure on $\mathcal{P}(\mathbb{N})$.

After the construction we show that numerosities and probabilities based on them are hypercountably additive (HCA). In fact, we shall see that it is perhaps more appropriate to call them hyperfinitely additive (HFA).

2.5.1 The construction

2.5.1.1 Step 1: Characteristic bit string

First, consider the indicator function or characteristic function of a subset A of $\mathbb{N}$: it tests whether a natural number is in the set A or not, where a positive answer corresponds to 1 and a negative to 0.

$$
\begin{aligned}
\chi_A : \quad \mathbb{N} \quad &\rightarrow \quad \{0,1\} \\
n \quad &\mapsto \quad \begin{cases} 0 & \text{if } n \in \mathbb{N} \backslash A \\ 1 & \text{if } n \in A \end{cases}
\end{aligned}
\tag{2.10}
$$

Now we can construct the function that assigns a characteristic bit string (CBS) to any subset of the natural numbers:

$$
\begin{aligned}
CBS : \quad \mathcal{P}(\mathbb{N}) \quad &\rightarrow \quad \{0,1\}^{\mathbb{N}} \\
A \quad &\mapsto \quad \langle \chi_A(1), \chi_A(2), \ldots, \chi_A(n), \ldots \rangle
\end{aligned}
\tag{2.11}
$$

In shorthand notation, we can refer to a sequence by its n-th element only, so $CBS(A) = \langle \chi_A(n) \rangle$.

2.5.1.2 Step 2: Partial sums of characteristic bit strings

Now we consider the sequence of partial sums of characteristic bit strings of a subset A of natural numbers, $SumCBS(A)$.

$$
\begin{aligned}
SumCBS : \quad \mathcal{P}(\mathbb{N}) \quad &\rightarrow \quad \mathbb{N}^{\mathbb{N}} \\
A \quad &\mapsto \quad \langle S_n \rangle
\end{aligned}
\tag{2.12}
$$

with

$$
\begin{aligned}
S_n &= \textstyle\sum_{m=1}^{n} \chi_A(m) \\
&= \chi_A(1) + \chi_A(2) + \ldots + \chi_A(n)
\end{aligned} \tag{2.13}
$$

So S_n has a value in $\{0, \ldots, n\}$ for all n. Alternatively, S_n can be written in terms of the counting function $\#$:

$$
\begin{aligned}
S_n &= \sum_{m=1}^{n} \chi_A(m) \tag{2.14} \\
&= \sum_{m \in \mathbb{N}} \chi_{A \cap \{1,\ldots,n\}}(m) \tag{2.15} \\
&= \#(A \cap \{1, \ldots, n\}) \tag{2.16}
\end{aligned}
$$

The sequence $\langle S_n \rangle$ seems to 'point to a value at infinity'. If we interpret this sentence in the framework of standard analysis, we should take the limit $n \to \infty$. For infinite sets A this results in $\lim_{n \to \infty} S_n = \infty$, which cannot be normalized. (If we consider the sequence $\langle \frac{S_n}{n} \rangle$ instead and its limit $n \to \infty$, we find asymptotic density again.) As we have seen however, in NSA we may alternatively interpret the whole sequence as one non-standard number.

2.5.1.3   Step 3: Numerosity as the equivalence class of the partial sum sequence

The equivalence class under a free ultrafilter of a sequence of partial sums of characteristic bit strings is a hypernatural number that can be interpreted as the size of the corresponding set, called its *numerosity*.

$$
\begin{aligned}
num: \quad \mathcal{P}(\mathbb{N}) &\to \quad {}^*\mathbb{N} \\
A &\mapsto \quad [\langle S_n \rangle]_{\mathcal{U}}
\end{aligned} \tag{2.17}
$$

with S_n as before.

Any finite set has a finite numerosity (value in $\mathbb{N}$, equal to $\#(A)$), whereas any infinite set has an infinite numerosity (value in ${}^*\mathbb{N} \backslash \mathbb{N}$). As an example, consider $A = \mathbb{N}$. In that case, $CBS(A) = \langle 1, 1, 1, \ldots, 1, \ldots \rangle$ and $num(A) = [\langle 1, 2, 3, \ldots, n, \ldots \rangle]_{\mathcal{U}}$, which is larger than any finite number. Thus we have shown that $num(\mathbb{N})$ is an element of ${}^*\mathbb{N} \backslash \mathbb{N}$. We may call this new number, $num(\mathbb{N})$, alpha (α).

Using equation (2.14), we can transform the definition for numerosity given by equation (2.17) in the following way:

$$
\begin{aligned}
num(A) &= [\langle S_n \rangle]_{\mathcal{U}} \\
&= [\langle \#(A \cap \{1, \ldots, n\}) \rangle]_{\mathcal{U}} \tag{2.18} \\
&= [{}^*\#({}^*A \cap [\langle \{1, \ldots, n\} \rangle])]_{\mathcal{U}}
\end{aligned}
$$

for all subsets A of $\mathbb{N}$.[9] Note that $[\langle \{1, \ldots, n\} \rangle]_{\mathcal{U}}$ is a hyperfinite set, which means that there is an infinite hypernatural number, N, such that this set is equal to

---

[9]The form on the second line makes it clear that our ultrafilter-based definition of the numerosity function is equal to that of the axiomatic approach developed in (Benci and Di Nasso, 2003b). In

$\{1, \ldots, N\}$. We can even be more precise, the hypernatural number N is equal to $[\langle 1, 2, 3, \ldots, n, \ldots \rangle]_{\mathcal{U}}$, which we will call α (following the terminology of Benci and Di Nasso). So we obtain:

$$num(A) = {}^{*}\#({}^{*}A \cap \{1, \ldots, \alpha\}) \tag{2.19}$$

Remark how similar this form is to the numerator of the asymptotic density function in equation (2.2). Thus, equation (2.19) is very suggestive for a probability function: all we need to do is normalize it.

2.5.1.4 Step 4: Non-standard probability

The construction is completed by normalization of the numerosity function. By dividing it by the numerosity of the sample space, $num(\mathbb{N}) = \alpha$, we can introduce the probability function of a lottery over $\mathbb{N}$ in a form that is similar to the probability measure of a finite lottery (eq. 2.1):

$$P_{num}: \quad \begin{array}{ccc} \mathcal{P}(\mathbb{N}) & \rightarrow & {}^{*}[0,1]_{{}^{*}\mathbb{Q}} \\ A & \mapsto & \frac{num(A)}{\alpha} \end{array} \tag{2.20}$$

P_{num} takes values on the unit interval of ${}^{*}\mathbb{Q}$ and may be interpreted as a hyperrational-valued probability function. It assigns an infinitesimal probability to any finite set and a larger probability to any infinite set.

Since algebraic operations on non-standard numbers are equivalent to the termwise application of the corresponding operation on the underlying sequence, from our construction we obtain:

$$P_{num}(A) = \frac{num(A)}{\alpha} = \left[\left\langle \frac{\#(A \cap \{1, \ldots, n\})}{n} \right\rangle \right]_{\mathcal{U}} \tag{2.21}$$

which makes the analogy with asymptotic density complete.

P_{num} is an external object, because its domain is $\mathcal{P}(\mathbb{N})$, which is an external set. This means that the function cannot be obtained directly by taking a standard function and applying the Transfer Principle to it.

2.5.2 Additivity of the probability function

In this paragraph, we will investigate to what extent our proposal for P_{num} satisfies SUM. We will show that the numerosity function is hypercountably additive (HCA) rather than countably additive (CA). It then follows directly that the probability measure based on it is HCA too.

---

the existence proof given in that paper, numerosity is related to the ultrafilter-construction of NSA as follows: $num(A) = [\langle \#(A \cap \{1, \ldots, n\}) \rangle]_{\mathcal{U}}$ (Benci and Di Nasso, 2003b, p. 62), where $\mathcal{U}$ is a free ultrafilter such that $\mathcal{U} = \{A \subseteq \mathbb{N} \mid \alpha \in {}^{*}A\}$ (Benci and Di Nasso, 2003b, p. 374).

2.5.2.1   Addition on $^*\mathbb{N}$

First we need to define finite addition on $^*\mathbb{N}$. The sum of two hypernatural numbers is defined as the star-map of the standard sum operation (on standard numbers) $+$: $^*+$. For any two sequences $\langle a_n \rangle$, $\langle b_n \rangle$ we have: $[\langle a_n \rangle]_{\mathcal{U}}\ ^*+[\langle b_n \rangle]_{\mathcal{U}} = [\langle a_n \rangle + \langle b_n \rangle]_{\mathcal{U}}$. Since it will be clear from the context whether we are summing standard or non-standard numbers, we may drop the * from the sum-symbol. Likewise, all finite sums are defined.

The countably infinite sum $\sum_{n\in\mathbb{N}}$ is not defined for non-standard terms: since $\sum_{n\in\mathbb{N}} = \lim_{m\to\infty} \sum_{n=1}^{m}$ and limits are only defined for standard numbers. Another way to see this is by taking into account that $\mathbb{N}$ is an external set, and summations over such sets are not defined.

An infinite addition that is relevant (always defined) for non-standard numbers is the hyperfinite sum: the equivalence class of a sequence of finite sums. We may also consider the summation over all of $^*\mathbb{N}$ —a hypercountable sum $\sum_{N\in{^*\mathbb{N}}}$— since the latter is an internal set.

2.5.2.2   *num* is not CA

CA relates a property of the domain of a measure to a property of its range. First of all, the domain has to be a σ-algebra, ensuring that the union of any countably infinite family of sets in the domain is also in the domain. Secondly, countable sums have to be defined on the range. CA then links the two by requiring that the measure of the union of a countable family is equal to the countable sum of measures of the each of the members of the family.

The domain of the numerosity function is $\mathcal{P}(\mathbb{N})$, which is indeed a σ-algebra. The range of the function however is the set of hypernatural numbers, for which the countably infinite sum is undefined. Therefore, the numerosity function cannot be countably additive. The same argument applies to *any* function that maps $\mathcal{P}(\mathbb{N})$ to a set of non-standard values.

2.5.2.3   *num* of a sequence of sets

So far, the numerosity function is only defined for individual subsets of $\mathbb{N}$ (eq. 2.17). Now, we define the numerosity of a sequence of disjoint subsets as follows:

$$num(\langle A_n \rangle) = {^*\#}\big({^*}(\langle A_n \rangle) \cap \{1, \ldots, \alpha\}\big) \tag{2.22}$$

The goal of this paragraph is to find an equivalent form of the above definition, that gives us more insight in the additivity of this function.

First, consider a sequence of (possibly empty) subsets of $\mathbb{N}$:

$$\langle A_1, A_2, \ldots, A_n, \ldots \rangle = \langle A_n \rangle_{n\in\mathbb{N}}$$

The star-map of this sequence is a hypersequence of internal subsets of $^*\mathbb{N}$:

$$^*(\langle A_n \rangle_{n\in\mathbb{N}}) = \langle\!\langle {^*}A_N \rangle\!\rangle_{N\in{^*\mathbb{N}}} \tag{2.23}$$

with $^*A_N = A_N$ for $N \in \mathbb{N}$.

Define the intersection with and the hypercounting function of a hypersequence of subsets of $^*\mathbb{N}$ componentwise:

$$\left(\forall S \in {}^*\big(\mathcal{P}(\mathbb{N})\big)\right) \quad \langle\!\langle {}^*A_N \rangle\!\rangle \cap S = \langle\!\langle {}^*A_N \cap S \rangle\!\rangle \tag{2.24}$$

$$^*\#(\langle\!\langle {}^*A_N \rangle\!\rangle) = \langle\!\langle {}^*\#({}^*A_N) \rangle\!\rangle \tag{2.25}$$

By applying equations 2.23, 2.24 and 2.25, the definition given in equation (2.22) is transformed to:

$$num(\langle A_n \rangle) = \langle\!\langle {}^*\#({}^*A_N \cap \{1, \ldots, \alpha\}) \rangle\!\rangle \tag{2.26}$$

Thus, the numerosity of an ω-sequence of sets is a hypersequence of hypernatural numbers; we may refer to its N-th element as $\big(num(\langle A_n \rangle)\big)_N$.

2.5.2.4   *num* and P_{num} are HCA

Now we arrive at the main result of this chapter: the numerosity function is hypercountably additive. To prove this, we need to show that for any family of disjoint subsets of $\mathbb{N}$, $\{A_n \mid n \in \mathbb{N}\}$:

$$num(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{N \in {}^*\mathbb{N}} \big(num(\langle A_n \rangle)\big)_N \tag{2.27}$$

As an essential step in the proof, we will use that the hyperextension of a countable union of a countable family of sets is equal to the hypercountable union of the hyperextension of the family (by Transfer, *cf.* Rubio, 1994, p. 104):

$$^*\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \bigcup_{N \in {}^*\mathbb{N}} {}^*A_N \tag{2.28}$$

Proof.

$$num(\bigcup_{n\in\mathbb{N}} A_n) = {}^*\#\left({}^*(\bigcup_{n\in\mathbb{N}} A_n)\cap\{1,\ldots,\alpha\}\right)$$

[By definition of *num* in equation (2.19)]

$$= {}^*\#\left(\bigcup_{N\in{}^*\mathbb{N}} {}^*A_N\cap\{1,\ldots,\alpha\}\right)$$

[By equation (2.28)]

$$= {}^*\#\left(\bigcup_{N\in{}^*\mathbb{N}} ({}^*A_N\cap\{1,\ldots,\alpha\})\right)$$

[Distributivity of intersection over union + Transfer]

$$= \sum_{N\in{}^*\mathbb{N}} {}^*\#({}^*A_N\cap\{1,\ldots,\alpha\})$$

[CA of counting function + Transfer]

$$= \sum_{N\in{}^*\mathbb{N}} \big(num(\langle A_n\rangle)\big)_N$$

[By equation (2.26) + remark below it]

$\square$

Because the numerosity function is HCA, so is the probability function obtained from it (provided that we first define P_{num} of a sequence of sets, as a normalized version of eq. 2.22). In particular, for the entire sample space $\mathbb{N}$, the infinitesimal probabilities of the countably infinite family of singletons do add up to unity.

The idea behind the proof is visualized in Figure 2.1 for the specific case of the countably infinite family of singletons of $\mathbb{N}$. In this example, we consider the countably infinite family of singletons of $\mathbb{N}$, whose union is $\mathbb{N}$. The numerosity function of a single set requires the extension of sequences (*SumCBS*) into hypercountably long ones (**SumCBS*) and their validation at position alpha. To determine the numerosity of a countably infinite union by addition, the countable family of standard sets has to be extended to a hypercountable family of non-standard sets. By looking at the last horizontal line, we see that the numerosity of $\bigcup_{n\in\mathbb{N}} A_n$ (here $\bigcup_{n\in\mathbb{N}}\{n\} = \mathbb{N}$) is determined as $^*\#(^*(\bigcup_{n\in\mathbb{N}} A_n)\cap\{1,\ldots,\alpha\})$, here $^*\#(^*\mathbb{N}\cap\{1,\ldots,\alpha\}) = @.$ In this case, the values in the column with number $M = \alpha$ add up to @ too. For each case, a similar table can be made, and the proof states that the same value is always obtained from comparing the αth position of the last row with the hypercountable sum of the αth column.

It is apparent from the example in Figure 2.1 that the lottery on $\mathbb{N}$ is HCA in a very specific sense:[10] for each family, there can always be found a hypernatural number, $K \in {}^*\mathbb{N}$ (α in the example, but possibly larger in other cases), such that the hypercountable sum decomposes in a hyperfinite sum and a hypercountable tail with zero-terms only. Thus we may call *num* and P_{num} hyperfinitely additive (HFA).

---

[10]This is completely analogous to the argument, given in section 2.2.1, showing that a finite lottery is only CA in a trivial sense.

Hypersequence: $<<*\#(*A_N \cap \{1,\dots,M\})>>$

Sequence: $<\#(A_n \cap \{1,\dots,m\})>$

Family of sets of natural numbers

Hyperfamily of sets of hypernatural numbers

(A_n)	$*A_N$	(m) M	1	2	3	...	n-1	n	n+1	...	α-1	α	α+1	...
$A_1 =$	$*A_1 = \{1\}$		1	1	1	...	1	1	1	...	1	1	1	...
$A_2 =$	$*A_2 = \{2\}$		0	1	1	...	1	1	1	...	1	1	1	...
$A_3 =$	$*A_3 = \{3\}$		0	0	1	...	1	1	1	...	1	1	1	...
	⋮											⋮		
$A_{n-1} =$	$*A_{n-1} = \{n-1\}$		0	0	0	...	1	1	1	...	1	1	1	...
$A_n =$	$*A_n = \{n\}$		0	0	0	...	0	1	1	...	1	1	1	...
$A_{n+1} = *A_{n+1} = \{n+1\}$			0	0	0	...	0	0	1	...	1	1	1	...
$\bigcup_{n\in\mathbb{N}}$ $\mathbb{N}$														
	$*A_{\alpha-1} = \{\alpha-1\}$		0	0	0	...	0	0	0	...	1	1	1	...
	$*A_\alpha = \{\alpha\}$		0	0	0	...	0	0	0	...	0	1	1	...
	$*A_{\alpha+1} = \{\alpha+1\}$		0	0	0	...	0	0	0	...	0	0	1	...
	⋮											⋮		
	$\bigcup_{N\in*\mathbb{N}}$ $*\mathbb{N}$	$*\Sigma_{N\in*\mathbb{N}}$	1	2	3	...	n-1	n	n+1	...	α-1	ⓐ	α+1	

Figure 2.1: *Illustration of the proof for the hypercountable additivity of the numerosity function for the specific case of the countably infinite family of singletons of* $\mathbb{N}$. *(Further explanation can be found in the main text.)*

2.6   Discussion

In this section, first we evaluate some consequences of our use of free ultrafilters to construct generalized probability functions for infinite lotteries. Subsequently we investigate the relation of a lottery on $\mathbb{N}$ to a different type of infinite lottery: the hyperfinite case. We also compare the non-standard description of a lottery on $\mathbb{N}$ to the best available real-valued approach, that of generalized asymptotic density; we see that SUM is lost in the latter as a result of an accumulation of rounding errors.

2.6.1   Non-constructiveness

Our approach may be criticized on mildly constructivist grounds. The existence proof for free ultrafilters uses Zorn's lemma and thus depends on the acceptance of the Axiom of Choice. Thus a hyperrational-valued probability measure requires free ultrafilters, which depend on the Axiom of Choice (AC).

To this objection we may reply that the more common solution based on FA, asymptotic density (paragraph 2.3) requires an extension of the limit concept measure that involves a free ultrafilter too. Therefore, this generally accepted solution is non-constructive too, as has been pointed out by Lauwers. To those who are unwilling to accept the Axiom of Choice, no measure is available that does any justice to the intuitions underlying fair lotteries. To the rest of us who have no objection to the Axiom of Choice, the HCA, hyperrational probability function is no less acceptable than the FA real-valued one. But there is something to be said in favour of the hyperrational probability function: it gives the SUM-intuition its due.

2.6.2   Non-uniqueness

Elga (2004) has remarked that there are often too many non-standard solutions: if we can use any infinite hypernatural number to model a problem, why should we prefer one rather than another? One answer is given by alpha-theory, which develops NSA from the idea of adding a new ideal number, α, to $\mathbb{N}$. This α can be interpreted as the numerosity of the set $\mathbb{N}$ (Benci and Di Nasso, 2003a, p. 357). Since numerosity theory introduces α as the size of the natural numbers, and constructs the other hypernatural numbers around it, α has a clear interpretation that is hardwired into the theory. It gives us a point of reference among the infinite number of infinite hypernaturals.

We have started from the ultrafilter-construction of $^{*}\mathbb{N}$ and then considered the equivalence class of the sequence $[\langle 1, 2, 3, \ldots \rangle]_{\mathcal{U}}$, which we interpret as α (consistent with the axiomatic approach). So, in the present context the accusation of arbitrariness boils down to the choice of a free ultrafilter $\mathcal{U}$. A different choice of free ultrafilter produces a different value of α and hence a probability function with the same standard part but infinitesimal differences.

To come back to the *Even* versus *Odd* example: the odd tickets always have a head-start compared to the even ones, for the simple reason that 1 comes before 2. Within our framework, it should not come as a surprise that the weight of this

very first ticket may result in an infinitesimal advantage for the whole set of odd tickets. After all, our approach is based on the idea that even in an infinite lottery each ticket has a non-zero probability. For all finite cases with an odd number of tickets, $P_n(Odd) = \frac{1}{2} + \frac{1}{2n} > \frac{1}{2}$ and $P_n(Even) = \frac{1}{2} - \frac{1}{2n} < \frac{1}{2}$, whereas $P_n(Odd)$ and $P_n(Even)$ are exactly $\frac{1}{2}$ for all finite lotteries with n even. Because the probability assignment in both cases is different for all finite lotteries, it leads to two different infinite lotteries as well: one in which the odd numbers are in the ultrafilter, such that $P_{num}(Odd) = \frac{1}{2} + \frac{1}{2\alpha}$ and $P_{num}(Even) = \frac{1}{2} - \frac{1}{2\alpha}$, and one in which the even numbers are in it, such that $P_{num}(Odd) = \frac{1}{2}$ and $P_{num}(Even) = \frac{1}{2}$.

It may look as if we can favor the second solution: a good probability measure should at least respect the limiting frequencies, and they are exactly $\frac{1}{2}$ for *Even* and *Odd*. Take into account however, that limiting frequencies are real-valued, and that our two non-standard solutions have exactly the same standard part, so they are both in accordance with this $\frac{1}{2}$ & $\frac{1}{2}$ solution: limiting frequencies do not rule out one of both.

By observing that $Even = \{n \in \mathbb{N} \mid n \mod 2 = 0\}$ and $Odd = \{n \in \mathbb{N} \mid n \mod 2 = 1\}$, the above analysis generalizes: for each $m \in \mathbb{N}$, m different scenario's emerge (corresponding to the issue whether $\{n \in \mathbb{N} \mid n \mod m = k\}$ is in the ultrafilter for either $0, 1, \ldots,$ or $m - 1$).

It is possible to impose additional constraints on the ultrafilter, such that the set corresponding to $m \mod 0$ is in the ultrafilter for all $m \in \mathbb{N}$, or equivalently: such that α is a multiple of any finite number (Benci and Di Nasso, 2003b, Mancosu, 2009). However, we see at present no convincing reasons for endorsing any particular constraints of this kind. We already remarked that the infinite lottery violates LA-BEL. Here we encounter another difference with finite lotteries: the solution is not unique. The problem stated as "Find a probability measure on all of $\mathbb{N}$ that satisfies FAIR, ALL and SUM" is highly underdetermined: there are as many different ways to draw a random number from $\mathbb{N}$ in a fair way as there are free ultrafilters, and the probability function P_{num} given in equation (2.20) should be seen as a whole family of solutions, all of which are, as far as we can presently see, equally relevant.

2.6.3   Lotteries on $\mathbb{N}$ versus hyperfinite lotteries

In non-standard measure theory, the emphasis lies on internal measures: measures that can be obtained as the star-map of a standard function. Since $\mathbb{N}$ is an external set within $^*\mathbb{N}$, no internal measure is appropriate to describe the probabilities concerning a lottery over $\mathbb{N}$.[11] The probability measure we propose in this chapter is indeed an external function, but it is closely related to an internal one: the probability function of a hyperfinite lottery.

Consider the finite set $A_n = \{1, \ldots, n\}$. Then $[\langle A_n \rangle]_{\mathcal{U}} = \{1, \ldots, \alpha\}$ is a hyperfinite set: an internal set and an initial sequence of $^*\mathbb{N}$. Let us consider this hyperfinite set as the sample space of a lottery. The probability measure for this hyperfinite

---

[11]With Loeb measures it is possible to transform the non-standard co-domain to a standard one (Cutland, 1983), but this still leaves us with a non-standard domain instead of $\mathcal{P}(\mathbb{N})$.

lottery can be obtained by Transfer of the standard measure of a finite lottery, given in equation (2.1). Thus we obtain the following internal probability measure:

$$
\begin{array}{rcl}
{}^{\star}P_{\alpha}: & {}^{\star}\mathcal{P}(\{1,\ldots,\alpha\}) & \to & {}^{\star}[0,1]_{{}^{\star}\mathbb{Q}} \\
& {}^{\star}A & \mapsto & \frac{{}^{\star}\#({}^{\star}A)}{\alpha}
\end{array}
\tag{2.29}
$$

Because this probability measure is obtained as the star-map of a standard probability measure on a finite sample space, which is FA, ${}^{\star}P_{\alpha}$ is HFA by Transfer.

As such, ${}^{\star}P_{\alpha}$ cannot be regarded as a satisfactory solution to the problem of denumerably infinite lotteries. The order type of the event space will be vastly different from the order type of $\mathbb{N}$. Thus, to use terminology introduced earlier, this probability function operates at best with a relabelling of the natural numbers. We have seen in our discussion of LABEL that probability functions on infinite event spaces simply cannot be taken to be invariant under relabelling. So the "internalization" of the problem of infinite lotteries in the non-standard universe does not solve the original problem. Instead, it is a solution to a different problem.

Note that although $\mathbb{N}$ is a subset of $\{1,\ldots,\alpha\}$, it is an external one (*cf.* paragraph 2.4.6), so it is not part of the domain of ${}^{\star}P_{\alpha}$. Therefore we cannot use conditionalization on ${}^{\star}P_{\alpha}$ to define a probability measure on the sample space $\mathbb{N}$.

If not by conditionalization, what is the relation between the hyperfinite lottery on $\{1,\ldots,\alpha\}$ and the hyperrational probability measure on $\mathbb{N}$? Any subset A of $\mathbb{N}$ has a hyperextension ${}^{\star}A \in {}^{\star}\mathcal{P}({}^{\star}\mathbb{N})$; the intersection of ${}^{\star}A$ with $\{1,\ldots,\alpha\}$ is an internal subset of $\{1,\ldots,\alpha\}$ and thus can be assigned a probability value by the internal measure ${}^{\star}P_{\alpha}$. This is precisely the inner working of P_{num}: a three-step-process that can be read off from equation (2.19).[12]

2.6.4 Non-standard probability and asymptotic density

All frequencies that are experimentally accessible concern finite sequences of trials only, and they can be expressed as rational numbers. Limiting frequencies are an infinite idealization of observable frequencies resulting in real-valued probabilities. Our approach is a different idealization that leads to hyperrational probabilities. We will show that its standard part is exactly equal to the limiting frequency. Therefore it is impossible to favour one solution over the other on mathematical or experimental grounds. So we are addressing a question of epistemology: the approaches encapsulate a different vision on probability and it is up to us to decide which one is most in concord with our intuitions.

2.6.4.1 Co-domain $\mathbb{R}$ versus ${}^{\star}\mathbb{Q}$

For a finite lottery, the probability measure takes values on $\mathbb{Q}$. However, this co-domain is not closed: a limit of a rational sequence is not always defined in $\mathbb{Q}$, but it is in $\mathbb{R}$. In general, the sample space may be infinite; to this end the range is extended to $\mathbb{R}$. In contemporary presentations, probability theory is usually introduced as a

---

[12]Our construction of P_{num} can be seen as a specific instantiation of the general suggestion made in (Benci et al., 2010, section 5).

special case of measure theory (*e.g.* Dudley, 2004); in this context, the idea that a probability measure is real-valued —which originally was part of an axiom (axiom III in Kolmogorov, 1933, p. 2)— is so basic that it remains a tacit assumption.

Using NSA, the natural extension of the co-domain is $^{*}\mathbb{Q}$ rather than $\mathbb{R}$. In order to assign sizes to all finite subsets of $\mathbb{N}$, we already need all of $\mathbb{N}$ (and zero), and to assign probabilities to all finite lotteries, we need all numbers in $[0,1]_{\mathbb{Q}}$. Thus, in order to assign sizes to all subsets of $\mathbb{N}$, finite as well as infinite, a larger set is required, which can be obtained using NSA: we consider $^{*}\mathbb{N}$ (and zero), or at least an initial part of it, $\{0,1,\dots,\alpha\}$ with $\alpha \in {}^{*}\mathbb{N}\backslash\mathbb{N}$. To assign probabilities to a countably infinite lottery (or all hyperfinite ones), we need $[0,1]_{*\mathbb{Q}}$ as the range.

Although NSA can be introduced axiomatically, for instance using alpha-theory, and thus need not be intrinsically more difficult than learning standard analysis, the latter was developed first and is still much more common. There are at least two reasons why mathematicians generally prefer the standard framework of classical analysis over the framework of nonstandard analysis. Firstly, nonstandard analysis was rigorously developed much later than standard analysis. So, even though it is not intrinsically more difficult to learn nonstandard analysis, it has the disadvantage of unfamiliarity. Secondly, because of the Transfer Principle, non-standard analysis does not yield any new information about the standard real numbers. However, the concept of a fair infinite lottery simply begs for the co-domain of the sought-for probability functions to be modelled using standard numbers. Thus the Transfer Principle cannot be used here to transport us back to the familiar shore of the standard real numbers. But we do stress that the standard but ultimately unsatisfactory $\mathbb{R}$-solution can be interpreted as the standard part of any suitable $^{*}\mathbb{Q}$-function.

2.6.4.2 Standard valued approximations and the failure of SUM

As we know from section 2.3, asymptotic density is only FA. Admittedly, CA is not achieved by the $^{*}\mathbb{Q}$ approach either, but the latter tells us why this is not so: the extension of FA for finite lotteries to the countably infinite case does not lead to CA, but rather to HFA (or HCA in a trivial sense). Or, put differently, this solutions shows us that we we can extend our finite SUM intuition to HFA. We now investigate how this relates to the FA of asymptotic density.

Let us approximate the hyperreal-valued probability measure (up to infinitesimals) by a real-valued measure. To this end, NSA provides the standard part function, *st* (paragraph 2.4.4). First, we will illustrate this by three examples concerning the lottery on $\mathbb{N}$:

- Single ticket: $A = \{n\}$ for some $n \in \mathbb{N}$. Then $P_{num}(A) = num(\{n\})/\alpha = 1/\alpha$. Since α is an infinite hypernatural, its inverse is an infinitesimal with standard part zero: $st\big(P_{num}(A)\big) = 0$.

- All but one ticket. $B = \mathbb{N}\backslash\{n\}$. Then $P_{num}(B) = num(B)/\alpha = (\alpha - 1)/\alpha = 1 - 1/\alpha$. So $st\big(P_{num}(B)\big) = 1$.

- Consider an arbitrary $m \in \mathbb{N}$. For all $k \in \{0,1,\dots,m-1\}$, let C_k be the set $\{n \mid n \mod m = k\}$. Then $st\big(P_{num}(C_k)\big) = st(1/m) = 1/m$ for all k.

In the examples, the standard part of the hyperrational-valued probability measure P_{num} equals the real-valued asymptotic density. This close connection should not come as a surprise, since P_{num} and P_{ad} are, respectively, the ultrafilter equivalence class (or alpha-limit) and the Hahn-Banach limit of one and the same ω-sequence. In fact, what we have observed in the examples holds in general:[13]

$$st \circ P_{num} = P_{ad} \tag{2.30}$$

Proof.
Fix an arbitrary $A \in \mathcal{P}(\mathbb{N})$. We need to show that $st\big(P_{num}(A)\big) = P_{ad}(A)$.

We know that $P_{ad}(A)$ is the Hahn-Banach limit of $\langle a_n \rangle = \langle \frac{\#(A \cap \{1,\dots,n\})}{n} \rangle$ (eq. 2.3). From equation (2.9), we know that in order to proof that $st\big(P_{num}(A)\big)$ is the Hahn-Banach limit of this sequence, we need to find $H_0, H_1 \in {}^*\mathbb{N} \backslash \mathbb{N}$ with $H_0 < H_1$) and an internal sequence of hyperreals $h_{H_0}, \dots, h_{H_1}$ such that $\sum_{N=H_0}^{H_1} h_N = 1$ and:

$$st\big(P_{num}(A)\big) = st\left(\sum_{N=H_0}^{H_1} h_N {}^* a_N \right)$$

First transform the left hand side using equations 2.6 and 2.21:

$$P_{num}(A) = \left[\langle \frac{\#(A \cap \{1,\dots,n\})}{n} \rangle \right]_{\mathcal{U}} = {}^* a_\alpha$$

Then we see that the above equality is equivalent to:

$$st({}^* a_\alpha) = st\left(\sum_{N=H_0}^{H_1} h_N {}^* a_N \right)$$

which holds if we choose $H_0 < \alpha$, $H_1 \geq \alpha$, $h_N = 0$ for all $N \neq \alpha$ and $h_\alpha = 1$.[14] $\qquad\qquad \square$

This relation elucidates why (generalized) asymptotic density has poorer additivity properties than the non-standard measure: each time the standard part function is applied, a rounding-off error (up to infinitesimals) is introduced. When adding a finite number of rounded values this still only introduces an infinitesimal error, but when adding an infinite number of approximated terms, the total error may be appreciable. In particular, when summing over the family of all singletons, we see SUM fail most dramatically: the total error is as high as $\alpha \cdot 1/\alpha = 1$, which is 100% error in case of probabilities.

---

[13]This is true for *any* P_{num} in the family of solutions corresponding to the freedom of choice in the free ultrafilter.

[14]For the special case where $\langle a_n \rangle$ converges, $st({}^* a_\alpha) = \lim_{n \to \infty} a_n$ (eqs. 2.6 and 2.8). Thus the proof for that case boils down to: $st\big(P_{num}(A)\big) = \lim_{n \to \infty} a_n = P_{ad}(A)$

2.7 Conclusion

In this chapter, we have argued that fair infinite lotteries can best be described using tools and concepts of non-standard analysis. We have constructed a uniform probability measure which is defined on the full power set algebra of $\mathbb{N}$, and which takes its values in the non-standard extension of the $[0,1]_{\mathbb{Q}}$-interval. The construction is closely related to that of asymptotic density measures. The resulting probability measure is uniform, it gives a non-zero probability to the event that a given ticket wins, and it is not just finitely but also infinitely additive.

Lewis (1986a) argued that every possibility should be assigned a non-zero probability. This has been taken by him and others as a reason for advocating probability functions that take their values in a non-standard $[0,1]$-interval. We do not advocate the principle of Regularity in general. But we have seen that in the context of the lottery on the natural numbers, it follows from desiderata that we are committed to on independent grounds. So it was natural for us to try to work out a non-standard approach in some detail.

Proposals to construct non-standard measures on infinite sample spaces have been made in the literature (Cutland, 1983). But instead of taking the natural numbers as their domain, these probability functions operate on a non-standard extension of the natural numbers. This entails that they do not provide a genuine solution to the problem of lotteries on the natural numbers; instead, they change the problem. The probability measure that we have constructed is of mixed origins. It stands with one leg in the classical universe, and with its other leg in the non-standard universe.

Asymptotic density and our probability measure are based on the same sequence of partial fractions, and both look at its behavior 'at infinity'. Only the formalization of this statement is achieved differently: by (a generalization of) classical limits (Schurz and Leitgeb, 2008) in the first case, and by free ultrafilter-based equivalence classes in the second. Asymptotic density takes values in the standard $[0,1]$-interval. The price for this is that such probability functions can only be finitely additive for fair lotteries on $\mathbb{N}$.

The solution that we propose meets the conceptual and intuitive requirements connected with a lottery on $\mathbb{N}$: it is fair, uniform, defined on the whole power set algebra of $\mathbb{N}$, and infinitely additive. We have seen that we do not end up with a unique solution. The probability functions that we propose are only determined up to the choice of an ultrafilter. As far as we can presently see, no choice of ultrafilter is superior to any other. We do not exclude that when more intuitive constraints are imposed, the class of satisfactory probability functions can be narrowed down further. But at present we see no way of doing so.

The most notable feature of our solution is that whereas it is infinitely additive, its additivity behavior is not adequately described by summing over the natural numbers, but by summing over a non-standard extension of the natural numbers. One might naively expect the relevant sum generally to be an ω-sum of non-standard weights. But it emerged that the probability of an ω-length family of events will be a sum of a much longer order type. In this context, it should be recalled that even Kolmogorov, the discoverer of the classical σ-additivity axiom, emphasized that the

condition is merely a useful assumption of idealization but is not contained in the meaning of probability.[15] It turned out that in the context of lotteries on the natural numbers, σ-additivity is not the *right* idealization.

The obvious next step would be to consider larger, nondenumerable lotteries, such as lotteries on $\mathbb{R}$. The description of fair lotteries on $\mathbb{R}$ is a mathematically significantly more difficult affair, and will be left for future work.

---

[15]Kolmogorov (1933, p. 15) states: "Infinite fields of probability occur only as idealized models of real random processes. *We limit ourselves, arbitrarily, to only those models which satisfy Axiom VI.*" (Emphasis in the original.) In Kolmogorov's paper, 'field' means 'algebra' and 'Axiom VI' refers to the 'Axiom of Continuity' which, together with FA, automatically leads to CA in cases where the algebra is a σ-algebra.

Chapter ⚂

Stratified Belief and Ultralarge Lotteries

> *There are no whole truths; all truths are half-truths. It is trying to treat them as whole truths that plays the devil.*
>
> Alfred North Whitehead (Price, 1954, p. 14)

> *If the doors of perception were cleansed everything would appear to man as it is—infinite.*
>
> William Blake (1790)

> *Five is a sufficiently close approximation to infinity.*
>
> Robert Firth

3.1 Introduction

Whereas the previous chapter dealt with the foundations of probability theory in relation to infinite lotteries, the current chapter deals with the epistemology of finite lotteries. The relation between the two cases will be examined in Chapter 4.

In order to study the epistemology of yes–no beliefs, in particular the conditions for their rational acceptability, in so far as they are based on probabilistic information, we will focus on a simple example of a lottery. Consider a fair lottery with N tickets, exactly one of which will be randomly selected as the winner. This game of chance can simply be described by a uniform probability function, which assigns a probability of $\frac{1}{N}$ to each ticket. Clearly, the description of a fair lottery does not pose a problem within probability theory, but the interplay of probabilities and rational beliefs triggers epistemological questions. We are interested in what is rational to

believe for a participant in such a lottery in the case that N is very large and the winning odds of a single ticket are correspondingly very small.

If you own only one ticket in such a large lottery, it may seem rational for you to believe that your ticket will not win. Buying another ticket does not increase your odds very much, so it still may seem rational for you to believe that none of your tickets will win. Suppose that you keep buying tickets, each with a very small probability of winning, and that you keep believing that none of your tickets will win. At some point, you will own all the lottery tickets and thereby you will be certain that one of them will win, which contradicts your belief that none of your tickets will win. This is the Lottery Paradox, originally stated by Kyburg (1961).[1]

The Lottery Paradox occurs when three *prima facie* plausible principles are combined: the Lockean Thesis, the Conjunction Principle, and the Law of Non-Contradiction. This is the first principle:

Lockean Thesis (LT, Informal Version) It is rational to believe a statement if the probability of that statement is sufficiently close to unity.

The second principle, the Conjunction Principle or CP, states that if it is rational to believe two statements, it is also rational to believe their conjunction. The Law of Non-Contradiction or LNC expresses the idea that it is never rational to believe a contradiction. According to Kyburg himself, it is the employed aggregation rule for beliefs, CP, that causes the paradox (Kyburg, 1961). Whereas Kyburg's argument that rational belief is not closed under conjunction was supported by Foley (1979) and Klein (1985), the idea that CP is the cause of the problem is now a minority position. Some doubt the Law of Non-Contradiction (Priest, 1998), but most contemporary authors are more suspicious of the Lockean Thesis. It has been suggested that LT be modified with a defeater-clause. It seems natural to assume that such a defeater can be made mathematically precise, but Douven and Williamson (2006) show that any formally precise defeater does not work to avoid the Lottery Paradox, reducing much of the initial appeal of this solution.

In this chapter, we will analyze the Lottery Paradox as an instantiation of vagueness. After all, the problem only occurs for a lottery ranging over a *large enough* number of tickets, making the probability of winning with a single ticket *small enough*. Also in the informal phrasing of LT, a vague element is present, where it states that the probability has to be *sufficiently close* to unity.

It is the goal of this chapter to find a formal solution to the Lottery Paradox that does justice to this vagueness. It may not seem likely that a formal solution of this type exists: what mathematical method can help us out if the problem is intrinsically vague?

We propose to apply relative or stratified analysis (Hrbacek, 2007, Hrbacek et al., 2010), a type of non-standard analysis. Based on stratified analysis, we will give a formalization of LT and refer to the resulting type of rational belief as 'Stratified Belief'. As it turns out, CP will have to be adapted too, in order to be compatible with this soritic version of LT.

---

[1]The paradox can be restated in terms of knowledge (Nelkin, 2000). However, here we will address only the original phrasing in terms of rational belief.

Regarding CP, the conclusion of this chapter is close to the position of Kyburg, Foley, and Klein: we find that the Conjunction Principle is too strong to be expected to hold for rational beliefs. However, we do argue in favor of a weakened form of CP. So, like Kyburg, we claim that you would be wrong to keep believing that none of your tickets will win: the repeated addition of an extra ticket with a small probability does not guarantee that the total probability of all the tickets that you own remains small. The total probability of winning will be considerable before you have bought all the tickets. Yet, knowing this does not tell you exactly when you should stop adding tickets or change your opinion. The question "When do the winning odds of a number of tickets cease to be small?" is not all that different from "When does a number of lottery tickets start to be a heap?" In the application of the aggregation rule, we see that induction fails at some point, making the property of rational acceptability of beliefs intransitive.[2]

In the context of the philosophy of probability, two varieties of probability are considered: objective (or physical) probability on the one hand, and subjective (or epistemic) probability on the other. The probabilities occurring in physics are taken to be objective[3] and are thought of as real numbers in the $[0, 1]$-interval. Subjective probabilities are often referred to as 'degrees of belief' in the Bayesian literature (Ramsey, 1931, Foley, 2009). Unlike objective probabilities, degrees of belief do not necessarily have a numerical value. However, in the case of a lottery or other situations in which all relevant information about the objective probabilities is explicitly available, the agent's subjective probability assignments should be equal to the objective probabilities. This requirement has been dubbed "the Principal Principle" by Lewis and we consider it as a minimal, necessary condition for rationality, underlying any attempt to formalize the notion of rational belief. Throughout this chapter, we will focus on the case in which the subjective probabilities are indeed equal to the objective ones. Therefore, we may use the term 'probability' without further qualification.

Notation Consider an N-ticket lottery, with N some natural number at least equal to 2. Here, we introduce some notation for probabilities of statements concerning such a lottery. Denote the set of N tickets as: $T_N = \{t_1, \ldots, t_N\}$. From this set, exactly one ticket will be randomly selected and assigned to be the winning ticket. If A is a subset of T_N, let $\varphi(A)$ denote the statement that one of the tickets in A is the winner. We introduce a similar notation for loss statements: if B is a subset of T_N, let $\psi(B)$ denote the statement that none of the tickets in B is the winner. Clearly, $\psi(B)$ is equivalent to $\varphi(T_N - B)$.

The assignment of probabilities (P) to win and loss statements can be done as follows:

---

[2]Intransitivity is a typical symptom of problems that are soritic in nature.

[3]Of course, even those probabilities are subjective to a certain extent: probability is a way to model a system about which we have insufficient information to predict its behavior with certainty or to summarize information about large numbers of particles.

$$
\begin{aligned}
P\big(\varphi(A)\big) &= \#(A)/N \\
P\big(\psi(A)\big) &= 1 - \#(A)/N
\end{aligned}
\tag{3.1}
$$

where $\#(A)$ denotes the number of tickets in the set A.

Equation (3.1) is not intended to define probability measures (in particular, $P \circ \psi$ is not additive, so it cannot be a measure), but rather to introduce some shorthand notation. If A is a singleton, say $\{t_i\}$, then we write $\varphi(A)$ as φ_i; we define ψ_i analogously. So, we write $P(\varphi_i)$ as shorthand for $P\big(\varphi(\{t_i\})\big)$ and $P(\psi_i)$ for $P\big(\psi(\{t_i\})\big)$.

Structure The chapter is structured as follows. In section 3.2, we frame the Lottery Paradox in the broader context of finding a way to relate real-valued probabilities to binary belief states. We specify three desiderata required for the conversion from probabilities to beliefs. Because the threshold-based model is a popular approach, in section 3.3 we review this model and show that it does not satisfy CP. In section 3.4, we show that statements about a lottery, for which the paradox may be invoked, show typical traits of vagueness. We claim that the Lottery Paradox occurs in the threshold-based model, precisely because the approach does not deal well with these soritic aspects. In section 3.5, we argue in favor of applying ideas from relative analysis to the epistemology of large lotteries. This leads to the main result of this chapter: our model of 'Stratified Belief' in section 3.6. In section 3.7, we discuss the relation of our proposed solution to contextualism and the epistemicist account of vagueness. We summarize our findings in section 3.8.

3.2   Mapping $[0,1]$ onto $\{0,1\}$

Underlying the Lottery Paradox, there is a more general question: how do we relate probabilistic information, represented by real numbers in the $[0,1]$ interval, to simple yes–no judgments (beliefs), which can be represented by the binary values $\{0,1\}$? Note that there is an asymmetry between i) either you believe p or you don't, and ii) either you believe p or you believe its negation. Here, the binary values refer to the first interpretation.[4] So, in accordance with Leitgeb (2010), we take beliefs to distinguish between three states: belief, disbelief, and suspension of judgement.

A first answer to the above question could be: we shouldn't (convert probabilities to unqualified beliefs). If we have detailed information in the form of probabilities, we should stick to that. Indeed, in the lottery case, it is easy to calculate the winning odds of any set of tickets simply by adding the individual probabilities. However, stating that the winning odds of a subset is 0.125, for instance, does not answer the

---

[4]It is clear that the second notion of belief is even more restrictive: it does not allow a representation of agnosticism, doubt, or suspension of belief, all of which would at least require either a third value besides 0 and 1 or the option not to assign any value at all. We will not deal with this issue here, because adding values shifts but does not remove the problem we want to study: under which conditions is it rational to believe p? We will take into account different strengths of belief in subsection 3.7.2.

question "Do you believe that all tickets in this set will lose?" There still seems to be a need for a translation.

Foley has argued that it is indeed indispensable to have some way of extracting simple true-false judgments out of detailed probabilistic information (Foley, 2009). He comes up with some very convincing examples: in court, for instance, the judge or jury has to choose between 'guilty' and 'not guilty', no matter how fine-grained the information on which they must base their conclusion. In daily life, simple answers are often required to facilitate communication. As Mencken put it: "The public ... demands certainties; it must be told definitely and a bit raucously that this is true and that is false. But there *are* no certainties" (Mencken, 1919, p. 46).

This situation is very analogous to that in image analysis, where black-and-white pictures are sometimes more useful than grey-scale ones. Even the reasons behind this are remarkably similar: conversion of grey-scale images to black-and-white images helps to make certain features stand out more, or to facilitate sending the file by e-mail—think of real-time images sent by a distant space craft—both of which can be seen as facilitating communication. We will encounter this analogy again in the next section.

At this point, we formulate the main requirements for dealing with the relation between probabilities and beliefs:

Desideratum 1 There should be a method to translate continuous probabilities into discrete beliefs.

Desideratum 2 There should be a rule to aggregate these beliefs.

Desideratum 3 The translation method and aggregation rule should be chosen such that together they do not lead to a Lottery Paradox.

First, we look at a candidate for Desideratum 1: a popular model that achieves the translation requirement is the threshold-based model for rational belief; it will be discussed in detail in the next section. Now, let us think of a rule that could be used for Desideratum 2. For logical truths, the aggregation rule is simple: if you start out with two or more true propositions, their conjunction is true as well; this rule of inference is called *Adjunction* or *Conjunction Introduction*. The *Conjunction Principle* (CP) states that something very similar holds for rational beliefs instead of logical truths: if two or more propositions are rationally acceptable, their conjunction is rationally acceptable as well. We can show however (in section 3.3.2), that using the threshold-based model for rational belief as Desideratum 1 and CP as Desideratum 2 makes it impossible to meet Desideratum 3. This is no new result: this is precisely the Lottery Paradox.

The approach of this chapter is to critically examine the threshold-model, and suggest an alternative model without explicit thresholds, which avoids the Lottery Paradox without abandoning CP completely. In subsection 3.6.1, we will propose a new option for Desideratum 1. In subsection 3.6.3, we will find a matching choice for Desideratum 2, such that also Desideratum 3 is fulfilled.

3.3   Threshold-based model of the Lottery Paradox

In this section we give an overview and critical examination of the wide-spread threshold-based model for rational belief. It rephrases the Lockean Thesis as follows:

Threshold Belief (Informal Version) It is rational to believe a certain statement if the probability of that statement is at least equal to a given threshold value.

This invites a crucial question: what is the value of this threshold used by actual people (in a given context)? And, what ought it be? Usually, the threshold is taken to be a number close to unity, such as 0.999. Achinstein (2003) considers a threshold of exactly $\frac{1}{2}$ to be an option, although he prefers the more liberal but less precise condition that the threshold be larger than $\frac{1}{2}$. He also mentions the less liberal but even more vague condition that the threshold be 'much larger than' $\frac{1}{2}$.

When we know that a statement has a sufficiently high probability, Threshold Belief tells us that it is rational to believe the statement fully; this transition is analogous to image processing, where a grey-scale picture can be converted into a black-and-white image by setting a threshold on the brightness: darker pixels become black and brighter pixels become white. In that context too, the question of finding the 'right' thresholds is a non-trivial one: in some situations the threshold and even the whole scale may be relative, which leads to adaptive thresholding and image enhancement, respectively (Shapiro and Stockman, 2002). Similar to adaptive thresholding, if the threshold-model is any good at all, it should allow for context-dependent thresholds. Image enhancement is analogous to dramatization: even in one context, it seems as though people use different thresholds in order to contrast cases. This may have little to do with rational beliefs, but if you take into account what it takes for humans to get a message across or remember it (Lang, 2000), context-depending thresholds definitely serve a function.

Many related questions could be raised, for instance concerning experimental accessability and measurement precision of the thresholds (Douven and Uffink, 2003), possible hysteresis (Égré, 2009) *etc.* However, we will not dwell on these points, for there are more substantial problems with this approach, which will cause us to discard it completely.

3.3.1   Formalizing the Lockean Thesis

We write $B(x)$ to denote the belief in statement x, and $B(x) \in R_\alpha$ to denote that it is rational for agent α to believe x. Then, the threshold-version of the Lockean Thesis can be formalized as follows.

Threshold Belief (Formal Version)

$$B(x) \in R_\alpha \quad \Leftrightarrow \quad P(x) \geq \theta \tag{3.2}$$

where θ is the threshold value or cut-off (a real value in $[0, 1]$).[5]

---

[5]Clearly, the range for θ can be made smaller. The interval given here is chosen such that it is broad enough to encompass all the threshold values one may want to consider. There are good

We could have used an agent's personal probability estimate P_α, but since we assume that all agents use the same probability function for a fair lottery (*cf.* the aforementioned Principal Principle), we have omitted the subscript.

3.3.2 Illustration of the failure of CP

We can easily show that:

$$B(\psi_1) \in R_\alpha \text{ and } B(\psi_2) \in R_\alpha \not\Rightarrow B(\psi_1 \wedge \psi_2) \in R_\alpha \tag{3.3}$$

This can be seen as follows. Applying equation (3.2) to the left-hand side:

$$B(\psi_1) \in R_\alpha \Rightarrow 1 - \frac{1}{N} = P(\psi_1) \geq \theta$$

$$B(\psi_2) \in R_\alpha \Rightarrow 1 - \frac{1}{N} = P(\psi_2) \geq \theta$$

Applying equation (3.2) to the right-hand side:

$$B(\psi_1 \wedge \psi_2) \in R_\alpha \Rightarrow 1 - \frac{2}{N} = P(\psi_1 \wedge \psi_2) \geq \theta$$

However, $1 - \frac{1}{N} > 1 - \frac{2}{N}$; thus $1 - \frac{1}{N} \geq \theta$ does not guarantee that $1 - \frac{2}{N} \geq \theta$ (unless when $\theta = 1$, but this case is of little interest for modeling belief anyway).

In words, this shows that, given a value for the threshold, θ, there exists a sufficiently large lottery such that the losing odds of any specific ticket are at least equal to the threshold, but the losing odds of two tickets are below it. If rational acceptability is identified with having probability above a certain threshold—that is to say, the 'if' in the informal phrasing of LT is interpreted as 'only if'—, the conjunction of only two rational beliefs may fail to be rationally acceptable.

3.4 The vague lottery: a heap of tickets

In this section, we step away from the existing threshold-based model of LT and take a fresh look at the Lottery Paradox. The first three observations of this section will fuel our search for a new formalization of LT, which will be the foundation of a new solution to the Lottery Paradox. The fourth observation foreshadows the idea that also CP will have to be adapted.

---

reasons to doubt that $\theta \leq \frac{1}{2}$ or $\theta = 1$ are viable choices (Achinstein, 2003). Moreover, Achinstein and many other authors suspect that θ should take a value much closer to 1 than to $\frac{1}{2}$.

3.4.1 Observation 1: The soritic lottery

In formal epistemology, the Lottery Paradox may be seen as an easier (clearer) problem to work on as opposed to cases in which the probabilities are not explicit, such as the 'Preface Paradox' (Makinson, 1965) that deals with a book containing a lot of statements, in each of which the author has 'very high' confidence. This viewpoint may be misleading: to invoke the Lottery Paradox, the number of tickets N just has to be 'large enough'. Another way of saying this is: the probability of winning $\frac{1}{N}$ has to be 'small enough'. 'Large enough' and 'small enough' are vague concepts: vagueness is at the heart of the problem (both for the Lottery and the Preface Paradox).

There is also vagueness in the informal version of the Lockean Thesis (LT), which mentions a probability *sufficiently close* to unity, but this vagueness is not reflected in the usual formalization of LT with thresholds. The threshold-model has stepped into the trap of illusory exactness: it attempts to set a sharp boundary around what is large, in particular what is a large enough probability. The idea that 'being large' is a well-defined property does not agree with our normal use of the concept. It is the position of this chapter that counter-intuitive results such as the Lottery Paradox follow from just this property. The Lottery Paradox is merely a symptom of this deeper problem.

This first observation suggests that we should formalize LT such that the vague aspect of it is respected and apply this vague version of LT to the lottery case. Now, the puzzle is how to deal formally with the kind of vagueness at issue here.

3.4.2 Observation 2: Contextual element

Yet another formulation of the lottery case, which although still vague, at least has the advantage of making clear that 'large' and 'small' are relative concepts: in order to make it plausible that it is rational to believe that none of your tickets will win, the number of tickets that you own has to be very small compared to the total number of tickets. Although it is less quantitative, in a way the latter statement is more informative than "You own two tickets in a fair 10,000,000-ticket lottery with one winner".

The observation that the vagueness involved in the Lottery Paradox is relative suggests that we should allow a contextualist element in the solution.

3.4.3 Observation 3: Analogy with fair lottery on ℕ

As Lavine (1995, p. 389) points out: "It is a familiar idea that our knowledge about the infinite is obtained by in some sense extrapolating or idealizing our knowledge of and about the finite." This idea can also be applied in the opposite direction: large finite phenomena are often modeled by infinite ones. In physics, a long thin cylinder may be taken to be an infinitely long wire, a large surface to be an infinitely large area (*e.g.* by using periodic boundary conditions), and so on. For our problem, it seems natural that a *sufficiently large* lottery behaves qualitatively the same way as does an *infinite* lottery.

If we want to describe a lottery on $\mathbb{N}$, it turns out that we do not have the freedom to assign fair odds to the tickets within the classical axiomatization of probability theory developed by Kolmogorov (1933). A fair lottery on $\mathbb{N}$ can be described if we drop the requirement of Normalization or that of Countable Additivity. If we, like de Finetti (1974), choose the latter option, the probability of winning of a single ticket (or any other finite subset of $\mathbb{N}$) is zero, the same probability as we assign to the impossible event. Real-valued, finitely additive probabilities are not fine-grained enough to distinguish between the impossible event and some possible but 'highly unlikely' events. Within the real numbers, there is no way of quantifying just how unlikely these events are; all we can express is that their probability is zero. Although there is a clear *qualitative* difference between possible and impossible events, some of the associated probabilities are *quantified* by the same number.[6] Clearly, this situation is unsatisfactory from an epistemological point of view. If we allow infinitesimals in the range of the probability function, it is possible to distinguish between the probability of the empty set and a non-empty set (*cf.* Chapter 2).

Thus, the infinite version of our problem requires the use of non-standard analysis (NSA), originally developed by Robinson (1966). The idea of NSA is to extend $\mathbb{N}$ and $\mathbb{R}$ to the strictly larger sets ${}^{\star}\mathbb{N}$ and ${}^{\star}\mathbb{R}$. ${}^{\star}\mathbb{N}$ is called the set of hypernaturals and contains infinite numbers, strictly larger than any natural number. ${}^{\star}\mathbb{R}$ is called the set of hyperreals, and next to infinite numbers it also contains their inverse— infinitesimals—which are smaller in absolute value than any strictly positive real.

If we consider a fair lottery on a finite subset of $\mathbb{N}$, the probability assignment is unproblematic: the winning odds of any non-empty subset are non-zero. However, in our reasoning about such a lottery we sometimes deal with very small probabilities as if they were zero. In such a situation, our beliefs are not fine-grained enough to distinguish between the impossible event and some possible but 'highly unlikely' events. Because of the analogy between the infinite lottery puzzle and the Lottery Paradox, we ask: can NSA be applied to solve the latter problem too?

It should come as no surprise that the study of an infinite lottery leads to a system in which one has infinite numbers and infinitesimals available. However, at first sight, there seems no good reason to bother with infinities when dealing with beliefs about a finite lottery, no matter how large. Although we do not want to introduce infinite numbers, this is not a valid objection to the use of NSA for this problem, since there are approaches to NSA that do not extend the standard sets into the transfinite, but work entirely within the standard sets $\mathbb{N}$ and $\mathbb{R}$. This is true for Nelson's 'internal set theory' (IST) (Nelson, 1977), as well as Hrbacek's relative analysis (inspired on IST) (Hrbacek, 2007).

A promising aspect of relative analysis is that it is able to cope with soritic concepts. Some vagueness is present in Robinson-style NSA, too. To see this, observe that the standard sets $\mathbb{N}$ and $\mathbb{R}$ have no largest element. Likewise, in ${}^{\star}\mathbb{N}$ and ${}^{\star}\mathbb{R}$ there is no smallest infinite number: although for any given number, it is easy to

---

[6]The problem is not limited to impossible and highly unlikely events: there are many more events that are qualitatively distinguishable (they differ up to a finite subset of $\mathbb{N}$) but get exactly the same probability assignment. In particular, probability one is not only assigned to the necessary event ('Some ticket of $\mathbb{N}$ will win'), but also to infinitely many other events: $\mathbb{N}$ minus any finite subset.

determine whether it is finite or infinite, the border between the finite numbers and infinite numbers cannot be pinpointed by giving a number at or near it. In $^{*}\mathbb{R}$, there is no largest infinitesimal, nor a smallest positive non-infinitesimal number. Stated informally, the result of the internal-set approach is that the vague border between the finite and infinite gets 'pushed down' into a vague border between standard and large size, whereas that between finite and infinitesimal 'scales up' into a vague distinction between standard and small size. With relative analysis, we will model the small probabilities that are indistinguishable from zero as 'relative infinitesimals' or 'ultrasmall numbers'.

3.4.4   Observation 4: Weakening CP

For statements supported by probabilistic considerations, it is not self-evident that CP should hold. First, let us recall how to calculate the probability of a conjunction. In the case of two tickets for the same lottery, their losses are dependent events and the probability of a loss statement decreases linearly with the number of tickets.[7] Nevertheless, if we start out believing a statement that has a sufficiently high probability, the aggregation of some additional statements with equally high probability will not dramatically change our (degree of) belief in the conjunction. So we do not expect to see CP fail completely either. For the aggregation of beliefs, we expect to find a weakened version of CP. In particular, for beliefs concerning a large lottery, what we expect intuitively is this:

Weakened Conjunction Principle for Beliefs (Informal Version) It is acceptable to aggregate a few rationally acceptable beliefs, but the conjunction of many rationally acceptable beliefs is not necessarily rationally acceptable.

In a threshold-based model, this intuitive rule is violated, for we cannot even allow the conjunction of two beliefs, as we have seen in subsection 3.3.2.

Another way of seeing that we may have to weaken CP is by observing the following statements:

- The probability of winning for one ticket is small.

- The probability of winning for two tickets is small.

- . . .

In mathematics, it is often tricky to correctly interpret a continuation with an ellipsis: it suggests a type of limit process. (The limit of classical calculus is not the only option; we will come back to this.) Here, the same care is needed, for the ellipsis does not generalize to the (incorrect) statement: "The probability of winning for all tickets together is small", but rather to:

---

[7]Consider two different tickets i and j in an N-ticket lottery. Then, the event that ticket i will not win has probability $P(\psi_i) = 1 - \frac{1}{N}$, while the event that both tickets i and j will not win has probability $P(\psi_i \wedge \psi_j) = 1 - \frac{2}{N}$.

- The probability of winning for a few tickets is small.

Whereas 'all tickets' refers to a definite number, 'few' is a vague term of course. As we have already seen in Observation 3, relative analysis may be able to deal with this formally.

3.5 Introduction to relative analysis

Observing that the Lottery Paradox is related to vagueness may seem like saying that it is a problem that escapes proper formalization. This is not true. For a first attempt at such a formalization, we may get inspiration from the praxis of physics.

3.5.1 Vagueness in physics

Physics is the prototype of a hard science, a fortress of exactness. Indeed, experimental physicists may put a lot of effort in high-precision measurements, such as of the mass of an elementary particle. Yet, physicists are also experts in estimating quantities: figuring out the right unit (dimensional analysis) and prefix, which expresses a power of 10. As long as they know where to locate a quantity on the logarithmic scale, most physicists do not need more precise values for their back-of-the-envelope calculations.

A way to represent structures at different scales of magnitude is given in Figure 3.1: this illustration is popular with researchers working in nanotechnology who want to present their work to a general audience and often adapt the original image to include examples of their specific field of study. Similar illustrations with logarithmic scales are used by astrophysicists to indicate how large their objects of research are.

Physicists also frequently use the words 'microscopic' and 'macroscopic'. What these terms mean, however, is ambiguous: it may depend on the field, or the more specific context in which they are used. For a computational physicist, used to simulating single molecules or a unit cell of a crystal, a mole of matter is definitely macroscopic, whereas it is microscopic for his colleague in astrophysics. Their close contact with the external world through experiments has sharpened their intuitive use of sliding scales of magnitude. Their heuristics include rules such as:

Small + Small = Small	Large + Large = Large
Few × Small = Small	Few × Large = Large
Small × Small = (Very) Small	Large × Large = (Very) Large
Small × Large = (undetermined)	

Although rarely made explicit, this type of rule is often employed in physical reasoning to estimate the order of magnitude of a quantity: if a certain contribution is (very) small compared to the effect one wants to describe, it can often be neglected. The above rules are very general, and could be applied to probabilistic problems, in
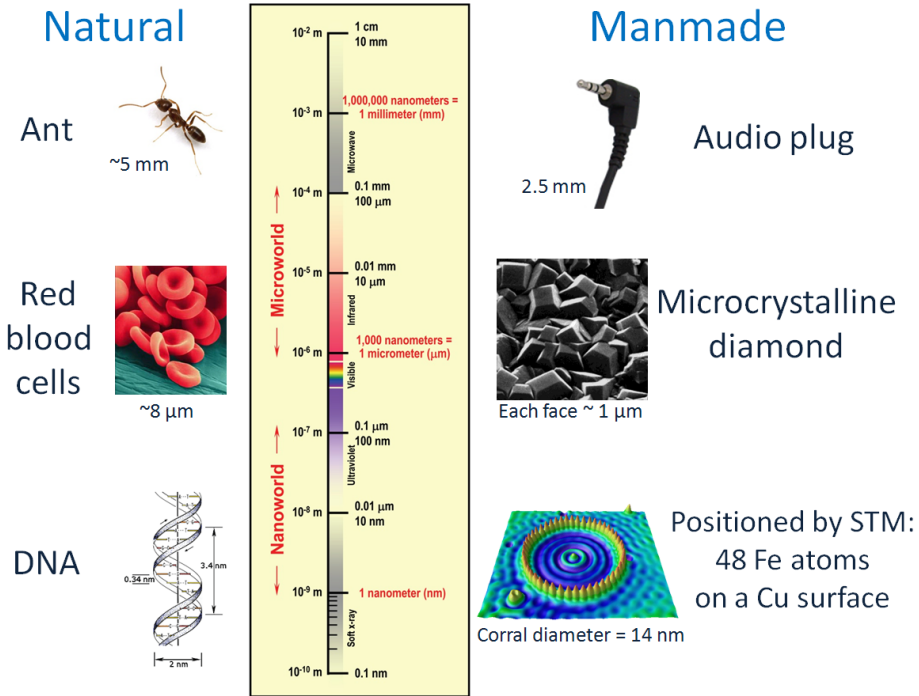
Figure 3.1: Illustration based on "The scale of things – Nanometers and more". Original designed by the Office of Basic Energy Sciences for the U.S. Department of Energy.

particular to lotteries, but this is a dangerous idea for as Hrbacek et al. (2010, p. 801) say: "Scales of magnitude play an important role in the thinking of physicists, but to a mathematician the concept seems incoherent." So, if we want to use the above rules to remedy the Lottery Paradox, we first need to find a consistent system to handle them.

As we already know from the previous section, by combining Observation 1, 2, and 3, we need a formalism that is capable of dealing with vagueness, provides contextual elements, and is a form of NSA.

The good news is that we do not have to develop a theory from scratch: there is a formal system available that precisely formalizes the contextual and vague concepts of largeness and smallness, and more generally different scales of magnitude. It is an approach to NSA developed in Hrbacek (2007), Hrbacek et al. (2010), called 'relative analysis' or 'stratified analysis'.[8]

---

[8]Of course, there are other formal systems that deal well with vagueness and imprecise probabilities, such as fuzzy logic, which has been applied to belief in Booth and Richter (2005), and interval-valued probabilities of Dempster-Schafer theory (Dempster, 1967), but we will not discuss those further, here.

3.5.2 Scales of magnitude

Because stratified analysis was developed to deal with soritic quantities, let us apply it to an example dealing with a heap, or rather a bucket, of sand. Relative to the bucket, which is mesoscopic (neither small nor large), a single grain of sand is negligibly small (microscopic). The whole beach, however, is gigantic (macroscopic) compared to the bucket. This is depicted schematically in Figure 3.2.

The numbers 0, 1, 2, 3, ... and $\frac{1}{2}$, $\frac{1}{3}$, ... are mesoscopic numbers in relative analysis. They are observable on the coarsest context level (to be defined below). In order to apply relative analysis, we may let the numbers correspond to a physical quantity. A bucket of sand is a conceivable quantity to us, humans: it is one "dose" of sand, so we may use it as a unit of sand (much like dm^3 is a unit of volume). In a different application (*e.g.* when describing the viewpoint of a small animal on the beach), it may be more useful to take one grain of sand as the unit (or mm^3 as the unit of volume). In this application, the grain of sand is mesoscopic, while both the bucket and the beach are macroscopic. Thus, one and the same quantity can be both gigantic and negligibly small, depending on the choice of unit (*i.e.* which physical interpretation is given to the number 1, which is always considered to be mesoscopic in relative analysis).

This example indicates that our observations are related to different scales. One aspect of this can be understood from Lavine's approach to finite set theory (Lavine, 1995). He uses a similar example (albeit with beans instead of grains of sand) as a physical model for learning addition. Although the bucket contains a finite number of grains, we have no idea how many: the bucket is an 'indefinitely large' supply of grains of sand. Likewise, the beach is indefinitely large compared to the bucket. Lavine points out that whenever necessary, we may take a larger supply of sand, but this does not mean we ever need an infinite amount. The context of set theory is less convenient to discuss 'indefinitely small': whereas we can consider larger and larger collections, at the bottom we find singletons, and cannot look at smaller scales. In other words, we cannot represent scales (of largeness and smallness) well by the counting numbers ($\mathbb{N}$) alone, but since the real numbers are closed under inversion, $\mathbb{R}$ may do better.[9]

3.5.3 Levels

Relative analysis formalizes the aforementioned scales of magnitude as levels, a new concept intended to correspond with the intrinsically vague concept of largeness. As was mentioned already, unlike Robinson-style NSA, which extends $\mathbb{N}$ and $\mathbb{R}$ to the strictly larger sets $^*\mathbb{N}$ and $^*\mathbb{R}$, relative analysis works within $\mathbb{N}$ and $\mathbb{R}$. To obtain the concepts of infinitesimals and infinite numbers within $\mathbb{R}$, relative analysis adds the new primitive binary predicate $\sqsubseteq$ to the language (Hrbacek, 2007).[10] The meaning

---

[9]Actually, $\mathbb{Q}$ would suffice for our current purpose.

[10]In Nelson's Internal Set Theory, a new *unary* predicate is introduced to signify "x is standard". To allow for multiple levels of standardness, Péraire and Wallet (1989) introduced a *binary* predicate to signify "x is standard *compared to* y"; they called their theory 'relative internal set theory' (RIST). Also Hrbacek's relative analysis is based on this binary predicate, for which he uses the symbol $\sqsubseteq$.

Grain	**Bucket**	**Beach**
< 1 mm	~1 dm	>1 km
Microscopic	Mesoscopic	Macroscopic

Negligibly small	Appreciable size	Inconceivably large
Ultrasmall	Standard	Ultralarge
Relative infinitesimal		Relatively infinite
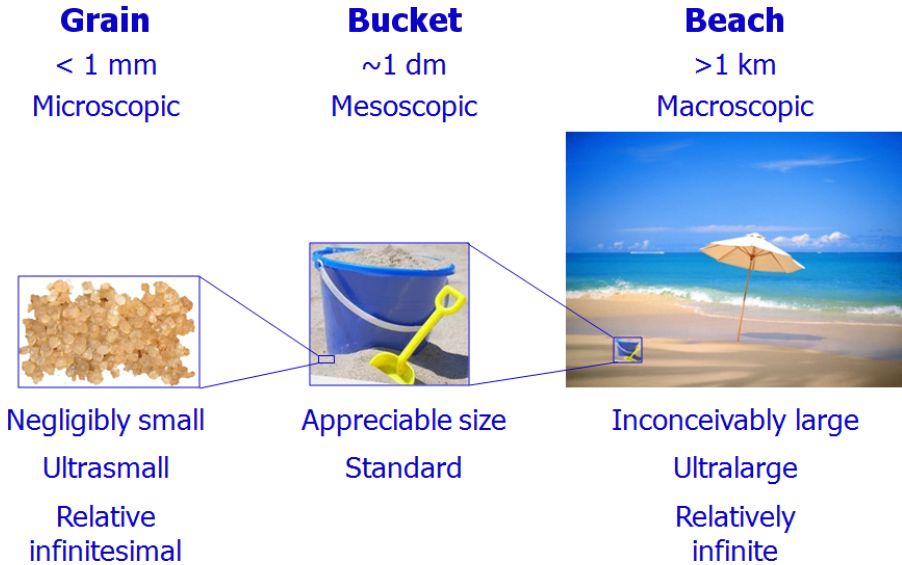
Figure 3.2: *A typical grain of sand is less than 1 mm in size. This is negligibly small compared to a typical bucket. The length of the beach is much larger than the dimensions of a bucket, so much that the factor is inconceivable to us.*

of this predicate is fixed by axioms, but informally $x \sqsubseteq y$ means that the number x is observable at every level where the number y is observable: the number on the left is observable at the coarsest level where that at the right is observable, and possibly at coarser levels too. To express that x and y are observable at the same level, we may write $x \sqsubseteq y$ and $y \sqsubseteq x$. To simplify the formalism, the axioms can be represented in terms of the aforementioned levels. So, "x belongs to the level of y" replaces $x \sqsubseteq y$. Levels stratify the set $\mathbb{R}$ into different levels or scales of magnitude; therefore, relative analysis is also referred to as stratified analysis.

3.5.3.1 Axioms for levels

In Hrbacek et al. (2010), there are eight axioms that fix the meaning of the level-concept. We paraphrase them here:

1. For every finite collection of real numbers, there is a coarsest level at which all the specified numbers are observable.

2. There are also finer levels, on which more numbers are available. Two levels can always be compared: we can always say which level is *at least as fine as* the other. Although levels are not sets (see subsection 3.5.5), $V_1 \sqsupseteq V_2$ is used to indicate that level V_1 is at least as fine as level V_2.

3. For any level, there exist non-zero real numbers that are ultrasmall compared to it.

4. Neighbor Principle: at every level there is a best approximation to any real number that is not ultralarge relative to that level.

5. Closure Principle: any number, function, operation, or set that is defined without mention of levels from parameters that are observable at a certain level is itself observable at that very same level. This level is called the context level or observation level.

6. Stability Principle: if a statement is true about its context level, then it also holds for any finer level. Using Stability, Closure can be generalized to the Transfer Principle, an indispensable tool in any approach to NSA (Benci et al., 2006a).

7. Definition Principle: for any internal statement (*i.e.* any statement that makes no reference to levels coarser than the context level) and any set A of real numbers, there exists a set B whose elements are exactly those elements of A for which the internal statement holds.

8. Density of levels: given two levels $V_1 \subset V_3$, there is a level V_2 such that $V_1 \subset V_2 \subset V_3$.

In relative analysis, the number 1 is always standard at every level. From the viewpoint of applications, this poses no limitation: after choosing any non-zero number as the unit of interest, one can always divide the whole scale by this value and thus achieve normalization. This points out the fractal-like structure of the real line: it is self-similar on all scales.[11]

3.5.3.2   Further terminology

Let us introduce some more of the vocabulary of relative analysis. Some numbers can be used to define a level V (*cf.* axiom 1): these numbers are called 'standard' compared to level V; they are also said to be 'observable at that level'.

Given a level V, there are non-zero numbers 'ultrasmall compared to' other numbers ($\ll_V$); ultrasmall numbers are 'relative infinitesimals'. Likewise, on level V there are numbers 'ultralarge compared to' other numbers ($\gg_V$, 'relatively infinite numbers'). A number that is not ultralarge compared to level V is called 'finite' compared to this level. Zero is the only number that can belong to a level while being ultrasmall compared to that level. A level does not contain any numbers that are infinite in comparison to it.

If the difference between two numbers is ultrasmall compared to level V, the numbers are 'ultraclose' to each other on level V ($\simeq_V$). In other words, they are indistinguishable (on that level), for their difference is negligible (on that level).

---

[11]With relative analysis, the fractal-like structure of $\mathbb{R}$ comes to expression in the fact that, for any level V, true formulas that quantify only over levels finer than V are exactly the same. However, not all formulas hold for all levels: for instance, a formula dealing with the existence of coarser levels will not hold at the coarsest level.

3.5.4 What levels are: a predicate on the domain

Like Lavine's indefinitely large sets within $\mathbb{N}$, levels on $\mathbb{R}$ are predicates on the domain. Thus, a level is a collection of real numbers. It contains all the numbers that have a unique name. At any given point in time, there can only be finitely many numbers that have a unique name. Therefore, the set of reals always contains infinitely many numbers that are larger than any uniquely named number (*cf.* Lavine's indefinitely large numbers), as well as infinitely many that are smaller than any uniquely named real number. Thus there is always 'room' for ultralarge and ultrasmall numbers. (Of course you can refer to these ultralarge or ultrasmall numbers indirectly, but you can never give all of them a unique name in finite time.)

When we *apply* relative analysis (as opposed to making a contribution to the development of its mathematical formalism), we may allow different criteria for numbers to be standard—not just uniquely named ones. Some examples:

- Distances in an image that are above the lower detection limit (resolution) and smaller than the upper detection limit (field of view).

- The quantity of sand measured as numbers of buckets, including partially filled buckets (fractions).

- Numbers of lottery tickets and their probabilities of winning.

3.5.5 What levels are not: sets

It is crucial to note that levels are not sets; in particular, the principle of mathematical induction does not hold for levels. For instance, adding or multiplying two numbers that are observable on a level V will result in a number that is standard compared to the level as well. The addition and multiplication can be generalized, but the result is only guaranteed to be standard for a relatively small (standard) number of terms or factors.

The set of real numbers, $\mathbb{R}$, forms a complete metric space, which means that every Cauchy sequence of real numbers converges to a real number. Unlike $\mathbb{R}$, the extended set $^{\star}\mathbb{R}$ is incomplete and as a consequence the limit concept of classical calculus is not available in it. However, within the various approaches to NSA, there is a different limit operation available (such as the alpha-limit in Alpha Theory of Benci and Di Nasso (2003a)).

Internal set theory and the related approach of Hrbacek do not introduce the non-standard extension $^{\star}\mathbb{R}$, but work on the standard set $\mathbb{R}$. Yet, there is a form of incompleteness that occurs within these approaches too: they introduce new predicates that do not necessarily have a defined limit. Mathematical induction is like a limit operation on sets; as noted at the beginning of this paragraph, it does not apply to levels. This elucidates the meaning of the ellipsis following the sentences "The probability of winning of one ticket is small", "The probability of winning of two tickets is small", ... Just as there is no last finite number in $^{\star}\mathbb{N}$, there is no sharp boundary around the number of tickets which collectively have a small probability of winning. Whereas the (incorrect) application of induction would lead us to a crisp,

but false conclusion ("The probability that the set of *all* tickets contains the winner is small"), the correct interpretation using levels leads to a sound generalization, albeit one containing a vague term ("The probability that the set of *a few* tickets contains the winner is small").

We may rephrase the analogy between extension-style NSA and relative analysis as follows. Asking what is the supremum of a set in $^*\mathbb{R}$ does not always make sense. Because $^*\mathbb{R}$ is incomplete, there is not necessarily an element of $^*\mathbb{R}$ at the point to which the limit seems to converge. The question posed may well point to a gap. Asking what is the supremum of a level in $\mathbb{R}$ is equally misguided. There is no such element; the question points to a gap in the system. There is no largest real number that is standard on a given level. The borders of a level are vague.

Some form of incompleteness—the presence of gaps in the set itself or gaps related to new predicates on the set—is necessary to invoke infinitesimals. If you try to make the vague borders of a level precise, relative analysis collapses to standard analysis. This points out that a model of beliefs based on relative analysis is incompatible with a model involving explicit, sharp thresholds.

3.5.6 The grain of sand, the bucket, and the beach

The sand example given at the beginning of subsection 3.5.2 can now be represented as: Grain of sand $\ll_V$ Bucket (standard) $\ll_V$ Beach, where the level V is that of a child playing on the beach (where V also contains the bucket).[12]

It is very likely that it does not make any difference to the child exactly how many grains of sand a bucket contains. If there were one grain more or less, he would not notice it, so he is indifferent to that. The same holds for any small number of grains that are added or removed, where 'small' means ultrasmall compared to the total number of grains of sand in the bucket. Clearly, the child will notice (and may care) if, for instance, all or half of the grains of sand in his bucket are removed.

The transition between will and will not notice is vague. In an experimental setting, this could mean that on one occasion the child will notice a certain difference, whereas he might not register this at another time. If you keep removing grains of sand one by one, you may get further than if you remove them all at once. We will not dwell on these general properties of vagueness, for there is another aspect that is more relevant to the application that we have in mind.

The child may weigh the bucket to estimate the number of grains of sand in it, or may even count the actual number of grains of sand.[13] In the first case, he will be able to tell the difference between at least some buckets that appear equally filled to the unaided eye. His increased ability to distinguish quantities can be represented by a finer level. In the second case, given enough time the child will be able to spot

---

[12]Observe that although a grain of sand is at a *smaller* scale of magnitude than a bucket and a beach is at a *larger* scale of magnitude, both appear at a *finer* level when the bucket is taken to be at the standard level. Thus, although the intuitive notion of scales of magnitude was an important motivation for developing a theory in terms of levels, the terms are not synonymous.

[13]It seems unlikely that the child would do any of this spontaneously, but he might be motivated to do so for a project at school or because there is a prize connected to getting the answer right. Of course, you are free to see this as an allegory of what is going on in science.

any difference in the filling of the buckets: it is at the finest level relevant for this example.[14]

3.6   Analysis of the Lottery Paradox using relative analysis

We are now ready to formulate a new definition for rational belief. This new definition is inspired by the Lockean Thesis, but developed along the lines of stratified analysis rather than a view based on thresholds.

3.6.1   Stratified model for rational belief

Whereas the sand example dealt with our sensory limitations (the (in-)ability to perceive differences in heaps of sand), forming rational beliefs about a lottery has to do with our mental ability to deal with numbers. Here, the problem is not that we do not see the difference between 0.999 and 1, but rather that we do not always attach any significance to the difference. Sometimes we deal with 0.999 as we would with 1, while in other contexts we may consider the difference as highly relevant. Even when we know probabilities are quantitatively different, our means of categorizing them qualitatively (large/small, comparable/different, ...) are limited. This limitation is useful, for we are finite beings with finite capacities.

Epistemology often deals with highly idealized agents, but in order to make sense of the Lottery Paradox, it is important to take into account at least this descriptive element: humans use scales of magnitude to make qualitative assessments. Given this limitation, we ask how we can deal with this as well as possible in the formation of rational beliefs. The use of scales may introduce a type of rounding error, which may go unnoticed when dealing with direct sensory information, but may produce some strange consequences if the starting information is precise and explicitly quantitative, as is the case with judgments based on probabilities.

If we have to judge whether or not we believe a certain event will take place, based on the numerically specified probability of the event, we have to compare the given probability value to unity. If the given value is zero or one, we can immediately answer that we do not or do believe that the event will happen. For intermediate values, we have to judge on a case-by-case basis whether or not the provided probability is close to unity: we will model this as being 'ultraclose to unity' in the sense of relative analysis.

---

[14]This idea of different levels in relative analysis is related to that of 'degrees of availability' in finite mathematics (Lavine, 1995). Of course, if the next task were to keep track of the number of atomic nuclei (or even smaller structures) in the bucket, and the child had an electron microscope (and a really long holiday ...) then an even finer level could be appropriate. This is possible, for we are working in the real numbers (a complete set with perfect self-similarity), where we have an unlimited supply of finer (axiom 3) levels available. We may also model this differently: by choosing the smaller structure of interest as the new unit of the context level.

Definition 1 (**Stratified Belief, SB**). *It is rational for agent α to believe x on a level V if and only if the probability of x is indistinguishable from unity on the context level of the agent. More formally:*

$$B(x) \in R_{\alpha,V} \quad \Leftrightarrow \quad P(x) \simeq_V 1 \tag{3.4}$$

Compared to the threshold-version of the Lockean Thesis, there is a new contextual element present here—that of a level—which will be discussed further in section 3.7.[15] Expressed in words, SB says: relative to a certain level V, it is rational for an agent to believe a proposition if and only if the probability of the proposition is ultraclose to unity as compared to the level. Other ways of formulating this condition for rational belief is that the probability should be indistinguishable from unity (on a given level) or equal to unity up to an infinitesimal (relative to that level).

Also in the context of probabilistic approaches to conditionals, infinitesimals have been applied to interpret conditionals by Adams (1966): the statement "If A then B" is read as "The conditional probability of B given A is larger than 1 minus an infinitesimal" or symbolically: $P(B|A) > 1 - \epsilon$.[16] The latter expression can also be used in default reasoning to represent the statement "Normally, if A then B (but there may be exceptions)". In the context of Adams' work, ϵ does not refer to an infinitesimal in the sense of NSA or relative analysis, but to the ϵ, δ-framework of standard analysis. However, the interpretation is quite similar to that of SB: a probability that is infinitesimally close to unity indicates 'almost certainty' (which includes full certainty).

Whereas the version of SB given in this section models rational belief as almost certainty, we will relax this condition in section 3.7.2.

3.6.2 Stratified belief applied to a large lottery

Now we can apply the definition of SB to the case of an (ultra)large lottery. For an N-ticket lottery, where N is large (as judged by a specific agent), consider level V_{lott} which contains 1 but not N.[17] In other words, on this level 1 is standard and N is ultralarge. In that case, $\frac{1}{N}$ is ultrasmall compared to 1 on this level ($N \gg_{V_{\text{lott}}} 1 \gg_{V_{\text{lott}}} \frac{1}{N}$), which shows that level V_{lott} is a good starting point for discussing the probabilities of an (ultra-)large lottery. Figure 3.3 illustrates how relative analysis helps us to understand what the number line looks like to such an agent: he only takes into account the standard numbers of the level V_{lott}. The standard numbers in the $[0, 1]$-interval are of particular interest, since, in the current view, they guide the agent in his belief-forming practices.

Since $\frac{1}{N}$ is ultrasmall, $\frac{1}{N} \simeq_{V_{\text{lott}}} 0$, we find that:

---

[15] Also in the discussion of Threshold Beliefs, it has been remarked that the threshold should be thought of as context-dependent (see *e.g.* Hawthorne and Bovens, 1999, p. 246); in SB, this consideration is taken into account as an explicit parameter.

[16] Thanks to Sonja Smets for the pointer.

[17] Before proceeding, we should check that such a level exists. Axiom 3 of Hrbacek et al. (2010) ensures that there always exists a number ϵ that is ultrasmall compared to the given level; thus, all numbers $N > \frac{1}{\epsilon}$ are sufficiently large for such a level to exist. We can give examples of values for N that are not large enough (such as $N = 2$), but because we cannot give an explicit example of ϵ,
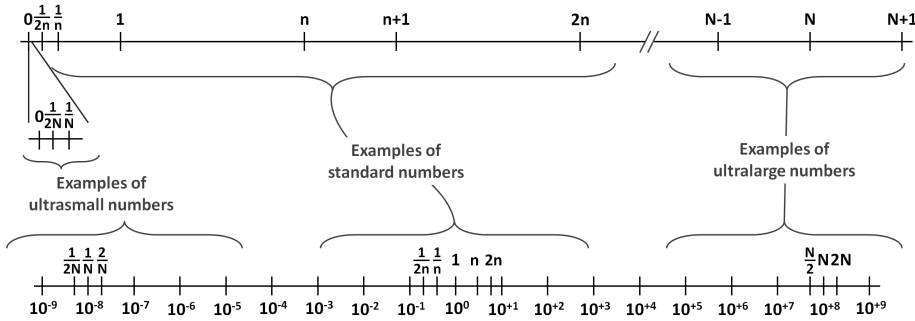
Figure 3.3: *An agent's mental picture of the real number line only contains the standard numbers of a certain context level (V_{lott}). N represents the total number of tickets in an ultralarge lottery (e.g. 10^{+8}); n is some standard number of tickets (e.g. 3). The inverse of N (e.g. 10^{-8}) is ultrasmall, whereas the inverse of n (e.g. $\frac{1}{3}$) remains standard. The lower part of the figure represents the same number line on a logarithmic scale. On such a scale, the context level extends equally wide to the left and to the right of the number 1: it shows the inversion between the ultralarge and the ultrasmall numbers as a mirror symmetry.*

$$1 - \frac{1}{N} \quad \simeq_{V_{\text{lott}}} \quad 1$$

Let us consider two groups of tickets: 1) the set of all but one lottery tickets ($N - 1$) and 2) the set of all tickets (N). The above equation formalizes that, on a certain level, an agent is not able to appreciate the difference between the winning odds of both sets of tickets.

We may interpret the same equation in terms of losing odds instead of winning odds: then it expresses that on this level, the agent also cannot distinguish between the losing odds in the situation 1) in which he owns one ticket and 2) in which he has no ticket at all. If the difference between having a ticket or not appears negligible to an agent, it is only rational for him to believe that his single ticket will not win.[18]

---

there also is no clear threshold for N being large enough—a remark that should sound familiar at this point.

[18] Because the winning odds of all but one ticket are equal to the losing odds of one ticket, on the same level they are either both distinguishable from unity or both indistinguishable from it. This may be a drawback if we want to model an optimistic person, who is less worried that he will not win if he owns all but one ticket than he is expecting to win if he owns just one ticket, and *vice versa* for a more pessimistic mind. Although this could be modeled using different levels for the two cases, or levels that are asymmetric under inversion, it is probably better to regard the psychological factors of risk seeking and risk aversion as outside the scope of what the model of stratified belief intends to capture.

3.6.3 Aggregation of stratified beliefs

The next question is whether the Conjunction Principle holds for this type of beliefs. We will see that it is allowed, but only within a level. This stratified conjunction principle (SCP) formalizes our intuitive weakening of CP given in section 3.2.

Although we started our search for a solution to the Lottery Paradox from the viewpoint that we need to adapt the formalization of the Lockean Thesis, it does turn out that we have to adapt CP too.

3.6.3.1 Aggregating beliefs concerning single lottery tickets

First we show that the conjunction of two rational beliefs (in the sense of SB), each concerning only one lottery ticket, amounts to a new rational belief.

Definition 2 (Stratified Conjunction Principle, SCP).

$$B(\psi_1) \in R_{\alpha,V} \text{ and } B(\psi_2) \in R_{\alpha,V} \quad \Rightarrow \quad B(\psi_1 \wedge \psi_2) \in R_{\alpha,V} \qquad (3.5)$$

To show that SCP indeed holds, apply equation (3.4) to the left-hand side:

$$B(\psi_1) \in R_{\alpha,V} \Rightarrow 1 - \frac{1}{N} = P(\psi_1) \simeq_V 1$$

$$B(\psi_2) \in R_{\alpha,V} \Rightarrow 1 - \frac{1}{N} = P(\psi_2) \simeq_V 1$$

and apply equation (3.4) to the right-hand side of SCP:

$$B(\psi_1 \wedge \psi_2) \in R_{\alpha,V} \Rightarrow 1 - \frac{2}{N} = P(\psi_1 \wedge \psi_2) \simeq_V 1$$

Since $1 - \frac{1}{N} \simeq_V 1 - \frac{2}{N}$, $1 - \frac{1}{N} \simeq_V 1$ guarantees that $1 - \frac{2}{N} \simeq_V 1$.

At first sight, SCP takes the same form as CP. To see that SCP is nevertheless a weakened form of CP, note that SCP does *not* generalize to the conjunction of any number of stratified beliefs. In particular, it does not hold that the conjunction of an ultralarge number of rational beliefs is rational. That is to say, $B(\psi_1) \in R_{\alpha,V}$ and ... and $B(\psi_N) \in R_{\alpha,V} \nRightarrow B(\psi_1 \wedge \ldots \wedge \psi_N) \in R_{\alpha,V}$ (because $1 - \frac{N}{N} = 0 \not\simeq_V 1$).

3.6.3.2 Generalization of SCP

Although SCP does not generalize to the conjunction of 'many' (*i.e.* an ultralarge number of) and in particular all beliefs, it does allow the conjunction of 'a few' (*i.e.* a standard number of) beliefs, or—which is equivalent—the conjunction of two beliefs, each concerning 'a few' (standard number of) tickets. As such, SCP can be considered as the formal counterpart of our intuitive weakening of CP stated in section 3.2.

Here, we will prove this slightly stronger version of SCP, that is valid for arbitrary events E_1 and E_2, not just singletons:[19]

---

[19] Thanks to Karel Hrbacek for suggesting this.

$$B\big(\psi(E_1)\big) \in R_{\alpha,V} \text{ and } B\big(\psi(E_2)\big) \in R_{\alpha,V} \Rightarrow B\big(\psi(E_1) \wedge \psi(E_2)\big) \in R_{\alpha,V}$$

The proof of this conjunction rule is as follows: by the definition of SB in equation (3.4), the two assumptions are equivalent to $P\big(\psi(E_1)\big) \simeq_V 1$ and $P\big(\psi(E_2)\big) \simeq_V 1$, respectively. Thus, $1 - \frac{\#(E_1)}{N} \simeq_V 1$ and $1 - \frac{\#(E_2)}{N} \simeq_V 1$, which imply that $\frac{\#(E_1)}{N} \simeq_V 0$ and $\frac{\#(E_2)}{N} \simeq_V 0$. If two numbers are ultraclose to zero, this is also true for their sum: $\frac{\#(E_1)}{N} + \frac{\#(E_2)}{N} \simeq_V 0$. Since $\#(E_1) + \#(E_2) \geq \#(E_1 \cup E_2)$, we find that $\frac{\#(E_1 \cup E_2)}{N} \simeq_V 0$. This implies that $1 - \frac{\#(E_1 \cup E_2)}{N} \simeq_V 1$, or in terms of probability: $P\big(\psi(E_1 \cup E_2)\big) \simeq_V 1$. Because of the definition of ψ, this is equivalent to: $P\big(\psi(E_1) \wedge \psi(E_2)\big) \simeq_V 1$. Finally, by the definition of SB, we conclude that: $B\big(\psi(E_1) \wedge \psi(E_2)\big) \in R_{\alpha,V}$.

3.6.4 Is the solution psychologically plausible?

In Figure 3.3, we have represented a person's mental picture of the number line. There is psychological evidence that people indeed use such a picture (Dehaene et al., 1999, p. 970): "Within the domain of elementary arithmetic, current cognitive models postulate at least two representational formats for number: a language-based format is used to store tables of exact arithmetic knowledge, and a language-independent representation of number magnitude, akin to a mental 'number line,' is used for quantity manipulation and approximation. ... [E]xact calculation is language-dependent, whereas approximation relies on nonverbal visuo-spatial cerebral networks." Applied to the Lottery Paradox, this finding suggests that the winning odds of a single lottery ticket in a very large lottery is represented in the brain in two different ways: one part of the brain registers that it is 'definitely different from zero', while the other part processes it as 'zero or approximately zero'.

Moreover, there is evidence that a logarithmic mental scale (as in the lower part of Figure 3.3) comes first in the human cognitive and cultural development, whereas the linear scale (represented in the upper part of Figure 3.3) is only acquired through formal education (Dehaene et al., 2008). Observe that, on a linear scale, the *absolute* error due to approximation is constant. On a logarithmic axis, however, errors due to approximation scale the same way as the quantities they apply to: in that case, the *relative* error is constant.

Even persons who have learned mathematics at school, and are thus able to think of quantities on a linear scale, still apply the log scale in situations that discourage counting (situations involving large and/or continuous quantities) (Dehaene et al., 2008). Thus, when confronted with a heap of sand, we are likely to visualize the amount of sand on a logarithmic scale. If we repeatedly remove one grain of sand from the heap, the relative difference increases as the heap becomes smaller (although the absolute difference is the same each time). When confronted with a number of grains that can easily be counted, we will use a linear scale and think in terms of the absolute difference. The discrepancy between our two types of mental number scales may at least be partially responsible for what strikes us as paradoxical in soritic cases.

Despite the experimental evidence for the approximation that occurs when humans reason about numbers of different scales of magnitude, which fits well with our model of Stratified Belief, a word of caution is also called for: SB is a very simple model. In particular, it only allows us to use one context level for one person at a given point in time. This means that the person's counting and reasoning capacities are modeled with one and the same context level. From the psychological point of view, it seems rather unrealistic that these mental abilities should be so perfectly balanced. We admit that SB is a crude model in this respect, but adding more realism to the model always comes with a price: it makes the formalism less transparent.

In a more advanced approach, we could use different context levels to indicate different mental capacities and attitudes. For instance, an additional context level could be included to reflect how much money a person has available to spend on buying lottery tickets: a large number of tickets may be considered ultralarge by a person because their price is much more than his total budget.[20]

3.7 Relation to philosophical theories and application to other problems

3.7.1 Relation to contextualism

Skepticism tells us that there is always the possibility that our whole life is an illusion and we are just brains in a vat (Putnam, 1981), or, that when we say "We will be there in one hour", the Earth will get hit by a huge meteor and we never get there. (Similar examples are considered by Harman (1986).) In daily conversation, it would be tiresome to always sum up these and similar highly unlikely events. Yet, it seems like we should, for there is no way to exclude these options with absolute certainty. The analysis in terms of levels gives a post-hoc justification of what we actually do: we usually treat the unlikely events as infinitesimals, but there is a finer level available—which may be relevant in scientific or philosophical contexts— on which even these minuscule probabilities are appreciable.[21] Moreover, we may compare the highly unlikely events: I judge the probability of my whole life being an illusion as ultrasmall, even compared to the probability of the Earth being hit by a large meteor in the next hour, which is itself ultrasmall (compared to the chance of a coin landing heads, for instance). Because we are free to use a different level in different situations, we may say to a friend that we believe that we will arrive in 60 minutes because our GPS says so, but deny that we believe that the extinction of the dinosaurs has been caused by anything other than a meteor. In both cases we use the verb "to believe", which suggests a fixed scale of belief, but apparently the scale does depend on the context.

---

[20]Thanks to Bryan Renne for this suggestion.

[21]Whereas conditionalizing on zero-probability events is problematic when using Kolmogorov's ratio formula (Hájek, 2003), it is unproblematic to conditionalize on events with a relatively infinitesimal probability, since these are just ordinary, non-zero real numbers. This is similar to the solution to the countably infinite lottery problem given in Chapter 2, where the use of NSA allows one to assign (actual) infinitesimal probabilities to single tickets.

Contextualism is "a theory according to which the truth-conditions of knowledge-ascribing and knowledge-denying sentences . . . vary in certain ways according to the context in which they are uttered" (DeRose, 2009, p. 2). The variable parameter is the epistemic standard. We hope that it is clear at this point that a rise in epistemic standard can be modeled as switching to a finer level of probability values. We will look at three examples: 1) the sand case, 2) the lottery case, and 3) the bank case, well-known in the contextualism literature and to be discussed in subsection 3.7.3.

With the sand example and the possibility of counting the individual grains, we saw that whereas some properties may be vague on one level, the region of vagueness may be shifted or removed by going to a finer level. We also indicated that such a change in context is typically triggered by a higher reward for the agent, since the refinement of the level usually requires more effort of the agent (observing for a longer time, using larger equipment, . . .).

Let us find an application of this in our lottery example: suppose a person is quick to say "I do not believe that my ticket will win." Of course, the probability of winning for any ticket is non-zero, so he can never be absolutely certain that he will not win, no matter how large the lottery is and how small the probability. Yet, there is a level at which the winning odds of a single ticket are indistinguishable from zero. In other words, the agent only makes a very small rounding error if he takes his probability of winning to be zero. Doing so may not be justified on the highest standards of rationality, but for practical purposes it seems to be good enough and certainly not completely irrational.

Of course, the agent is by no means forced to ignore the small difference of his winning odds compared to zero. Suppose the same agent notices that he has just been robbed and really needs the money. He may now be more receptive of the very small chance that he might win the lottery with his single ticket, which would solve his current financial problems. This change of context may make him reconsider his previous statement, a process that can be modeled as a transition to a finer level.

Relative analysis may also be considered in relation to contextual identity. As we have seen, Stratified Belief formalizes a probability being 'sufficiently close' to unity in LT as that probability being 'ultraclose' to 1 on a certain level ($\simeq_V 1$). In other words, it requires the probability to be indistinguishable from 1. Whether or not this is the case, depends on the context level for all values of the probability except for those *exactly* equal to 1. In the philosophical literature, there are some well-known paradoxes related to identity, such as the Ship of Theseus. It has been suggested that there is a contextual element to the identity predicate involved in these paradoxes. Some examples of the context-dependence of words as 'same' (identical) and 'different' (not identical) are discussed in Crawshay-Williams (1957, p. 22–24). Also Douven and Decock (2009) comment on the vagueness and relativity of the identity predicate. We conclude here that the relation $\simeq_V$ can be used to formalize these ideas of contextual identity.

3.7.2 Relation to the epistemicist account of vagueness

3.7.2.1 Threshold-free scales of belief

Since largeness is a vague concept, beliefs about topics in which largeness is a crucial element, such as a large lottery, 'inherit' some of this vagueness. Therefore, a model for belief about ultralarge lotteries has to be construed within a framework capable of handling vagueness, and, as we have seen, relative analysis provides such a framework. We have used it as the basis for a threshold-free model of rational belief: the model of Stratified Belief.

It may be argued that our model for Stratified Belief—and associated scales of belief—can be interpreted as having 'hidden' thresholds too. In that case, however, it should be noted that it is of crucial importance that the thresholds are implicit. Any attempt to make their value explicit makes the stratified analysis core of the model collapse back into standard analysis, which means we get back all the threshold-related problems (such as the total failure of CP).

So far, we have only discussed yes–no beliefs. Now we return to the topic of qualified belief and subjective probability, quickly passed over in the Introduction. In our language, there are various words and word combinations to express subtle differences in the strength of our beliefs. We may say: 'I suspect that', 'I believe that', or 'I am convinced that' to give only three examples in the order of firmness of belief.[22]

These various expressions could be thought of as expressing different degrees of belief. In the threshold-based model, a higher degree of belief can be be made to correspond to the probability being at least equal to a higher threshold value. In our model of Stratified Belief, we replace the degrees with the threshold-free notion of scales of belief: a higher scale of belief corresponds to the probability being indistinguishable from unity still at a finer level.[23]

3.7.2.2 Vague thresholds

We will now show how we can adapt Stratified Belief to encompass weaker forms of belief. This adapted form of SB, SB_θ, can also be used to model the epistemicist account of vagueness (Williamson, 1994, 1997), which assumes that the threshold does have a sharp value, but that this value is essentially inaccessible to us.

Let us fix an approximate threshold θ (in the interval $]\frac{1}{2}, 1]$) and define SB_θ by:[24]

---

[22] As indicated before, we do not consider the knowledge-version of the Lottery Paradox. Although 'I know that' would certainly rank as a very strong expression of firmness of belief, knowledge requires something more than firm belief.

[23] It may seem appealing to let knowledge correspond to the finest level. Axiom 3, however, states that for any level there are numbers that are ultrasmall compared to it. Therefore, for any level there is a finer level, and 'the finest level' does not exist (an illustration of incompleteness). However, in the case of a specific N-ticket lottery there are levels at which all the relevant knowledge is available: those which contain N and (thus) $\frac{1}{N}$. The point is that there is no such level with this property for *general N*.

[24] Thanks to Karel Hrbacek for this suggestion.

Definition 3 (Stratified Threshold Belief, SB$_\theta$).

$$B(x) \in R_{\alpha,V} \Leftrightarrow P(x) \gtrsim_V \theta$$

where $P(x) \gtrsim_V \theta$ means that either $P(x) > \theta$ or $P(x) \simeq_V \theta$.

Applied to an ultralarge lottery, the Stratified Conjunction Principle would remain valid for two singletons, or for an arbitrary event E_1 and a singleton (but no longer for arbitrary E_2). Observe that the initial formulation of Stratified Belief is a special case of SB$_\theta$ with $\theta = 1$.

If we compare SB$_\theta$ to the usual threshold-model, we see that although the threshold θ is a specific real number, it functions merely as an arbitrarily chosen representative of all the numbers $r \in \mathbb{R}$ such that $r \simeq_V \theta$.[25] The vagueness works in two ways: in determining what the threshold is or ought to be, and in the formation of beliefs. Therefore, SB$_\theta$ can be used to model thresholds that have no precise value, as well as thresholds that do have a precise value but which is inaccessible to us. Using the latter interpretation, SB$_\theta$ can be thought of as a formal representation of the epistemicist account of vagueness.

3.7.3 Application of Stratified Belief to similar problems

The model for Stratified Belief was motivated by cases in which information in terms of objective probabilities is explicitly available. It replaces the idea of degrees of belief with scales of belief. However, the former notion is not restricted to cases in which the objective probabilities are known, and actually fits better with cases in which they are not. In this subsection, we apply the analysis in terms of Stratified Belief to three examples in which the exact values of the objective probabilities are not known. Moreover, the first example is usually presented in terms of knowledge rather than (rational) belief. Nevertheless, stratified belief does seem to provide a sensible account of this case, too.

In the contextualism debate, there are many examples without explicit probabilities. A popular example is Keith DeRose's 'bank case': given circumstantial evidence, we may claim to know that the bank is open on Saturday if little is at stake, but when the stakes are higher, we may deny doing so. When the stakes are higher, the epistemic standard rises, and this may be modeled by using a finer level. The (implicit) probability value for the bank being open on Saturday is so close to unity that on the coarse level used in a case with low stakes, it is indistinguishable from unity. One may be aware that there is no absolute certainty, but the difference is a relative infinitesimal and therefore inappreciable: it is rational to believe or say one knows the bank to be open. When the stakes are higher, there is an incentive to reconsider the importance of the difference between the relevant probability and unity. Using a finer level, a previously infinitesimal difference becomes appreciable and it becomes rational to say that one does not know that the bank will be open.

---

[25]Note that these numbers—like a level—do not form a set. In the context of extension-style NSA, this would be called the 'halo' of the number θ. Here it is again a predicate, not a set, and level-dependent.

In the contextualism literature, there are more complicated examples (with different speakers as well different contexts), but because the second axiom of relative analysis ensures that it can always be established which of two levels is at least as fine as the other, these cases could be analyzed in terms of levels, too.

A problem similar to the Lottery Paradox arises in thinking about elections: it is rational for anyone to believe that the impact of his or her single vote on the result of the elections is negligible, because it is only one out of an (ultra-)large number of votes. However, the combined impact of all individually negligible votes is not negligible at all. This is called the paradox of (non-)voting (Owen and Grofman, 1984). Clearly, the same analysis given here for the ultralarge lottery can be applied to the paradox of voting: the paradox can be blocked by noting that the unrestricted application of mathematical induction would be unreasonable.

3.8   Conclusions

We want to model the formation of rational beliefs related to probabilistic information, expressed as real numbers, but our mental capabilities are finite and do not allow us to form beliefs with such infinite precision. This means that even if we recognize that some real-valued probabilities are different, we cannot always form distinct beliefs based on those numbers. If we think of the integers on a number line, we imagine them as clearly and evenly separated. If we think of the rational numbers or the real numbers, we may imagine the position of some simple fractions ($\frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \ldots$) and some well-known irrational numbers ($e, \pi, 2\pi, \ldots$). However, if we try to focus on one specific number, zero for instance, we have to admit that the position we think of as the position of this number is not precise enough and actually contains a whole cloud of numbers, all very close to zero. In other words, infinitely many numbers that are quantitatively different from zero are nevertheless indistinguishable from it in our mental image of the real line. The same is true for any other real number that we may want to consider. Unlike $\mathbb{N}$, $\mathbb{Q}$ and $\mathbb{R}$ are dense sets. If we try to form ourselves a mental picture of these numbers on a line, it is as if we look at them through frosted glass.

Imposing levels on the standard set of real numbers, as relative analysis does, is a way to model what we see through the glass. We do have the capability of focusing on a region of interest, so as to be able to distinguish more numbers in that region. In relative analysis, this is modeled as a transition to a finer context level (or a different choice for the unit). Although there is no limitation on the amount of 'zoom', at each point in time the zoom is limited to some finite factor and we can never look at the real numbers directly, in all their infinite depth; we cannot look behind the frosted glass.

Our search for a model of rational beliefs based on probabilistic information was motivated by the Lottery Paradox. We observed that vagueness is an essential ingredient to that paradox, that there is a contextual aspect to this vagueness, and that a similar problem exists for infinite lotteries. We also observed that we expect a weakened form of the Conjunction Principle to hold for rational beliefs. The com-

bination of these observations led us to the application of relative analysis. Based on the relative analysis framework, we formulated a soritic, contextual version of the Lockean Thesis. This led us to a new definition of rational belief as 'almost certainty' (including absolute certainty), which we called Stratified Belief.

We also investigated the aggregation of this type of beliefs. Kyburg's own response to the Lottery Paradox was to abandon CP (Kyburg, 1961), whereas many later authors have tried to rescue (part of) it (Wheeler, 2007). We found that a weakened version of CP indeed holds for stratified beliefs. Because the aggregation is restricted to 'a few' (a standard number of) beliefs, the Lottery Paradox does no longer occur. Based on the lottery example, we may compare aggregating beliefs to doing a calculation based on rounded values: it is better to avoid this, but if the rounded values are all you have, some calculations still give reasonably good outcomes. All of this can be stated more rigorously using the language of relative analysis.

One of the observations that led to our solution, was the analogy with an infinite lottery. In an infinite lottery, it is possible that a specific ticket will win, but the real-valued probability assigned to this possibility is zero, exactly the same as for the impossible event. This leads to some counter-intuitive results (such as the failure of Countable Additivity), which can be blocked using NSA, by allowing infinitesimals as the value of the probability function. In a finite lottery, no matter how large, the probability of any single ticket is strictly larger than zero. However, in our mental representation of it and our resulting rational beliefs, we may not always be able to distinguish between an event with a very small probability and the impossible event. This leads to the counter-intuitive result that even the conjunction of two such rational beliefs is not guaranteed to be rational. Again, NSA can be applied to solve the puzzle: using relative analysis, probabilities that are indistinguishable from zero can be modeled as ultrasmall numbers or relative infinitesimals. As was already mentioned, the concept of beliefs based on this framework does allow the conjunction of, for example, two beliefs.

In short, by looking at the Lockean Thesis from the viewpoint of relative analysis, we have found a new way to incorporate its inherent soritic and contextual nature in the formalization. This leads to a definition for and aggregation rule of rational beliefs, whose combination does not lead to a Lottery Paradox. However, we do not regard dealing with the Lottery Paradox as the end goal of this chapter. It serves merely as a case study, prompting reflection about the presence of vagueness in our statements and judgments concerning topics about which we have precise probabilistic information. It is our conviction that the proposed model of Stratified Belief can be a useful tool for epistemology in general. In particular, the model matches well with various contextualist approaches—in epistemology as well as metaphysics—that have appeared in the last few decades.

Apart from this application to rational beliefs, relative analysis may have a further role to play in formal philosophy: it is a powerful mathematical system capable of dealing with vagueness and its contextual aspect. To what extent this is a contribution to the philosophy of vagueness in general may be worth further reflection.

Chapter ⚃

Ultralarge and Infinite Lotteries

Our minds are finite, and yet even in these circumstances of finitude we are surrounded by possibilities that are infinite, and the purpose of human life is to grasp as much as we can out of the infinitude.
Alfred North Whitehead (Price, 1954, p.160)

'Can you do Addition?' the White Queen asked. 'What's one and one and one and one and one and one and one and one and one and one?'
'I don't know,' said Alice. 'I lost count.'
Lewis Carroll (1871, Chapter IX)

One, two, three . . . infinity.

George Gamow (1947)

4.1   Introduction

This chapter is intended to summarize the findings of Chapters 2 and 3, to emphasize the analogy between the problems and the solutions discussed there, and to develop the epistemology of an infinite lottery.

The problem of describing a fair infinite lottery (subsection 4.1.1) and the Lottery Paradox (subsection 4.1.2) are usually considered to be unrelated issues, appearing in different subdisciplines of philosophy. In this chapter, however, we will emphasize the analogy between them. This approach allows us to diagnose them as suffering from an 'adding problem', which is in both cases due to the accumulation of rounding errors (section 4.2). Because the diagnosis is the same, we can proscribe the same

treatment, too: we will achieve this with the help of infinitesimals, which are available in non-standard analysis (NSA) (section 4.3). For each of the two cases, we can select an approach to NSA that suits the application best. In section 4.4, we will combine our findings to analyze the epistemology of an infinite lottery.

4.1.1 Case 1: Failure of Countable Additivity in an infinite lottery

In standard probability theory, with the axioms as introduced by Kolmogorov (1933), it is not possible to describe a fair lottery on a countably infinite sample space.

To see this, consider $\mathbb{N}$ as the sample space (usually denoted by Ω). The axiom of Normalization requires that the probability of the full set of tickets equals one. To model a fair lottery means that all the tickets have the same probability. If this is so, the probability of a single ticket cannot be any number larger than zero, for then the infinite sum ranging over all the tickets diverges, and either Normalization or Countable Additivity (CA) fails. However, the probability cannot be zero either, because then the infinite sum of all the individual probability values evaluates to zero, which is not equal to one as demanded by Normalization and CA. Hence, we cannot assign *any* probability to the individual tickets: Kolmogorov's approach to probability theory cannot describe a fair infinite lottery.

It is puzzling that the formalism is not flexible enough to deal with this case, in particular since a countably infinite lottery corresponds to a sample space of the lowest infinite cardinality.

Various solutions have been proposed to solve this problem:

(1) Give up the property of Countable Additivity.

(2) Give up the axiom of Normalization.

(3) Deny that there exists a fair lottery on $\mathbb{N}$.

(4) Assign an infinitesimal probability to the single tickets, rather than zero.

The most prominent advocate of option (1) was de Finetti (1974), who argued to replace CA by the weaker requirement of Finite Additivity. Solution (2) was proposed by Rényi (1955). If we hold on to Kolmogorov's system, including Normalization and CA, we arrive at option (3): probability measures on countable domains are always biased. This solution has some strange consequences, to say the least (see for instance Kelly, 1996, p. 321–323).

We find solution (4), advocated by Lewis and Skyrms (1980), to be the most plausible one. It is not immediately clear which changes have to be made in Kolmogorov's axioms in order to realize this solution. A detailed analysis (as the one presented in Chapter 2) shows that not only has the range of the probability function to be adapted (from the standard unit interval of the real numbers to a non-standard set), but that even then the property of Countable Additivity cannot hold (although it can be replaced by another form of infinite additivity). Hence, it turns out that (1) and (4) are not completely independent solutions after all.

4.1.2 Case 2: Failure of Conjunction Principle for rational beliefs concerning a large, finite lottery

The probability assignment for a fair, finite N-ticket lottery with exactly one winner is trivial: every ticket has a probability of winning equal to $1/N$. From an epistemological viewpoint, however, the case where $N \gg 1$ is problematic: a paradox can be obtained for rational beliefs about such a lottery.

Consider a very large, but finite fair lottery. If you own just one ticket, it may seem rational for you to say: "My single ticket will not win." Suppose that you receive an additional ticket for the same lottery. It may still be rational for you to say: "My two tickets will not win." Likewise for three tickets, and so on. If you generalize the previous statement to "Likewise for all (N) tickets", you obtain a paradox, for you know the lottery to have a winner for certain.

This Lottery Paradox was first constructed by Kyburg (1961), who concluded that the Conjunction Principle (CP) does not hold for rational acceptability. Many later authors have tried to rescue at least part of CP (Wheeler, 2007).

Although we are in favor of a formal approach to epistemology, the popular threshold-based model for rational beliefs—which states that it is rational to believe a statement if the probability of that statement is larger than some positive threshold smaller than one—does not seem adequate. In particular, it does not do justice to the intrinsic vagueness of both the Lockean thesis—the thesis stating that it is rational to believe a statement if the probability of that statement is 'sufficiently close to unity'—and the Lottery Paradox itself, which, after all, only occurs in cases in which the number of tickets is 'sufficiently large'. The phrasings 'sufficiently close to unity' and 'sufficiently large' are strongly suggestive of a vague and contextual element, both of which are absent in threshold-based models.

Intuitively, we would expect a weakened version of CP to hold, allowing us to aggregate at least 'a few'—again a vague and context-dependent term—rational beliefs. Applied to the example of the Lottery Paradox, the maximal generalization that can rationally be obtained is: "Any set containing a few ticket will not win"—not 'many', in particular not 'all'. In the threshold-based model, however, CP fails and there is no weakened version of it available either.

4.2 The analogy between the two cases

Though appearing in different contexts, both cases have common characteristics:

(a) They both involve the assignment of fair probabilities to a denumerable lottery.

(b) They deal with either a 'sufficiently large' or an 'infinite' sample space, two concepts that are intimately related.

(c) They both involve a vague distinction.

(d) They both involve a limit process.

(e) They both pose an adding problem.

Whereas characteristic (a) is immediately clear, we will now provide further motivations for the statements (b)–(e).

Characteristic (b) claims that the properties of being finite but sufficiently large and of being infinitely large are intimately related. This is not on original claim: Lavine (1995) has argued that our whole concept of infinity has to have originated, somehow, from our experience with finite quantities (since our senses do not allow us to experience any infinite quantity). In particular, Lavine sees 'infinity' as a way of dealing with the 'indefinitely large' numbers. Although the former has the benefit of being a context-independent concept, it is not strictly necessary, and Lavine (1995) argues to practice mathematics only in terms of the latter, finite concept. Those who do not endorse a finitistic attitude towards mathematics may employ Lavine's claim for the analogy between 'indefinitely large' and 'infinite' in both directions: whereas a generalization occurs from the finite to the infinite realm in the development of NSA, we will apply NSA to the infinite lottery first and use this result as an inspiration for dealing with the Lottery Paradox, involving a large but only finite domain.

The presence of vague distinctions (c) has already been demonstrated for the Lottery Paradox, but may seem an unwarranted claim when considering the infinite lottery problem. Admittedly, the claim there is of a more subtle nature: it says that the distinction between finite and infinite is vague. As we will see, in NSA there are infinite hypernatural numbers, which are larger than any finite natural number. Just like there is no last (*i.e.* largest) finite natural number, there is no first (*i.e.* smallest) infinite hypernatural one. Hence, the distinction between finite and infinite, clear as it may seem, does not amount to a sharp threshold between the two, but a transition that can only be indicated by points of ellipsis.

An ellipsis often indicates a type of limit process, which brings us to (d). The limit process involved in the infinite lottery seems clear: we want to idealize a fair lottery on N as N becomes larger, in such a way that the answer is no longer dependent on how large N is precisely. Kolmogorov's framework—in particular, his Axiom of Continuity, which has CA as a consequence—implies using a classical limit, which is always defined on the real numbers. However, this is not the only limit operation mathematics has to offer. As we will see in the next section, one approach to NSA has a limit operation available that evaluates to an infinitesimal rather than zero, for the winning odds of a single ticket in an infinite lottery. Likewise, we may interpret another approach to NSA as blocking the generalization of the statements in the Lottery Paradox to a statement about 'many' (in particular, 'all') tickets, while allowing the corresponding statement in terms of 'a few' tickets.

The adding problem referred to in (e) refers to the failure of CA on the one hand and to the failure of CP, even in a weakened form, on the other. We will now show that the problem seems to be due to the accumulation of rounding errors in both cases. To see this, all we need is a basic understanding of error propagation. For instance, if two or more rounded values are added, the total error on the sum is equal to the sum of the errors on the terms. On its own, a rounding error it just that: one small error. However, in a sum with many terms, all of which have a small error associated with them, the total error is of the form 'many $\times$ small', which is not guaranteed to be small.

In the next section, we will investigate whether we can employ the analogies observed here to resolve the two issues with the same solution strategy.

4.3 Solution using non-standard analysis

The solution we will pursue is to represent the aforementioned rounding errors by infinitesimals. Whereas it has long been thought that infinitesimals are an inconsistent concept, Robinson (1966) has developed NSA as a framework that allows us to deal with infinitesimals—as well as infinite numbers—consistently.

4.3.1 Infinite additivity for a fair lottery on $\mathbb{N}$

Because the infinite lottery case seems the most natural one to apply NSA to, we will consider this first. In fact, there are multiple approaches to NSA that can be followed here. A model in terms of free ultrafilters can be constructed, as is done in Chapter 2. One may also start from the numerosity theory of Benci and Di Nasso (2003b). Since this approach starts from an axiomatic basis, it allows us to stress the concepts behind the theory rather than the technical details, as seems appropriate within the scope of a relatively short chapter.

Numerosity offers a way to assign sizes to sets such that a set is guaranteed to have a larger size than any of its proper subsets (Benci and Di Nasso, 2003b). This entails that sets of equal cardinality may have different numerosities, whereas the opposite does not hold. In particular, numerosities provide a way of distinguishing the size of infinite subsets within the natural numbers. In order to turn the numerosity function into a probability function, all that remains to be done is to take care of Normalization. Because numerosities are hyperreal numbers, they are closed under division and, thus, Normalization poses no difficulty. The numerosity of the full set of natural numbers is denoted by α (which is an infinite hypernatural number). Hence, we divide the numerosity of a subset of $\mathbb{N}$ by α to define a probability function that is defined on the whole power set of $\mathbb{N}$. Because the numerosity of a singleton is 1, the probability of a single ticket in a fair lottery on $\mathbb{N}$ is $1/\alpha$, an infinitesimal hyperrational number.[1]

What this approach amounts to in terms of a limit operation is the alpha-limit, which can be considered as a finer limit than the classical limit of standard analysis: sequences with equal classical limits, may have infinitesimally different alpha-limits. In particular, whereas the empty set and any finite subset of $\mathbb{N}$ both result in zero as the limiting probability in the classical approach (called asymptotic density, see *e.g.* Schurz and Leitgeb, 2008), only the finite set has probability zero in terms of the alpha-limit. Thus, for the infinite lottery, NSA makes it possible to assign infinitesimal probabilities to each ticket, which in turn allows us to have an infinite additivity property—though not CA (for details, see Chapter 2).

---

[1]A hyperrational number is a fraction of hypernatural numbers. The set of hyperrational numbers form a proper subset of the hyperreal numbers.

4.3.2 Conjunction Principle within a level for Stratified Beliefs concerning an ultralarge lottery

For the finite lottery, the application of NSA is slightly more indirect. After all, it is not the probability function itself that is unable to distinguish between the winning odds of an almost empty set of tickets and that of the empty set. The problem seems to arise only when we start to form beliefs based on those probability values. We propose a threshold-free model for beliefs based on an alternative formulation of NSA, called relative or stratified analysis (Hrbacek, 2007). We derive a rule that tells us when CP holds and when it is violated: it is allowed within a level, which is a concept from relative analysis intended to model the mesoscopic scale of magnitude. These levels always contain only a finite number of numbers, though some may be 'finer'—*i.e.* contain more numbers, both larger and smaller ones. A level always contains the number 1 and all quantities that are neither too small nor too large to be observable (depending on the context), but these quantities are not known with infinite precision. Real numbers that are larger than any number of a certain level are relatively infinite or ultralarge numbers compared to that level; positive real numbers that are smaller than any non-zero number belonging to the level are relative infinitesimals or ultrasmall numbers. In our model, the probability of winning for a single ticket in a fair, ultralarge lottery is a relative infinitesimal; this means that it is indistinguishable from zero (on that level). Details of the solution may be found in Chapter 3.

One may wonder whether it is really rational to believe that some non-empty set of possible outcomes will not occur, in exactly the same way as one believes that the impossible event will not occur. The answer is 'it depends', which is precisely the reason for the context-dependent element of levels in the model. If an agent is interested in, for instance, comparing the winning odds of someone owning a small set of tickets to that of someone who has no ticket at all, the very nature of the question is aimed at the difference between a small non-zero and a zero probability. In such a case, it is not likely that the agent will ignore the non-zero probability of a non-empty set. However, it may be rational for the agent to approximate an almost impossible event as having zero probability when comparing it to a certain or almost certain event. In that case, the error introduced by the approximation is of no significance to the conclusion.

The role of a level is similar to that of quantifier domain restriction, known from the philosophy of language.[2] Typically, when we refer to 'everbody' or 'everywhere', we do not mean 'all the people on Earth' or 'the whole Universe', but rather a more specific subgroup of people or a certain region on Earth, respectively. Clearly, the meaning of such a quantifier depends on the context in which the statement is uttered. Likewise, a level makes it possible to quantify only over the numbers that are observable in a given context.

---

[2]Thanks to Lorenz Demey for this suggestion. For a survey of quantifier domain restriction, consult for instance Stanley and Szabó (2000). The relation between domain restriction and vagueness is explored by Pagin (2010).

4.3.3 Visual summary of the solution

Figure 4.1 visualizes the analogy between the two cases. Figure 4.1(a) shows how the hyperrational probability values assigned to a fair lottery on $\mathbb{N}$ correspond to the real-valued approximation of (generalized) asymptotic density (as discussed by Schurz and Leitgeb, 2008). Whereas the former are fully additive, the latter is only finitely additive. Figure 4.1(b) shows why the full additivity of probabilities does not necessarily lead to full 'additivity' (CP, in this case) of rational beliefs based on precise knowledge of the probability function: in a context in which the total number of lottery tickets is considered to be ultralarge, the probability of one (or a few) tickets is ultrasmall and rounded to zero.

4.4 The epistemology of an infinite lottery

In this final section, we will illustrate how we can combine the two approaches to describe rational beliefs about an infinite lottery. In that case, two rounding processes occur: one in Step 2 and another in Step 3.

Step 1. If we want to determine the winning odds of a specific ticket or set of tickets in a fair, countably infinite lottery, we may express this probability as a hyperrational number, in particular, as a normalized numerosity. This solution is exact (although there is some freedom in the choice of the non-standard model).

Step 2. If we want to apply our definition of Stratified Belief to a hyperrational probability value, we cannot do so directly, for it works only on real-valued probabilities. Taking the standard part of the outcome, however, results in a real-valued approximation to the probability. In other words, the value is rounded up to an infinitesimal a first time.

Step 3. Subsequently, we can apply the usual definition of Stratified Belief on this already rounded probability value, which results in an additional rounding, this time up to a *relative* infinitesimal.

According to this solution, it is equally rational to believe that your tickets will not win if you own any finite number of tickets in a fair infinite lottery, as it is rational to do so for a finite lottery in which you own 'a few' tickets. Admittedly, the probability of winning is not zero in either case, but beliefs are modeled here as almost certainty, which may be completely false in some (*i.e.* an infinitesimal fraction) of all cases.
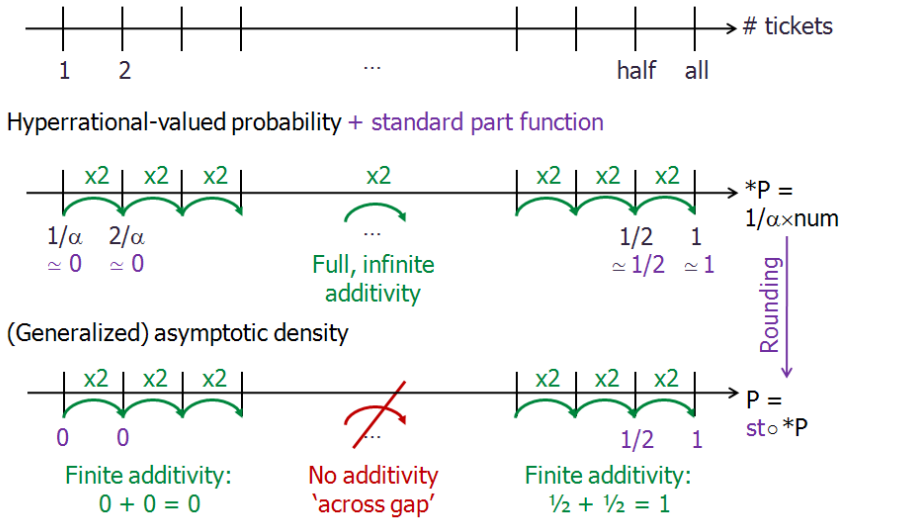
(a) Infinite lottery: hyperrational and real probability.
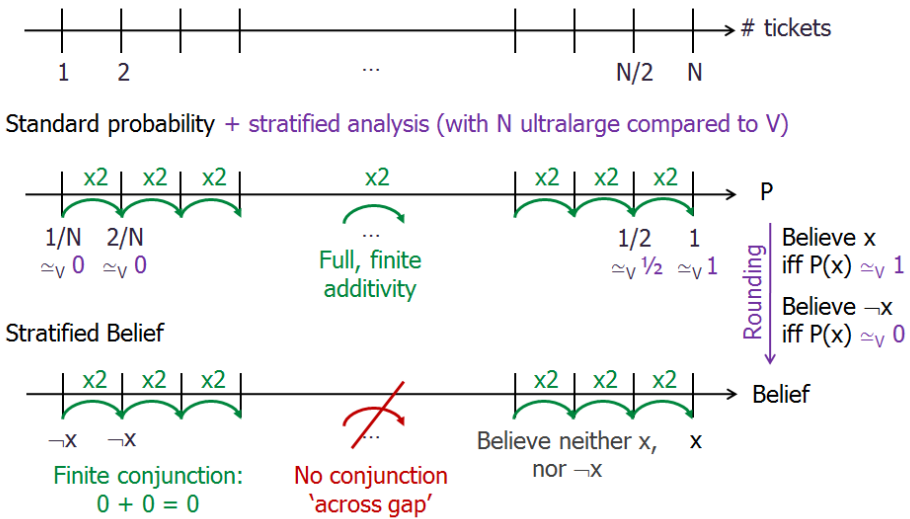
(b) Finite lottery: real probability and Stratified Beliefs.

Figure 4.1: *Illustration of the analogy between (a) an infinite lottery and (b) a large but finite lottery. Before rounding off the (relative) infinitesimals, the probability functions are fully additive. For the rounded values, an appropriately weakened additivity rule applies.*

Chapter ⚄

Determining the Probability of Inconsistencies in Theory Updating under Bounded Confidence

I refer to the theoretical assessment of probabilities concerning the future as psychohistory.

Isaac Asimov (1988)

Go not to the elves for counsel, for they will say both yes and no.

J. R. R. Tolkien (1954, p. 111)

He thinks by infection, catching an opinion like a cold.

John Ruskin

5.1 Introduction

So far, we have discussed the foundations of probability theory and epistemological issues closely associated with it. In the current chapter, we will *apply* probability theory to solve a question in formal epistemology. We aim to contribute to the blossoming field of opinion dynamics, which investigates groups of epistemically interacting agents in an artificial society, in particular how the opinions or belief states of the agents evolve through time as a result of their interactions.

Analytical sociology aims at describing and understanding social patterns and processes. An important methodology to achieve this goal is the 'structural individualism' approach, which looks for the explanation of social mechanisms at the

level of individual agents: their characteristics, behavior, and interactions. To quote Hedström and Bearman (2009, p. 13):

> In order to understand collective dynamics we must study the collectivity as a whole, but we must not study it *as* a collective entity. Only by taking into account the individual entities, and most critically the relations between them and their activities, can we understand the macro structure we observe.

On this view, macro-level phenomena should be studied in terms of models that apply to the micro-level: it is a bottom-up approach.

The use of agent-based models has become increasingly popular in social dynamics research (Macy and Willer, 2002, Macy and Flache, 2009). The branch of social dynamics we are interested in here is opinion dynamics, which investigates groups of epistemically interacting agents. There is evidence from psychological experiments that agents adjust their opinion when they are informed about the opinion of another agent (see for instance Wood, 2000). Partly inspired by such accounts, opinion dynamics studies how the opinions or belief states of agents evolve over time as a result of interactions with other agents, which may lead to cluster formation, polarization, or consensus on the macro-level.

Social psychologists have also investigated whether groups are better at finding the correct answer than individual agents. Their results depend on the precise context, in particular the type of task. When the task is to estimate a quantity, aggregating individual estimates may lead to a result that is closer to the truth than any of the individual guesses; this effect is called 'wisdom of the crowds' (Surowiecki, 2004, Douven and Kelp). When the task involves an intellective problem—in particular, one that has a demonstrably correct answer—groups of interacting agents perform better than the best agent among an equivalent number of agents who work individually; this type of cooperative problem solving is called 'collective induction' (Laughlin et al., 2002).

In general, the processes involved in opinion dynamics are very complex. Apart from analytical results, also computer simulations of agent-based models are used to study these large, complex systems. Simulations allow researchers to perform pseudo-experiments in situations in which real-life experiments are impossible, unpractical, or unethical to perform. This gives rise to the relatively young field of 'computational sociology' (Macy and Willer, 2002).

Contributions to opinion dynamics have come from mathematicians, physicists, economists, philosophers, computer scientists, and social psychologists: so far, not many sociologists have been involved. Indeed, many seminal contributions have been published in physics journals (for instance Sznajd-Weron and Sznajd, 2000, Galam, 2002, Lorenz, 2005). Although the methodology of opinion dynamics is akin to that of statistical physics, the application is clearly sociological. We hope that our contribution will help to familiarize sociologists with this interdisciplinary field.[1]

---

[1]Maybe the situation is changing already, as in 2010 a PhD-dissertation in sociology was devoted entirely to the topic of opinion dynamics (Mäs, 2010).

In most approaches to opinion dynamics, an agent's belief state is modeled as an opinion on either a single issue or multiple unrelated issues. We propose a model for opinion dynamics in which an agent's belief state consists of multiple interrelated beliefs. Because of this interconnectedness of the agent's beliefs, his belief state may be *inconsistent*. Our goal is to study the probability that an agent ends up with an inconsistent belief state. In order to explain how the beliefs are interrelated, how the agents revise or update their belief state, and how this leads to the possibility of inconsistency, we briefly review six aspects of our model: the content of an agent's opinion, the update rule according to which agents adjust their own opinion upon interaction with others, aspects related to the opinion profile, the time parameter, the group size, and the main research question. The notions introduced in the current section will receive a more formal treatment further on.

We compare our model with earlier approaches, in particular with the arguably best-known model for studying opinion dynamics, to wit, the Hegselmann–Krause (HK) model (Hegselmann and Krause, 2002). In this model, the agents are trying to determine the value of an unspecified parameter and only hold one belief about this issue at any point in time. In the most basic version of the HK model, an agent updates his belief over time by averaging the beliefs of all those agents who are are within his 'bound of confidence'; that is, the agents whose beliefs are not too distant from his own. The model we present can be regarded as an extension of the HK model.

5.1.1 Content of an agent's opinion

Agent-based models of opinion dynamics come in two flavors: there are discrete and continuous models. In discrete models (such as Sznajd-Weron and Sznajd, 2000), an agent's belief is expressed as a bit, 0 or 1 (*cf.* Ising spin model in physics). The opinion may represent whether or not an agent believes a certain rumor, whether or not he is in favor of a specific proposal, whether or not he intends to buy a particular product, or which party the agent intends to vote for in a two-party system. This binary system can be generalized into discrete models that allow for more than two belief states (*cf.* multi-spin or Potts spin), which makes it possible to model multiple attitudes towards a single alternative or to represent preferences among multiple options (*e.g.* Bernardes et al., 2001, Stauffer, 2002). In continuous models (such as Deffuant et al., 2000, Hegselmann and Krause, 2002), the agents each hold a belief expressed as a real number between 0 and 1. This may be used as a more fine-grained version of the discrete models: to represent the agent's attitude towards a proposal, a political party, or the like. In such models, values below 0.5 represent negative attitudes and values above 0.5 are positive attitudes. Alternatively, the continuous parameter may be employed to represent an agent's estimation of a current value or a future trend.

These models can be made more realistic by taking into account sociological and psychological considerations. For instance, they may be extended in a straightforward manner to describe agents who hold beliefs on multiple, independent topics (such as economic and personal issues, see Sznajd-Weron and Sznajd, 2005). As such, the

models can account for the observation that agents who have similar views on one issue (for instance, taste in music) are more likely to talk about other matters as well (for instance, politics) and thus to influence each other's opinion on these unrelated matters.[2] Nevertheless, it has been pointed out in the literature that these models are limited in a number of important respects, at least if they are to inform us about how real groups of agents interact with one another (Douven, 2010, Douven and Riegler, 2010). One unrealistic feature of the current models is that the agents only hold independent beliefs, whereas real agents normally have much richer belief states, containing not only numerous beliefs about possibly very different matters, but also beliefs that are logically interconnected.

In the discrete model that we propose, the belief states of the agents no longer consist of independent beliefs; they consist of *theories* formulated in a propositional language (as will be explained in Section 5.2.1). We will show that this extension comes at a cost. Given that the agents in earlier models hold only a single belief, or multiple, unrelated beliefs, their belief states are automatically self-consistent. This is not true for our model: some belief states consisting of interrelated beliefs are inconsistent.

5.1.2 Update rule for opinions

The update rule specifies how an agent revises his opinion from one point in time to the next. A popular approach is to introduce a *bound of confidence*. This notion—which is also called 'limited persuasion'—was developed first for continuous models, in particular the HK model (Hegselmann and Krause, 2002),[3] and was later applied to discrete models as well (Stauffer, 2002). Moreover, the idea of bounded confidence can be extended to update rules for belief states which are theories: such an HK-like update rule will be incorporated into our current model.

There is some empirical evidence for models involving bounded confidence. In a psychological experiment, Byrne (1961) found that when an agent interacts with another agent, the experience has a higher chance of being rewarding and thus oft leading to a positive relationship between the two when their attitudes are similar, as compared to when their attitudes differ. According to this 'Similarity Attraction Paradigm', in future contacts, people tend to interact more with people who hold opinions similar to their own. Despite this evidence, some readers may not regard updating under bounded confidence as a natural way for individuals to adjust their opinions in spontaneous, face-to-face meetings. Those readers may regard the agents as experts who act as consultants in a Delphi study.[4] In such a setting, the agents do

---

[2]This has been implemented for continuous opinions by Fortunato et al. (2005), Jacobmeier (2005), Pluchino et al. (2006), Lorenz (2008), and for discrete opinions by Sznajd-Weron and Sznajd (2005).

[3]See also Hegselmann and Krause (2005), Hegselmann and Krause (2006), and Hegselmann and Krause (2009). For related approaches, see Deffuant et al. (2000), Weisbuch et al. (2002), and Ramirez-Cano and Pitt (2006). This approach was partly inspired by DeGroot (1974) and Lehrer and Wagner (1981).

[4]A Delphi study is a survey method typically applied to forecast future trends, to estimate the value of a parameter, or to come to a consensus regarding a decision. Multiple experts answer a

not interact directly, but get feedback on each other's opinions only via a facilitator. When the facilitator informs each expert only of the opinion of those other experts that are within their bound of confidence, an HK-like update rule seems to apply naturally.

5.1.3 Opinion profile

An opinion profile is a way to keep track of how many agents hold which opinion. This can be done by keeping a list of names of the agents and writing each agent's current opinion behind his name. An anonymous opinion profile can be obtained by keeping a list of possible opinions and tallying how many agents currently hold a opinion; we will employ the latter type of profile. Opinion dynamics can be defined as the study of the temporal evolution of opinion profiles.

5.1.4 Time

Many studies in opinion dynamics investigate the evolution of opinion profiles in the long run. Usually, a fixed point or equilibrium state is reached. Hegselmann and Krause (2002), for instance, investigate whether iterated updating will ultimately lead groups of agents to full or partial consensus. Mäs (2010) also investigates consensus-versus cluster-formation, as a function of the sociological make-up of the group under consideration.

For sociologists, the behavior of opinion profiles at intermediate time steps may be more relevant than its asymptotic behavior. Research on voting behavior, for example, should focus on intermediate time steps (Bernardes et al., 2002); after all, elections take place at a set date, whether or not the opinion profile of the population has stabilized at that point in time.

In our study, we calculate the probability that an agent comes to hold an inconsistent opinion by updating. We do not investigate the mid- or long-term evolution of the opinion profile, but focus on the opinion profiles resulting from the very first update. In other words, we consider the opinion profile at only two points in time: the initial profile and the profile resulting from one round of updates.

5.1.5 Group size

Another interesting parameter to investigate in opinion dynamics is the group size. We are interested in updates which lead to inconsistent opinions, which may occur already for groups as small as three agents (see Section 5.2.4 below). The social

---

questionnaire separately from each other. There are additional rounds in which the experts receive additional information, including feedback on the answers gathered in the previous rounds. It has been observed that the experts tend to adjust their initial opinion in the direction of a consensus. The Delphi method was initiated by the RAND corporation for a study commissioned by the US Air Force (Dalkey and Helmer, 1963): the goal of this first published Delphi study was to estimate the number of bombs required to realize a specified amount of damage. Currently, Delphi studies are particularly popular in health care and nursing; see Keeney et al. (2011) for a review of applications in that context.

brain hypothesis of Hill and Dunbar (2003) states that 150 relations is the maximum people can entertain on average: Lorenz (2008) presents this as an argument to model groups of agents of about this size. Whereas this figure seems moderate from the sociological point of view, this is not necessarily the case from a mathematical viewpoint. As observed in Lorenz (2008, p. 323), "[c]omplexity arises with finite but huge numbers of agents." Therefore, opinion dynamics is often studied in the limit of infinitely many agents, which makes it possible to express the equations in terms of 'density of agents'. We will not do this in our current study: because of the previous observations, we should at least investigate the interval of 3 up to 150 agents.

5.1.6 Research question

As we have remarked, the agents in our model may end up in an inconsistent belief state, even when all agents start out holding a consistent theory. The main question to be answered in this chapter is: how *likely* is it that this possibility will materialize? More exactly, we want to know what the probability is that an agent will update to an inconsistent belief state and how this probability depends on the number of atomic sentences in the agents' language and on the size of their community. To this end, an analytical expression is given and evaluated numerically, both exactly and using statistical sampling. It is shown that, in our model, an agent always has a probability of less than 2 % of ending up in an inconsistent belief state. Moreover, this probability can be made arbitrarily small by increasing the number of atomic sentences or by increasing the size of the community.

5.2 Preliminaries

In this section, we first present the logical framework we assume throughout this chapter. Then we specify the representation of the opinion profile and the employed update rule. Finally, we relate our work to previous research on judgment aggregation and the so-called discursive dilemma.

5.2.1 Logical framework

5.2.1.1 Language and consequence relation

The agents in our model will have to judge a number of independent issues; we use the variable M for this number (where $M \in \mathbb{N}^+$). Throughout this section, we will illustrate our definitions for the case in which $M = 2$, the easiest non-trivial example. Each issue is represented by an atomic sentence. If the agents are bankers, the issues may be investment proposals; if they are scientists, the issues may be research hypotheses. As an example, one atomic sentence could be 'It will rain tomorrow', and another 'We should invest in the food industry'. Atomic sentences can be combined using three logical connectives: 'and', 'or', and 'not'. The collection of sentences that can be composed in this way is called the language L.

We assume a classical consequence relation for the language, which, following standard practice, we denote by the symbol $\vdash$. If A is a subset of the language (a set of sentences) and a is an element of the language (a particular sentence), then $A \vdash a$ expresses that 'a is a logical consequence of A'. That the consequence relation is classical, means that it obeys the following three conditions: (1) if $a \in A$ then $A \vdash a$; (2) if $A \vdash a$ and $A \subseteq B$ then $B \vdash a$; and (3) if $A \vdash a$ and for all $b \in A$ it holds that $B \vdash b$, then $B \vdash a$. Semantically speaking, that a is a logical consequence of A means that, necessarily, if all the sentences in A are true, then so is a.

5.2.1.2 Possible worlds

If we were to know which of the atomic sentences are true in the world and which are false, we would know exactly what the world is like (at least as far as is expressible in our language, which is restricted to a finite number of aspects of the world). The point is that our agents do not know what the world is like. Any possible combination of true–false assignments to all of the atomic sentences is a way the world may be, called a *possible world*.

Formally, a possible world is a subset of the atomic sentences. Therefore, a language with M atomic sentences allows us to distinguish between $w_{\max} = 2^M$ possible worlds: there is exactly one possible world in which all sentences are true; there are M possible worlds in which all but one of the sentences are true; there are $\binom{M}{2}$ possible worlds in which all but two of the sentences are true; and so on.

We may represent a possible world as a sequence of bits (bit-string). First we have to decide on an (arbitrary) order of the atomic sentences. In the bit-string, 1 indicates that the corresponding atomic sentence is true in that world, 0 that it is false. Let us illustrate this for the case in which there are $M = 2$ atomic sentences: call 'It will rain tomorrow' atomic sentence $m = 0$ and 'We should invest in the food industry' atomic sentence $m = 1$. Then there are $w_{\max} = 4$ possible worlds, $w \in \{0, \ldots, 3\}$, which are listed in Table 5.1. Also the numbering of the possible worlds is arbitrary, but for convenience we read the sequence of 0's and 1's as a binary number. The interpretation of possible world $w = 2$, for example, is that sentence $m = 0$ is false and sentence $m = 1$ is true: in this possible world, it holds that it will not rain tomorrow and we should invest in the food industry.

Table 5.1: *With $M = 2$, there are $w_{\max} = 2^M = 4$ possible worlds, $w = 0, \ldots, w = 3$.*

	$m = 1$	$m = 0$
$w = 0$	0	0
$w = 1$	0	1
$w = 2$	1	0
$w = 3$	1	1

5.2.1.3 Theories

A theory is a subset of possible worlds.[5] Let us explain this: an agent believes the actual world to be among the possible worlds that are in his theory; he has excluded the other possible worlds as live possibilities. To see that a theory may contain more than one specific possible world, consider an agent who is sure that 'We should invest in the food industry' is true, but has no idea whether 'It will rain tomorrow' is true or false. If these are the only atomic sentences in his language, the agent holds a theory with two possible worlds. Given that we can order the possible worlds, we can represent theories as sequences of 0's and 1's, which in turn can be read as binary numbers. (This procedure is similar to the one used above for representing possible worlds by binary numbers.) Note that there are $t_{\max} = 2^{w_{\max}}$ theories that can be formulated in a language with M atomic sentences.

Table 5.2 below illustrates this set-up for the case where $M = 2$. In that table, theory $t = 0$ is the inconsistent theory, according to which all worlds are impossible; syntactically, it corresponds to a contradiction. We know beforehand that this theory is false: by ruling out all possible worlds, it also rules out the actual world. Theory $t = 15$ regards all worlds as possible; syntactically, it corresponds to a tautology. We know beforehand that this theory is true—the actual world must be among the ones that are possible according to this theory—but precisely for that reason the theory is entirely uninformative. The other theories are all consistent and of varying degrees of informational strength. The most informative ones are those according to which exactly one world is possible; a little less informative are those according to which two worlds are possible; and still less informative are the theories according to which three worlds are possible.

In Table 5.2, we have numbered the theories by interpreting their bit-string notation as a binary number. The reverse order of the worlds in the top line is so as to make world w correspond with the wth bit of the binary representation of the theory.

5.2.2 Opinion profile

So far, we have focused on the belief state of a single agent, which is expressed as a theory. Now, we consider a community of N agents. The agents start out with (possibly different) information or preferences, and therefore may vote for different theories initially. The only assumption we make about the agents' initial belief states is that they are consistent. Subsequently, the agents are allowed to communicate and adjust there preference for a theory accordingly. In particular, we model what happens when the agents communicate with all other agents whose belief states are 'close enough' to their own—that are within their bound of confidence, in Hegselmann and Krause's terminology—and update their belief state by 'averaging' over the close enough belief states, where the relevant notions of closeness and averaging are to receive formally precise definitions. The totality of belief states of a community at a

---

[5]By defining a theory in terms of possible worlds, we have chosen for a semantic approach. The equivalent syntactical approach would define a theory as a subset of sentences in the agents' language closed under the consequence relation for that language.

Table 5.2: *With $M = 2$, there are $w_{max} = 2^M = 4$ possible worlds, $w = 0, \ldots, w = 3$, and $t_{max} = 2^{w_{max}} = 16$ different theories, $t = 0, \ldots, t = 15$. The penultimate column gives the sum of bits (bit-sum), s_t, of each theory. The last column represents the opinion profile of the community.*

	$w = 3$	$w = 2$	$w = 1$	$w = 0$	s_t	opinion profile
$t = 0$	0	0	0	0	0	n_0
$t = 1$	0	0	0	1	1	n_1
$t = 2$	0	0	1	0	1	n_2
$t = 3$	0	0	1	1	2	n_3
$t = 4$	0	1	0	0	1	n_4
$t = 5$	0	1	0	1	2	n_5
$t = 6$	0	1	1	0	2	n_6
$t = 7$	0	1	1	1	3	n_7
$t = 8$	1	0	0	0	1	n_8
$t = 9$	1	0	0	1	2	n_9
$t = 10$	1	0	1	0	2	n_{10}
$t = 11$	1	0	1	1	3	n_{11}
$t = 12$	1	1	0	0	2	n_{12}
$t = 13$	1	1	0	1	3	n_{13}
$t = 14$	1	1	1	0	3	n_{14}
$t = 15$	1	1	1	1	4	n_{15}

given time can be represented by a string of t_{max} numbers, $n_0, \ldots, n_{t_{max}-1}$, where the number n_t indicates how many agents hold theory t at that time. We may also represent these numbers as a vector, $\vec{n}$. We refer to this string or vector as the (anonymous) opinion profile of the community at a specified time. Because each agent has exactly one belief state, the sum of the numbers in an opinion profile is equal to the total number of agents, N. Also, given that initially no agent has the inconsistent theory as his belief state, n_0 is always zero before any updating has taken place. Later this may change. By updating, an agent may arrive at the inconsistent theory; we shall call such an update a *zero-update* (because the inconsistent theory is represented by a string of only 0's).

In most opinion dynamics studies, a random opinion profile is used as a starting point. Because our question deals with a probability in function of the initial opinion profile, we explicitly take into account *all possible* initial opinion profiles, or—where this is not possible—take a large enough statistical sample out of all possible initial opinion profiles. The different opinion profiles can be thought of as resulting from the individual choices the agents make regarding which world or worlds they deem possible. Here, we assume that the adoption of a theory as an initial belief state can be modeled as a sequence of 2^M independent tosses of a fair coin, where the agent is to repeat the series of tosses if the result is a sequence of only 0's. As a consequence, all consistent theories have the same probability—namely, $1/(t_{max} - 1)$—of being

adopted as an initial belief state by an agent. That is to say, we are studying what in the literature are sometimes referred to as 'impartial cultures' (*cf.* Section 5.2.4). Furthermore, the agents are assumed to choose independently of each other.

5.2.3   Update rule

Theorists have studied a variety of update rules, depending on the application the authors have in mind. For instance, to model gossip communication, Deffuant et al. (2000) use a rule in which updates are triggered by pairwise interactions. To model group meetings, the updates should rather be simultaneous within the entire group of agents. During a conference, the agents meet each other face-to-face; in that case, additional effects should be taken into account, such as the 'primacy effect', which demonstrates that the order in which the agents' opinions are publicly announced may influence how the others revise their opinion.

As mentioned before, we may think of our group of agents as a group of scientists, bankers, or other experts who act as consultants in a Delphi-study. The choices in the selection of the update rule follow from that. Delphi-studies are typically conducted in a way such that the experts do not have any direct interaction (Linstone and Turoff, 1975). Thus, we need a model with simultaneous updating but without primacy effects: in this respect, the update rule of the HK model (Hegselmann and Krause, 2002) applies to this situation in a natural way.

Another relevant aspect of the HK model is that an agent may not take into account the opinions of *all* the agents in the group. This may occur when the agent knows all the opinions but does not want to take into account the opinions of agents who hold a view that is too different from the agent's own, or because the facilitator of the Delphi-study only informs the agent about the opinions of experts who hold an opinion similar to the agent's.

In order to quantify what counts as a similar opinion, we introduce the 'maximal distance' or 'bound of confidence', D. This parameter expresses the number of bits that another agent's opinion may maximally differ from one's own if that agent's opinion is to be taken into account in the updating process. To quantify the difference between two theories, we use the so-called Hamming distance of the corresponding bit-strings, defined as the number of digits in which these strings differ (Hamming, 1950).

It is possible to consider heterogeneous populations, where agents may have different bounds of confidence, as in Hegselmann and Krause (2005). Because Hegselmann and Krause (2005) report no qualitative difference between the homogeneous and the heterogeneous case, we choose the simpler, homogeneous approach: D has the same value for all agents in any population we consider. We investigate the influence of the value of D on the probability of updating to the inconsistent theory. By an agent's 'neighbors' we refer to the agents whose opinions fall within the bound of confidence of the given agent. Note that, however D is specified, an agent always counts as his or her own neighbor.

At this point, we still have to specify *how* agents update on the basis of their neighbors' belief states. Like Hegselmann and Krause in most of their studies (Hegselmann

and Krause, 2002, 2006, 2009), we choose the arguably simplest and also the most plausible averaging method, which is to determine an agent's new belief state based on the straight average of his neighbors' belief states. Our update rule for theories is a bitwise operation in two steps—averaging and rounding. First, each bit is averaged by taking into account the value of the corresponding bit of an agent's neighbors. In general, the result is a value in $[0,1]$ rather than in $\{0,1\}$. Hence the need for a second step: in case the average-of-bits is greater than $\frac{1}{2}$, the corresponding bit is updated to 1; in case the average-of-bits is less than $\frac{1}{2}$, the corresponding bit is updated to 0; and in case the average-of-bits is exactly equal to $\frac{1}{2}$, the corresponding bit keeps its initial value.[6]

In the current study, we are only interested in the probability of arriving at an inconsistent conclusion after a single update. However, again following Hegselmann and Krause, one could also designate one of the theories expressible in the agents' language as the truth and allow the agents to gather evidence which points toward the truth. One could then study the interplay between convergence to the truth and avoiding inconsistencies. We plan to implement this in future research.

We want to keep the model as simple as possible. Therefore, we will not implement any of the following additional parameters: trustworthiness of agents, physical closeness and/or social network (such as in Galam, 2002), and other psychologically relevant aspects (such as bias, self-justification, and getting tired of repeated updating; see Tavris and Aronson (2007)). Because we only have one update rule, there is no need to consider mixed groups and/or agents changing their update rule over time, which would complicate matters even further. In general, pure cases have the drawback of being less realistic, at the benefit of showing more clearly the effect of a single parameter.

5.2.4 Comparison with related work

The possibility of inconsistent outcomes resulting from a voting procedure, similar to updating beliefs, has already received some discussion in the literature on the so-called discursive dilemma. According to majority voting on a set of interrelated propositions, a proposition should be made part of the collective judgment if it is part of a majority of the individual judgments. The discursive dilemma (Pettit, 2001), which is a more general form of the doctrinal paradox (Kornhauser and Sager, 1986), shows that this voting procedure may result in an inconsistent collective judgment even if all the individuals judgments are consistent.[7] This is immediately relevant to our present concerns, given that majority voting falls under the definition of averaging to be employed in our update rule.

The original example of the doctrinal paradox is stated in the context of a legal court decision. It presents three judges who have to vote on three propositions: two

---

[6]The precise definition of the update method is stated in Appendix A. Relative to an agent who initially holds theory t_{ref}, Equation (5.3) expresses how to calculate the straight average of the wth bit within the agent's bound of confidence. Equation (5.4) formalizes the subsequent rounding step of the update rule for the wth bit of the agent's theory.

[7]A good review of the discursive dilemma can be found in List and Puppe (2007).

premises P and Q, and a conclusion R. The connection between the premises and the conclusion is motivated by legal doctrine and formalized as $R \leftrightarrow (P \wedge Q)$. A vote is called consistent if it satisfies this rule, and inconsistent otherwise. Readers will have no difficulty assigning consistent individual judgments to the judges which, given proposition-wise majority voting, nevertheless give rise to an inconsistent collective judgment.

In Pettit (2001), it is argued that this type of paradoxical result can occur in group decisions on interrelated propositions in other contexts as well. The example can be generalized to cases with more than two premises (Pettit, 2001), more than three judges (List, 2005), or to cases with a disjunctive connection rule $R \leftrightarrow (P \vee Q)$ rather than a conjunctive one (List, 2005). There are many impossibility results to be found in the literature (List and Puppe, 2007), that show that—given some plausible conditions—there is no way of aggregating consistent individual judgments so as to guarantee a consistent group judgment.

List (2005, p. 5) addresses the question of how serious the threat posed by this paradox is. He gives a probabilistic analysis of the discursive dilemma. In the general case of k premises, $P_1, \ldots, P_k$, and one conclusion, R, there are 2^k combinations of the premises being true or false (possible worlds). Each time, the truth or falsehood of the remaining proposition, the conclusion, is determined by the conjunctive connection rule $R \leftrightarrow (P_1 \wedge \cdots \wedge P_k)$. The simplest case is that in which all the probabilities are equal that an agent holds a particular opinion out of the 2^k consistent ones; this situation is called an *impartial culture*. List (2005) considers this case as well as variants thereof. In his paper, he also analyzes the case of an *impartial anonymous culture*, which takes every anonymous opinion profile to be equally likely (rather than every individual choice of the agents). Following the literature on the Condorcet jury theorem, he assumes identical probabilities for all agents and independence between different agents. The focus of List (2005) is mainly on convergence results (in particular, on the probability of inconsistency in the limit of the number of agents going to infinity), although some results are stated in terms of a finite (but always odd) number of agents.

While we also intend to give a probabilistic analysis of the occurrence of inconsistencies in what may be interpreted as group judgments and make the result general for the number of atomic propositions considered by the agents, there are some differences between our model and that of List (2005) that merit highlighting.

First, we want to model agents that—in the terminology of the discursive dilemma—vote on a theory. The type of inconsistencies encountered in the doctrinal paradox can be avoided relatively easily by having the jurors vote either on the premises only (and then derive the conclusion from the collective judgments on the premises) or on the conclusion only (Pettit, 2001). For voting on theories, which are by definition closed under derivability, there is no quick fix available to avoid the problem that agents reach the inconsistent theory by majority voting (or, in our terms, updating by averaging). Therefore, the question regarding the probability of this event seems all the more pressing.

Second, we are interested in calculating the probability of an inconsistency for a completely general number of agents (odd as well as even), rather than in convergence

results (in the limit of infinitely many agents).

And third, rather than considering majority voting where the relevant majority has to be relative to the whole group of agents, we assume an update rule that admits of greater and smaller bounds of confidence, which effectively comes to requiring a majority only relative to a subgroup of agents. This also implies that if one agent comes to hold an inconsistent theory, this need not be so for all agents in the group.

On a more practical level, we note that, because we already take into consideration three parameters (number of atomic sentences, number of agents, and the bound of confidence parameter related to the update rule), we confine our discussion to impartial cultures.

5.3   The probability of inconsistencies

We now turn to the question of how probable it is that an agent with a consistent initial belief state updates to the inconsistent belief state by averaging the (also initially consistent) belief states of that agent's neighbors. More precisely, we consider a fixed update rule—a fixed way of averaging belief states—and study the effects on the said probability of the following parameters:

1. the number M of atomic sentences of the agents' language;
2. the number N of agents in the community;
3. the bound of confidence, D, which is the maximal Hamming distance for one agent to count as a neighbor of another.

The analytical solution consists of many nested sums. In the next section, we evaluate the analytical expression numerically. Because exact calculations are only feasible for small populations, we extend the calculations by a simulation based on statistical sampling.

Given M atomic sentences, N agents, and a maximal Hamming distance or bound of confidence D, we want to calculate, first, the fraction of agents who update to the contradiction, when we consider all agents in all possible initial opinion profiles, $F_{AG}(M, N, D)$; and second, the fraction of all possible opinion profiles that have at least one agent who updates to the contradiction, $F_{OP}(M, N, D)$. In other words, $F_{AG}(M, N, D)$ is the probability for an agent to update to the inconsistent theory in a single update under the assumption that nothing is known about the opinion profile—only the parameters M, N, and D are known. Likewise, $F_{OP}(M, N, D)$ is the probability that at least one agent in the entire population will update to the inconsistent theory in a single update. Clearly, the latter probability should be at least as great as the former.

Readers who are interested in the details of the derivation of the analytical expressions for these probabilities are referred to Appendix A. Here, we state the result of the derivation, introduce previously undefined parameters occurring in it, and clarify the overall form of the expressions for $F_{AG}(M, N, D)$ and $F_{OP}(M, N, D)$:

$$F_{\mathrm{AG}}(M,N,D) \;=\; \sum_{n_1=0}^{N} \sum_{n_2=0}^{N-n_1} \cdots \sum_{n_{t_{\max}-2}=0}^{N-(n_1+n_2+\cdots+n_{t_{\max}-3})} \frac{N!}{n_0!\,n_1!\cdots n_{t_{\max}-1}!} \;\times$$

$$\frac{1}{(t_{\max}-1)^N} \sum_{t=0}^{t_{\max}-1} \left(\frac{n_t}{N} \prod_{w=0}^{w_{\max}-1} \mathrm{INV}\big[\langle B_w(t)\rangle\big] \right); \quad (5.1\mathrm{a})$$

$$F_{\mathrm{OP}}(M,N,D) \;=\; \sum_{n_1=0}^{N} \sum_{n_2=0}^{N-n_1} \cdots \sum_{n_{t_{\max}-2}=0}^{N-(n_1+n_2+\cdots+n_{t_{\max}-3})} \frac{N!}{n_0!\,n_1!\cdots n_{t_{\max}-1}!} \;\times$$

$$\frac{1}{(t_{\max}-1)^N} \mathrm{ZUP}(M,N,D,\overrightarrow{n}). \quad (5.1\mathrm{b})$$

Because these expressions take the form of nested sums, in order to explain them, we should start by looking at the last part, which is the actual core of the equation. The expression for the population-based fraction, $F_{\mathrm{OP}}(M,N,D)$, is very similar to that for $F_{\mathrm{AG}}(M,N,D)$ except for that most central part. First we look at the expression for the agent-based fraction, $F_{\mathrm{AG}}(M,N,D)$.

At the heart of the expression for $F_{\mathrm{AG}}(M,N,D)$, we find the function $\langle B_w(t)\rangle$, which specifies how a given agent in a fixed opinion profile updates the bits of his theory: it calculates the average of the wth bit for an agent in opinion profile $\overrightarrow{n}$ with bound of confidence D whose initial opinion is theory t. Because we can do this for all bits, we can determine whether or not this agent updates to the inconsistent theory; the expression

$$\prod_{w=0}^{w_{\max}-1} \mathrm{INV}\big[\langle B_w(t)\rangle\big] \quad (5.2)$$

evaluates to 1 if this is the case, and to 0 otherwise.

As a next step, we need to count the zero-updates for all the agents in the opinion profile, not just for one: $\sum_{t=0}^{t_{\max}-1}$ sums over all possible initial opinions and the factor n_t takes into account how many agents hold each of these opinions initially. We divide by N for normalization.

Moreover, we need to take into account all different, anonymous opinion profiles $\overrightarrow{n}$, not just a particular one:

$$\sum_{n_1=0}^{N} \sum_{n_2=0}^{N-n_1} \cdots \sum_{n_{t_{\max}-2}=0}^{N-(n_1+n_2+\cdots+n_{t_{\max}-3})}$$

sums over all possible anonymous opinion profiles. Thus, at the right-hand side of these summations, all the n_t's have a fixed value, meaning that there, the full opinion profile, $\overrightarrow{n}$, is specified. The weight factor $\frac{N!}{n_0!\,n_1!\cdots n_{t_{\max}-1}!}$ takes into account that certain individual choices of agents result in the same anonymous opinion profile. With this weight factor, we consider an impartial culture; omitting it would result in an impartial *anonymous* culture (see also List, 2005). The remaining factor $\frac{1}{(t_{\max}-1)^N}$

is yet another normalization factor: it divides the result by the number of different (non-anonymous) opinion profiles.

We have seen that for the agent-fractions, the central expression (5.2) calculates the fraction of agents in the particular opinion profile $\overrightarrow{n}$ which perform a zero-update. Let us now look at the opinion-based fractions: there, the central expression is replaced by

$$\text{ZUP}(M, N, D, \overrightarrow{n}),$$

which evaluates to 1 if the corresponding agent-based term (5.2) is non-zero, and to 0 otherwise. Because the rest of the expression is identical for $F_{\text{AG}}(M, N, D)$ and $F_{\text{OP}}(M, N, D)$, this ensures that $F_{\text{AG}}(M, N, D) \leq F_{\text{OP}}(M, N, D)$.

5.4 Numerical evaluation of the probability of inconsistency

It is far from trivial to estimate the outcome of the expressions for $F_{\text{AG}}(M, N, D)$ and $F_{\text{OP}}(M, N, D)$ or to analyze their limiting behavior as N and/or M become large. To obtain an idea of the quantitative output and qualitative behavior of the formulae, we evaluate them numerically.

5.4.1 Exact calculations

Because the number of computations required to evaluate Equation (5.1) is considerable, we have written a computer program (in Object Pascal) capable of evaluating the expression for the number of atomic sentences $M = 1$ to $M = 3$.[8] Here we summarize the results.

If M is equal to 1, there are no opinion profiles in which any agent updates to the contradiction, no matter which values N and D have. For $M = 2$, we obtained exact results for $N = 2$ up to $N = 21$, as shown in Panels A and C of Figure 5.1. For $M = 3$, we could obtain exact results for $N = 1$ up to $N = 4$, as shown in Panels B and D of Figure 5.1.

If $D = 1$ or $N = 2$, a zero-update never occurs. For $D > w_{\max} = 2^M$, the number of zero-updates is equal to that for $D = 2^M$ (data not shown). Because for values of $D = 2^M$ onward, all agents have all other agents as their neighbors, increasing D further makes no difference.

The lowest values for M, N, and D that may result in a zero-update are: $M = 2$, $N = 3$, and $D = 2$. (In agreement with the doctrinal paradox, we find that a zero-update can occur for three agents.) This is a case that can be checked on paper, and can thus be used for testing the validity of the analytical expression and the program. Checking an example by hand is also a good way to get to understand the model better.

---

[8]For larger values of M, both the required time and memory become too large for easy computation.
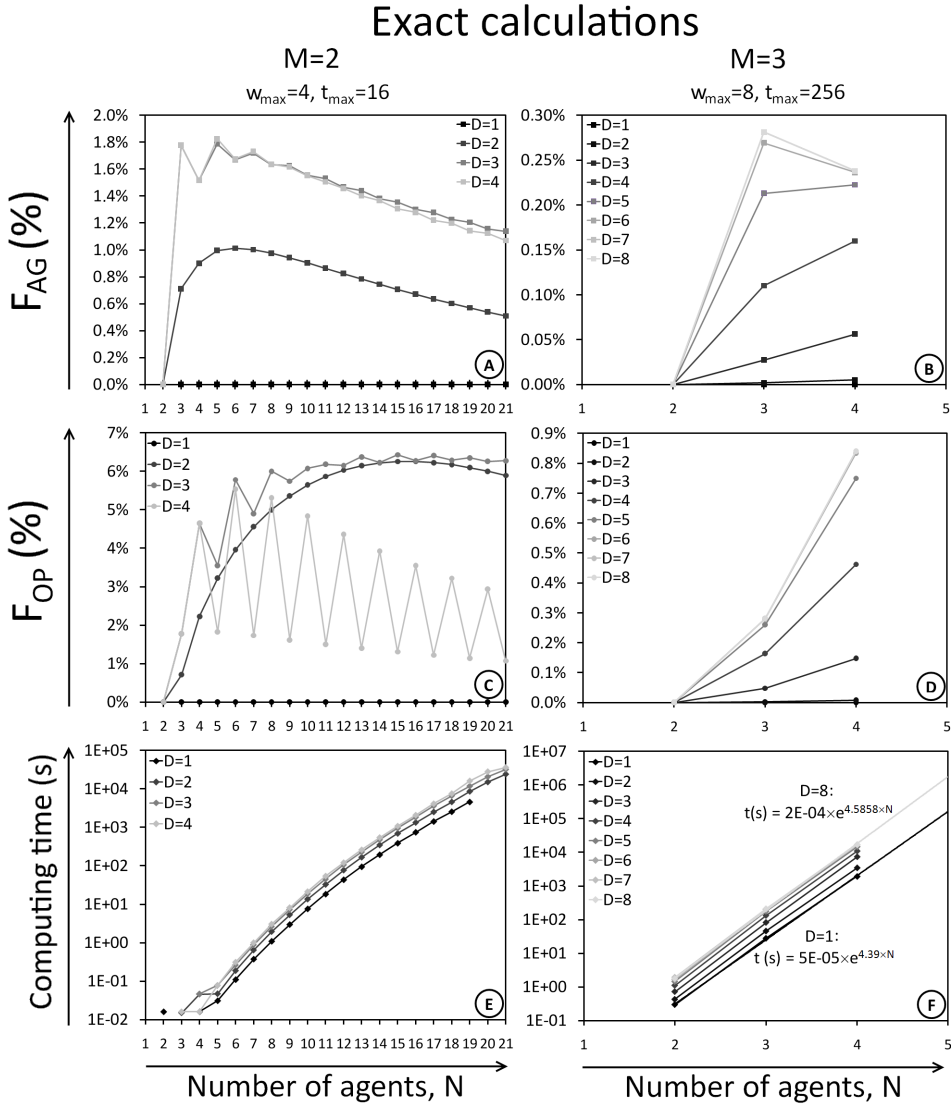
Exact calculations

Figure 5.1: *Results of exact calculations. The number M of atomic sentences equals 2 in the graphs at the left hand side (A, C, and E), and 3 in those at the right (B, D, and F). All graphs are presented as a function of the number of agents, N. The different curves in each graph represent different bounds of confidence, D. (A, B) Fraction (%) of agents who update to the inconsistent theory. (C, D) Fraction (%) of opinion profiles with at least one agent who updates to the inconsistent theory. (E, F) Computation time on one CPU in seconds on a semi-logarithmic scale.*

For $M = 2$, we can consult Table 5.2. We see that there are four theories with a bit-sum $s_t = 1$: $t = 1$, $t = 2$, $t = 4$ and $t = 8$. The distance, d, between any two of these theories equals 2. Consider an opinion profile with one agent holding $t = 1$, one agent holding $t = 2$, and one agent holding $t = 4$. If $D \geqslant 2$, then all agents are each others' neighbors. Therefore, to update, they all take the same average: $\frac{1}{3}\big((t = 1) + (t = 2) + (t = 4)\big) = \big(0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}\big)$. After rounding, their new opinion becomes $(0, 0, 0, 0)$; that is, they all arrive at $t = 0$, the inconsistent theory. For $N = 3$, populating three out of four of the theories $t = 1$, $t = 2$, $t = 4$, and $t = 8$ may happen in four different ways (given anonymity). If we keep track of the identity of the agents within each of these four opinion profiles, the agents may choose their belief states in six different ways, giving rise to $4 \times 6 = 24$ non-anonymous configurations that lead at least one agent, and thereby in fact all agents, to the inconsistent theory in just one update. For $M = 2$ and $N = 3$, there are 3 375 possible non-anonymous opinion profiles; with $D = 2$, the aforementioned 24 opinion profiles are the only starting points from which to arrive at the inconsistent theory. Therefore, the opinion-profile-fraction is $F_{\mathrm{OP}}(2, 3, 2) = 24/3\,375 = 0.711\,\%$. Because all three agents update to the inconsistent theory, the agent-fraction is exactly equal to this: $F_{\mathrm{AG}}(2, 3, 2) = 0.711\,\%$. These results are identical to the calculated value represented in the graphs.

For $D = 3$ and $D = 4$, the previous four anonymous (or 24 non-anonymous) configurations still lead to a zero-update, but there are additional possibilities which lead to the same result, to wit, those in which one agent occupies one of the six theories that have bit-sum $s_t = 2$ (namely, $t = 3$, $t = 5$, $t = 6$, $t = 9$, $t = 10$, and $t = 12$) and the other two agents each occupy a theory of bit-sum 1 such that the single 1-bit of the latter corresponds with a 0 in the first agent's theory. For instance, the combination of one agent holding $t = 3$ with another holding $t = 4$ and the third holding $t = 8$ leads to an average of $\frac{1}{3}\big((t = 3) + (t = 4) + (t = 8)\big) = \big(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}\big)$, which becomes $t = 0$ after rounding. As is readily seen, higher bounds of confidence and larger population sizes soon become too complex to check by hand. That is why computer calculations are indispensable for this type of research.

As for agent fractions, for $M = 2$ (Panel A of Figure 5.1), all curves have a similar peak shape. The values for $D = 3$ and $D = 4$ are very similar, and almost double as compared to those for the smaller bound of confidence $D = 2$. The former curves exhibit an 'odd–even wobble' near the top: for an odd number of agents a larger fraction of the population updates to the inconsistent theory than one would expect based on the behavior of even-numbered groups of similar size.

For $M = 3$ (Panel B), it is clear again that higher maximal distances give rise to higher fractions. Also the odd–even wobble seems present, especially for higher maximal distances, but we need more data to confirm this.

As for opinion profile fractions, for $M = 2$ (Panel C), the curves $D = 3$ and $D = 4$ again exhibit an odd–even wobble. Curiously, here the trend is opposite to that observed in Panel B: for an even number of agents a larger fraction of the population updates to the inconsistent theory than one would expect based on the behavior of odd-numbered groups of similar size. Furthermore, $D = 3$ and $D = 4$ are not as similar as is the case in panel B: for $D = 3$, the wobble decreases as the curve attains a maximum or plateau, whereas for $D = 4$ the wobble only decreases along with the

overall amplitude decrease of the curve.

The onset of the curves for $M = 3$ (Panel D) seems to indicate a similar reversed odd–even wobble, but again we need more data to confirm this.

Finally, a word on computing time: Although exact results for higher numbers of agents than shown here are attainable in principle, they come with ever increasing computational costs. Panels E and F of Figure 5.1—note the log-scale on the vertical axis—show that the required computing time increases nearly exponentially with the number of agents. For $M = 3$, we have extrapolated the computing time to $N = 5$ for $D = 1$ and $D = 8$, and found that one additional data point would require about 5 (for $D = 1$) to 21 (for $D = 8$) days of computation. Of course, these results are machine-dependent, but the exponential trend is intrinsic, since it is related to the fast increase in the number of terms in Equation (5.1). Obtaining more data is thus limited by practical constraints, unless we approach the problem differently.

5.4.2   Extending the numerical analysis by statistical sampling

Instead of calculating all possible opinion profiles and counting how many agents update to the inconsistent theory, we now consider a statistical approach: we have adapted the program used for calculating the exact results of the previous section to draw a random sample from all possible opinion profiles and calculate the fractions of agents and opinion profiles within the sample that update to the inconsistent theory. If the sample size is sufficiently large, these sample fractions are good estimates of the respective fractions in the complete set of opinion profiles.

We have done tests with different sample sizes, looking for a good trade-off between low noise on the data and acceptably low computational costs. All results presented in Figure 5.2 were obtained using samples of 10^3 sets of 10^3 opinion profiles each, that is 10^6 opinion profiles in total per data point. (Smaller sample sizes such as 10^4 opinion profiles per data point require a computation that is 100 times faster, but produce curves that are visibly noisy.)

Because we have some exact results, we can use these to assess the statistical program: the onset of the curves in panels A–D of Figure 5.2 corresponds well with the data presented in the respective panels of Figure 5.1.

As can be seen in panels E and F of Figure 5.2, the computation time increases at first, but then remains almost constant: a typical calculation no longer depends on the number of agents, N, but only on M and D. (The outliers are due to periods of standby time of the computer that was used for the calculations.) Thus, the approach with statistical sampling makes it possible to investigate the curves up to a much higher value of N than using the exact formula.

We have plotted the curves for $M = 2$ up to $N = 200$, where all the curves have long passed their maximum and are decreasing smoothly. For $M = 3$, we have plotted the curves up to $N = 2\,500$, because the maximum for $D = 6$ is only obtained at around $N = 1\,800$. Because the curves are smooth (apart from the odd–even wobble), instead of computing every point, from $N = 110$ onwards we have increased the step size to $\Delta N = 100$.
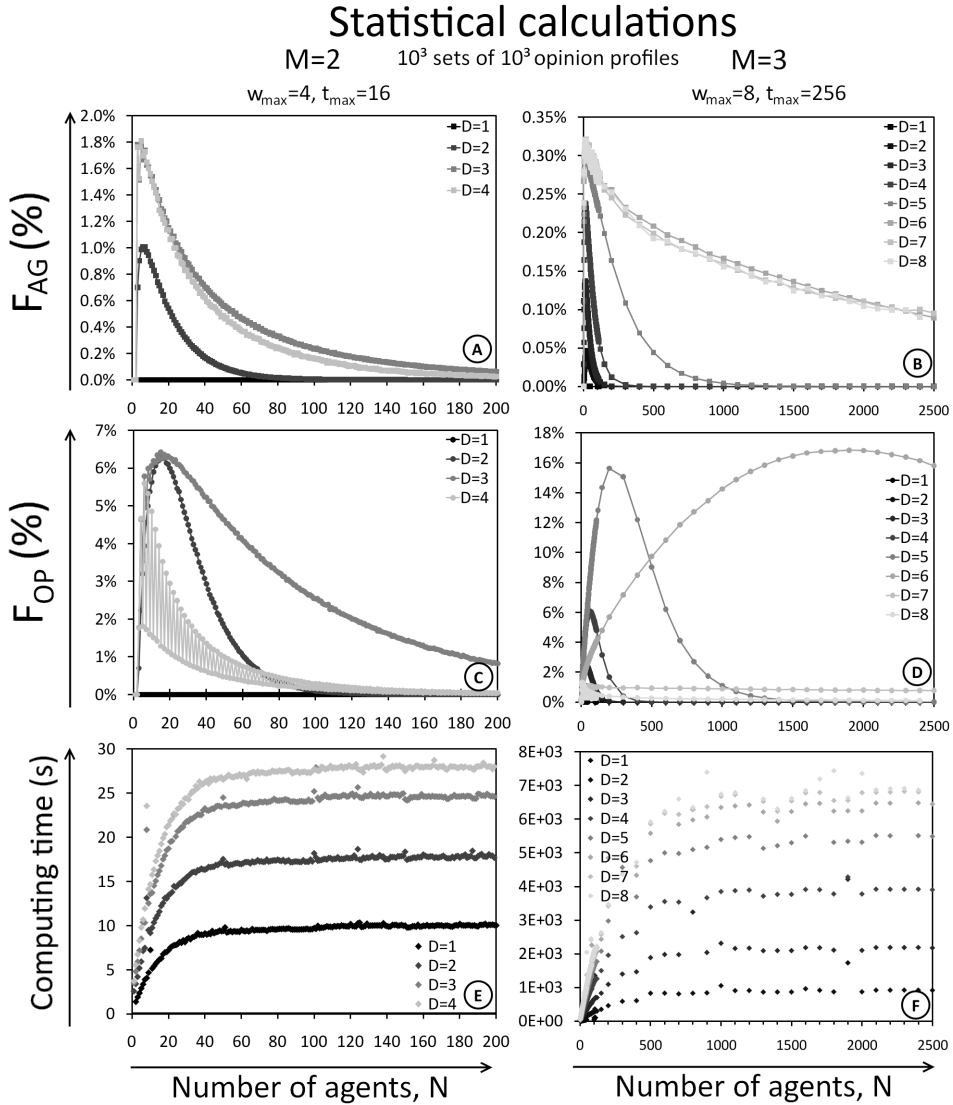
Figure 5.2: *Results of calculations based on statistical sampling, consisting of 1 000 sets of 1 000 opinion profiles each. (A, B) Fraction (%) of agents who update to the inconsistent theory. (C, D) Fraction (%) of opinion profiles with at least one agent who updates to the inconsistent theory. (E, F) Computation time on one CPU in seconds (linear scale). The number M of atomic sentences is two in the graphs at the left hand side (A, C, and E), and three in those at the right (B, D, and F). All graphs are presented as a function of the number N of agents. The different curves in each graph represent different bounds of confidence, D.*

First let us first examine the odd–even wobble that we noticed in the onset of the curves from the exact calculations.

For $M = 2$, in Figure 5.2.A we see that for the agent-based fractions, the oscillation is only present for low N-values. In Figure 5.2.C, we see that for $D = 2$ there is no odd–even wobble. For $D = 3$, it is present before the maximum in the curve, but not beyond it. For $D = 4$, the oscillation is well pronounced throughout the curve, although the amplitude of the oscillation diminishes as the curve drops.

For $M = 3$, in the agent-based fractions in Figure 5.2.B we see no odd–even wobble in the decreasing tails of the curves. Figure 5.3 provides a detail of the curves in panel D for the region near the origin where all curves overlap. There is no wobble visible for $D \in \{2, \ldots, 5\}$. For $D = 6$, there is an oscillation for N up to about 30 (long before the curve attains its maximum). For $D = 8$, the oscillation seems to go on for all values of N (much like curve $D = 4$ for $M = 2$). Because the start of curve $D = 7$ overlaps with that of $D = 8$, we present the former in a separate graph (*cf.* Figure 5.4). $D = 7$ is the only case in which we can see an amplitude modulation (like that in interfering sound waves, where the phenomenon is known as 'beats'): the oscillation seems to disappear at about $N = 34$ but its amplitude increases again for larger N until $N = 100$. There, a new oscillation starts, but because we have lowered the sampling from thereon to $\Delta N = 100$, we cannot examine it further.

We realize that this odd–even wobble cries out for an explanation. Currently, however, we have no conclusive explanation for it. For now, we present it as a puzzle.
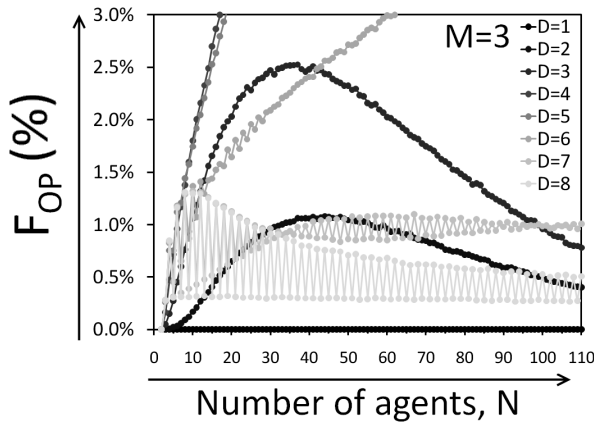

Figure 5.3: *Detail of Panel D in Figure 5.2: fraction (%) of opinion profiles with at least one agent who updates to the inconsistent theory in the case where $M = 3$.*

Let us now consider the positions for which the various curves become maximal. This is a worst-case analysis, because the maximal fractions correspond to situations with the highest probability for an agent to obtain the inconsistent theory by following the studied update rule, and for a population to have at least one agent who updates to the inconsistent theory. In Table 5.3, the maximal fractions for $M = 2$ and $M = 3$
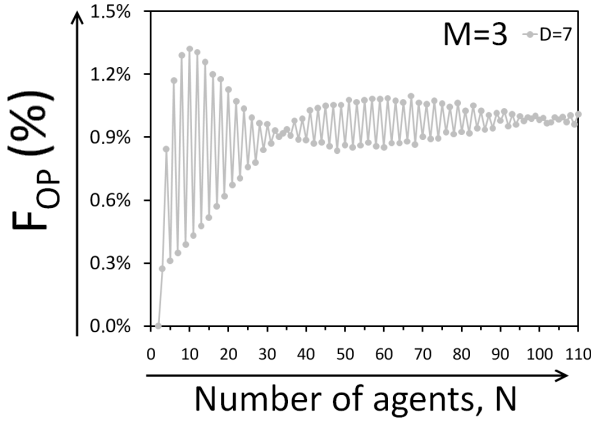
Figure 5.4: *Separate graph of maximal distance $D = 7$: fraction (%) of opinion profiles with at least one agent who updates to the inconsistent theory in the case where $M = 3$.*

are listed in terms of agents (left) and opinion profiles (right).[9] From the table, we see that compared to $M = 2$, for $M = 3$ the maximum occurs for larger N. Moreover, the percentage at the maximum is much larger for the opinion-profile-based fractions, but much smaller for the agent-based ones.

At first blush, this may seem a strange combination, so let us explain what is going on here: looking at the set of all possible opinion profiles (for the same N), there is a certain number of them that contains at least one agent who updates to the inconsistent theory. However, as the population grows larger, the average number of agents who update to the inconsistent theory decreases. Zero-updates require 'asymmetrical' opinion profiles, with all agents more or less evenly distributed at the lower bit-sum theories. (Recall the example for $M = 2$ and $N = 3$.) However, if there are many more agents than theories, $N \gg t_{\max}$, then there are many more combinations of individual choices that lead to a situation with a similar number of agents at each theory (low and high bit-sum) than there are combinations which produce opinion profiles with the agents primarily present at low-bit-sum theories. This is a consequence of the statistical law of large numbers: due to symmetry reasons, in the 'average opinion profile' (each possible belief state instantiated by $N/(t_{\max}-1)$ agents) no zero-updates are possible and, as N increases, the probability of an initial opinion profile being close to the average opinion profile also increases. Thus, large interacting groups act as a protective environment to keep the belief states of the group members consistent.

The general impression of the obtained agent- and opinion-profile-based curves in Figure 5.2 is that they vary smoothly in N. It seems that their behavior can be described effectively by an equation that has a substantially simpler form than Equa-

---

[9]For $M = 2$, all maxima occur for $N < 100$; hence, the data points represented in Figure 5.2 suffice. For $M = 3$, in all but two cases the maximum occurs for $N < 110$ and the data used to make Figure 5.2 suffice; only for $M = 3$ and $D = 5$ or $D = 6$ additional data points had to be calculated.

Table 5.3: *Maximal probability of obtaining the inconsistent theory after one update expressed as fraction (%) of agents (left) and opinion profiles (right).*

M	D	N	F_{AG} (%)
2	1	/	/
2	2	6	1.0081
2	3	3	1.7824
2	4	5	1.8112
3	1	/	/
3	2	22	0.0462
3	3	14	0.1368
3	4	11	0.2380
3	5	15	0.2986
3	6	13	0.3210
3	7	19	0.3215
3	8	13	0.3202

M	D	N	F_{OP} (%)
2	1	/	/
2	2	15	6.2930
2	3	15	6.4298
2	4	6	5.5932
3	1	/	/
3	2	44	1.0805
3	3	37	2.5250
3	4	59	6.0837
3	5	235	15.633
3	6	1780	16.867
3	7	10	1.3213
3	8	8	1.3079

tion (5.1). We tried to fit a continuous function to the discrete curves in Figure 5.2: the results are presented in Appendix B.

5.5 Concluding remarks

We have presented what can plausibly be regarded as an extension of the HK model, which currently is the most popular model for studying the dynamics of epistemically interacting agents. The extension consisted of equipping the agents with the capability of holding belief states significantly richer than the single beliefs the agents in the HK model have. As we pointed out, and as already followed from earlier work on the discursive dilemma, the extension has a price (apart from the greater mathematical and computational complexity), to wit, updating is not guaranteed to preserve consistency of an agent's belief state. The main goal of this chapter was to measure this price by determining the probability that an agent in our model indeed updates to the inconsistent theory. We investigated the effect on this probability of three key parameters: the number of agents in a community; the number of atomic sentences taken into consideration by the agents; and the bound of confidence, determining which agents count as an agent's neighbors.

Taking a social engineering perspective, and based on our results, one can make the following general recommendations for avoiding a zero-update: (i) make (if possible) the number M of atomic sentences large; (ii) avoid (if possible) even-numbered groups of agents; (iii) for a given M, choose (if possible) the number N of agents well below or above the maximum in the curves such as are given in Figure 5.2 for the specific cases of $M = 2$ and $M = 3$; and (iv) let (if possible) the agents adopt either a very low or a very high bound of confidence, D, relative to 2^M.

As we saw, apart from the trivial cases with $N = 2$ or $M = 1$, the probability for an agent (or a population) to reach the inconsistent theory after one update is always non-zero given the update rule we considered. But the good news is that an agent always has a probability $< 2\%$ of ending up in the inconsistent belief state. By making either the number of agents or the number of atomic sentences large enough, this probability can be made arbitrarily small. Seen on the scale of the whole population, the probability that one agent ends up with a contradiction is—of course—more complex; however, for sufficiently large populations, that value also decreases with N.

Because the update rule of our model seems to apply naturally to the case of a group of experts participating in a Delphi-study, the recommendations derived from this model may be useful in the design of a Delphi-study in which the experts have to state their preferences in the form of a theory. It may well be that the recommendations made here to lower the probability of inconsistencies differ from those that promote other desired features of communication among agents, such as their ability to converge to the truth. Because the present model is a very simple one, we do not claim to have captured all relevant aspects of opinion-revision of experts in Delphi-studies or of real-life communication of scientists or people in general. However, while a limited number of variables makes it easier to investigate and interpret the outcomes obtained in a model, we intend to investigate in future research the effect of some other parameters that have been held fixed in the present study.

Appendix A to Chapter 5:
Derivation of analytical expressions

This appendix describes how Equation (5.1) for the agent- and population-based fraction of zero-updates is derived. In the course of this derivation, notions such as 'possible world', 'opinion profile', and 'update rule' which have been introduced in the main text in informal terms receive a formal definition applicable to our model.

First, we take the M atomic sentences to be numbered (arbitrarily) from 0 to $M-1$. We can thus characterize possible worlds by means of bit-strings of length M: if the bit $b_m(w)$, with $m \in \{0, \ldots, M-1\}$, is 1/0, the atomic sentence m is true/false in world w. There are $w_{\max} = 2^M$ such bit-strings.

These 2^M bit-strings of length M can be numbered as well, most conveniently by their binary value. Each agent is in a particular belief state, which can be thought of as a set of worlds that the agent deems possible. Thus, the belief state of an agent can be represented as a longer bit-string of length $w_{\max} = 2^M$ with a single bit for each world equal to 1 or 0 depending on whether the agent deems that world possible or not. This results in $t_{\max} = 2^{w_{\max}} = 2^{2^M}$ different theories and thus, correspondingly, possible belief states. Each theory t, with $t \in \{0, \ldots, t_{\max}-1\}$, can be written as $w_{\max}$ bits, $B_w(t)$, with $w \in \{0, \ldots, w_{\max}-1\}$, such that:

$$t = \sum_{w=0}^{w_{\max}-1} B_w(t)\, 2^w.$$

One readily verifies that, given this notation, theory $t = t_{\max} - 1$ assigns 1 to all possible worlds indeed, and $t = 0$ assigns 0 to all possible worlds.

The sum of bits of a theory t—the 'bit-sum', written as s_t—indicates the number of bits equal to 1 in theory t. This can be stated formally as follows:

$$s_t = \sum_{w=0}^{w_{\max}-1} B_w(t).$$

Recall that, to specify the entire community of all N agents and their belief states, we can count the number of agents whose belief states are represented by a given theory t and denote it as $n_t \in \{0, \ldots, N\}$, with $\sum_{t=0}^{t_{\max}-1} n_t = N$. Hence, the entire opinion profile is specified by[10]

$$\overrightarrow{n} = \langle n_0, n_1, \ldots, n_{t_{\max}-1} \rangle.$$

Recall further that, because the inconsistent theory $t = 0$ is excluded as an initial belief state for all agents, n_0 is always 0 in the initial opinion profile.

To determine the new opinion profile of the whole community of agents after one update, we first formalize how a single 'reference' agent updates his belief state, t_{ref}, on the basis of the belief states of the other agents in the community.

---

[10]Because we do not keep track of which agent endorses which theory, this is an *anonymous* opinion profile. With the update rule used in this chapter, agents who initially are in the same belief state, will always update to the same belief state, so no information is lost by this choice of notation.

As a first step, the reference agent has to calculate the Hamming distance of the bit-string representing his belief state, t_{ref}, to that of the bit-strings representing the belief states of the other agents. The Hamming distance between t_{ref} and another bit-string t is equal to the bit-sum of the difference bit-string, which can be found by applying an exclusive-or (XOR) operator: $t \text{ XOR } t_{\text{ref}}$. Whereas this operation is familiar in information theory, here we prefer the equivalent algebraic procedure of performing a bit-wise addition followed by modulo 2:

$$d(t, t_{\text{ref}}) \; = \; \sum_{w=0}^{w_{\max}-1} \left(\left(B_w(t) + B_w(t_{\text{ref}}) \right) \mod 2 \right).$$

In a community with opinion profile $\vec{n}$, the number of agents with a belief state at a distance d from t_{ref} is called $a_d(t_{ref}, \vec{n})$:

$$a_d(t_{\text{ref}}, \vec{n}) \; = \; \sum_{t=0}^{t_{\max}-1} n_t \; \delta_{d, d(t, t_{\text{ref}})},$$

where $\delta_{d, d(t, t_{\text{ref}})}$ is a Dirac delta, which is 0 if the two indices are different and 1 if they are equal.

Now, the reference agent may count the number of agents A within his or her bound of confidence, $d(t, t_{\text{ref}}) \leqslant D$ (which, it will be recalled, necessarily includes him- or herself) as follows:

$$A \; = \; \sum_{d=0}^{D} a_d(t_{\text{ref}}, \vec{n}) \; = \; a_0(t_{\text{ref}}, \vec{n}) + \ldots + a_D(t_{\text{ref}}, \vec{n}).$$

The next step is to determine how to update any specific bit in the reference agent's belief state, $B_w(t_{\text{ref}})$. To stay as close as possible to the HK model, the agents in our model take the arithmetic mean (straight average) over the corresponding bits of the belief states of all their neighbors. We denote the average using angle brackets:[11]

$$\left\langle B_w(t_{\text{ref}}) \right\rangle \; = \; \frac{1}{A} \sum_{d=0}^{D} \sum_{t=0}^{t_{\max}-1} B_w(t) \; n_t \; \delta_{d, d(t, t_{\text{ref}})}. \tag{5.3}$$

Based on the above average, the agent decides how to update the value of the wth bit. If the average is smaller than $\frac{1}{2}$, the agent sets this bit to 0; if it is larger than $\frac{1}{2}$, to 1; and if it is precisely equal to $\frac{1}{2}$, the agent will keep his or her initial value for that bit. So, in a sense, we have majority voting here, with the important proviso that the majority is taken relative to an agent's neighbors and not (necessarily) relative to all agents in his or her community.

---

[11]Given that the average depends on the bound of confidence D and the opinion profile of the community $\vec{n}$, we could put those next to the brackets as a subscript. However, to keep the notation light, we have omitted this.

The update rule for the wth bit of an agent who initially holds theory t_{ref} is formalized as a function UPD:

$$
\text{UPD}\big[\big\langle B_w(t_{\text{ref}})\big\rangle\big] = \begin{cases} 1 & \text{if} \quad \big\langle B_w(t_{\text{ref}})\big\rangle > \frac{1}{2} \\ B_w(t_{\text{ref}}) & \text{if} \quad \big\langle B_w(t_{\text{ref}})\big\rangle = \frac{1}{2} \\ 0 & \text{if} \quad \big\langle B_w(t_{\text{ref}})\big\rangle < \frac{1}{2}. \end{cases} \tag{5.4}
$$

To make counting zero-updates more convenient, we introduce a function INV (for 'inverse') that has value 1 if the corresponding updated bit is equal to 0 and vice versa:

$$
\text{INV}\big[\big\langle B_w(t_{\text{ref}})\big\rangle\big] = \begin{cases} 0 & \text{if} \quad \big\langle B_w(t_{\text{ref}})\big\rangle > \frac{1}{2} \\ 1 - B_w(t_{\text{ref}}) & \text{if} \quad \big\langle B_w(t_{\text{ref}})\big\rangle = \frac{1}{2} \\ 1 & \text{if} \quad \big\langle B_w(t_{\text{ref}})\big\rangle < \frac{1}{2}. \end{cases}
$$

An update to the contradiction corresponds to updating all bits to 0. We can count those events by multiplying the result of INV over all bits: due to the former definition, this product will only be 1 if all bits are updated to 0.

In order to determine the opinion-profile-based fraction, we introduce a function ZUP (for 'zero-update') that is 1 if there is at least one agent in the community who updates to the contradiction, and 0 otherwise:

$$
\text{ZUP}(M, N, D, \vec{n}) = \begin{cases} 0 & \text{if} \quad \sum_{t=0}^{t_{\max}-1} \left(\frac{n_t}{N} \prod_{w=0}^{w_{\max}-1} \text{INV}\big[\big\langle B_w(t)\big\rangle\big]\right) = 0 \\ 1 & \text{if} \quad \sum_{t=0}^{t_{\max}-1} \left(\frac{n_t}{N} \prod_{w=0}^{w_{\max}-1} \text{INV}\big[\big\langle B_w(t)\big\rangle\big]\right) > 0. \end{cases}
$$

Now we can determine the agent- and opinion-profile-based fractions of zero-updates, $F_{\text{AG}}(M, N, D)$ and $F_{\text{OP}}(M, N, D)$, by summing over all combinations of the agents' belief states, $\vec{n}$. Because each theory t, with $t \in \{0, \ldots, t_{\max} - 1\}$, is equally likely to be chosen by all agents, we need to sum over all possible combinations of choices. For the sum-indices we use the following notation: $t(n)$ is the theory representing agent n's initial belief state. The requisite functions can then be written as follows:

$$
F_{\text{AG}}(M, N, D) = \sum_{t(0)=1}^{t_{\max}-1} \cdots \sum_{t(N-1)=1}^{t_{\max}-1} \frac{1}{(t_{\max}-1)^N} \times
$$
$$
\sum_{t=0}^{t_{\max}-1} \left(\frac{n_t}{N} \prod_{w=0}^{w_{\max}-1} \text{INV}\big[\big\langle B_w(t)\big\rangle\big]\right); \tag{5.5a}
$$

$$
F_{\text{OP}}(M, N, D) = \sum_{t(0)=1}^{t_{\max}-1} \cdots \sum_{t(N-1)=1}^{t_{\max}-1} \frac{1}{(t_{\max}-1)^N} \, \text{ZUP}(M, N, D, \vec{n}). \tag{5.5b}
$$

Because the number of terms is equal to the number of ways the agents can choose a theory as their belief state, that is (Rosen, 2000, p. 55),

$$
P^R(t_{\max} - 1, N) = (t_{\max} - 1)^N,
$$

each term is weighted by the inverse of this.

Equation (5.5) does not have the exact same form as Equation (5.1). To simplify the evaluation of $F_{\mathrm{AG}}(M, N, D)$ and $F_{\mathrm{OP}}(M, N, D)$, we can reduce the number of terms drastically by only summing over all *different* anonymous opinion profiles and introducing an additional weight function (multiset coefficient—see Rosen, 2000, p. 55) $\frac{N!}{n_0! n_1! \cdots n_{t_{\max}-1}!}$:[12]

$$F_{\mathrm{AG}}(M, N, D) = \sum_{n_1=0}^{N} \sum_{n_2=0}^{N-n_1} \cdots \sum_{n_{t_{\max}-2}=0}^{N-(n_1+n_2+\cdots+n_{t_{\max}-3})} \frac{N!}{n_0! n_1! \cdots n_{t_{\max}-1}!} \times$$
$$\frac{1}{(t_{\max}-1)^N} \sum_{t=0}^{t_{\max}-1} \left(\frac{n_t}{N} \prod_{w=0}^{w_{\max}-1} \mathrm{INV}\big[\langle B_w(t) \rangle\big] \right);$$

$$F_{\mathrm{OP}}(M, N, D) = \sum_{n_1=0}^{N} \sum_{n_2=0}^{N-n_1} \cdots \sum_{n_{t_{\max}-2}=0}^{N-(n_1+n_2+\cdots+n_{t_{\max}-3})} \frac{N!}{n_0! n_1! \cdots n_{t_{\max}-1}!} \times$$
$$\frac{1}{(t_{\max}-1)^N} \mathrm{ZUP}(M, N, D, \vec{n}).$$

This concludes the derivation of Equation (5.1).

The number of terms in Equation (5.1) is the multiset coefficient, which represents the number of ways to choose N out of $t_{\max} - 1$ with repetition (see Rosen, 2000, p. 55):

$$C^R(t_{\max} - 1, N) = \binom{N + t_{\max} - 1 - 1}{N} = \frac{(N + t_{\max} - 2)!}{(N)!(t_{\max} - 2)!}.$$

This number is smaller than or equal to the number of terms in Equation (5.5), $P^R(t_{\max} - 1, N)$.[13]

---

[12]There is no summation over $n_{t_{\max}-1}$, because we know by the condition $\sum_{t=0}^{t_{\max}-1} n_t = N$ that $n_{t_{\max}-1} = N - (n_1 + n_2 + \cdots + n_{t_{\max}-2})$.

[13]To illustrate how drastic the reduction in computational cost is, consider the case where $M = 2$ and $N = 12$: Equation (5.5) has $P^R(t_{\max}-1, N) = 129\,746\,337\,890\,625$ terms, whereas Equation (5.1) has 'only' $C^R(t_{\max} - 1, N) = 9\,657\,700$ terms. This means that the number of evaluations in this example is more than 13 million times smaller for Equation (5.1) as compared to Equation (5.5).

Appendix B to Chapter 5: Fitting curves to the statistical data

We tried to fit different asymmetric peak shapes—such as Poisson, Weibull and log-normal—to the data series presented in Figure 5.2. The fitting procedure was performed with commercial software (SigmaPlot). Although no single equation resulted in least-square fits with good R^2 values for (almost) all agent- and opinion-profile-based D-curves, a three parameter log-normal distribution gave the best overall result. Its distribution function is given by:

$$f(N) = Ae^{-\frac{1}{2}\left(\frac{\log \frac{N}{C}}{B}\right)^2}$$

The goal of a least-squares fit is to determine the values of the parameters—here A, B, and C—such that the sum of the squares of the distance between the data points and the value of the fit-curve is minimal.

To avoid deterioration of the fit quality due to the odd–even wobble, we have split the data sets into separate files for the values at odd and at even numbers of agents prior to the fitting procedure. The results for all odd and even, agent- and opinion-profile-based D-curves can be found in Table 5.4 for $M = 2$ and in Table 5.5 for $M = 3$. For $M = 2$, $R^2 > 0.9$ for all curves. For $M = 3$, there are eight curves with $R^2 < 0.9$, five of which with R^2 between 0.8 and 0.9. For the three remaining cases with $R^2 < 0.8$ (F_{AG} for N odd and $D = 2$, F_{OP} for N even and $D = 7$, and F_{OP} for N odd and $D = 8$) the values for the parameters A, B, and C are not shown.
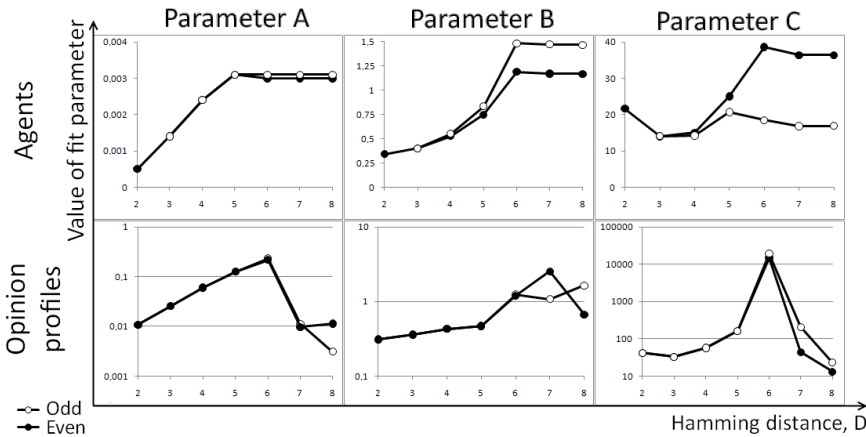

Figure 5.5: *Values of parameters obtained by fitting a log-normal curve to the agent- and opinion-profile-based D-curves for the case M = 3. The filled dots represent the values for an even number of agents, the open dots for odd numbers. For the agent-based values, a linear scale is used; for the opinion-profile-values, a logarithmic scale.*

A graphical representation of the case where $M = 3$ can be found in Figure 5.5.

These curves suggest a relation among the obtained parameter values, especially for the agent-based curves. For instance, the fit parameters for the even and odd case correspond well for low D-values, but diverge at higher values, and more drastically so ranging from parameter A over B to C. In all curves, the behavior changes at $D = 5$ or $D = 6$. For the agent-data, parameter A starts off with a linear trend in D, reaching a plateau from $D = 5$ on. For the opinion profile data, the initial trend of parameter A is exponential linear on the log-scale); this only changes at $D = 6$. For parameters B and C, the initial behavior in terms of D increasingly deviates from linearity (or exponential behavior in the opinion profile case). Though suggestive, the number of data points is insufficient to predict the shape of D-curves for higher values of M.

Table 5.4: *$M = 2$. Values of R^2 and the three fit parameters with standard error for a log-normal fit to the agent- and opinion-profile-based D-curves, for an even and odd number of agents.*

F_{AG} at even positions		
D	2	3	4
R^2	0.946	0.9736	0.9317
A	0.0157 ± 0.0004	0.0099 ± 0.0002	0.0152 ± 0.0004
B	$0.475 \ \ \pm 0.015$	0.3654 ± 0.0094	$0.511 \ \ \pm 0.017$
C	$9.02 \ \ \ \pm 0.42$	$7.86 \ \ \ \pm 0.22$	$9.39 \ \ \ \pm 0.50$

F_{AG} at odd positions		
D	2	3	4
R^2	0.9975	0.9988	0.9982
A	$0.0102 \pm 7\text{E-}5$	$0.0181 \pm 9\text{E-}5$	0.0182 ± 0.0001
B	0.4051 ± 0.0035	0.6744 ± 0.0047	0.6169 ± 0.0052
C	$6.819 \ \ \pm 0.076$	$4.663 \ \ \pm 0.085$	$4.778 \ \ \pm 0.095$

F_{OP} at even positions		
D	2	3	4
R^2	0.9892	0.9516	0.9516
A	0.0660 ± 0.0008	0.0651 ± 0.0009	0.0500 ± 0.0015
B	0.3342 ± 0.0050	$0.577 \ \ \pm 0.014$	$0.423 \ \ \pm 0.014$
C	$14.44 \ \ \pm 0.20$	$16.28 \ \ \pm 0.50$	$8.03 \ \ \ \pm 0.35$

F_{OP} at odd positions		
D	2	3	4
R^2	0.9898	0.9934	0.9982
A	0.0663 ± 0.0008	0.0639 ± 0.0003	0.0182 ± 0.0001
B	0.3320 ± 0.0048	0.5444 ± 0.0045	0.6169 ± 0.0052
C	$14.51 \ \ \pm 0.20$	$18.24 \ \ \pm 0.18$	$4.778 \ \ \pm 0.095$

Table 5.5: *M = 3. Values of R² and the three fit parameters (with standard error) for a log-normal fit to the agent- and opinion-profile-based D-curves, for an even and odd number of agents.*

F_{AG} at even positions						
D	2	3	4	5	6	7	8
R^2	0.9982	0.9956	0.9831	0.9592	0.8699	0.8747	0.8751
A	0.0005 ± 2E-5	0.0014 ± 1E-5	0.0024 ± 3E-5	0.0031 ± 5E-5	0.0030 ± 4E-5	0.0030 ± 4E-5	0.0030 ± 4E-5
B	0.3414 ± 0.0018	0.4024 ± 0.0035	0.5294 ± 0.0096	0.749 ± 0.025	1.190 ± 0.046	1.172 ± 0.045	1.170 ± 0.045
C	21.66 ± 0.10	13.97 ± 0.14	15.04 ± 0.37	25.0 ± 1.1	38.6 ± 3.6	36.4 ± 3.4	36.4 ± 3.4

F_{AG} at odd positions						
D	2	3	4	5	6	7	8
R^2	0.341	0.9973	0.9948	0.9724	0.9891	0.9919	0.9873
A	/	0.0014 ± 8E-6	0.0024 ± 2E-5	0.0031 ± 4E-5	0.0031 ± 1E-5	0.0031 ± 1E-5	0.0031 ± 1E-5
B	/	0.4021 ± 0.0028	0.5518 ± 0.0058	0.835 ± 0.024	1.486 ± 0.020	1.4734 ± 0.018	1.470 ± 0.023
C	/	14.05 ± 0.11	14.21 ± 0.21	20.71 ± 0.95	18.47 ± 0.84	16.84 ± 0.69	16.91 ± 0.87

F_{OP} at even positions						
D	2	3	4	5	6	7	8
R^2	0.9982	0.9938	0.9888	0.9444	0.9892	0.1241	0.8755
A	0.0108 ± 4E-5	0.0255 ± 0.0002	0.0601 ± 0.0005	0.1268 ± 0.0034	0.217 ± 0.019	0.0097 ± 0.0002	0.0114 ± 0.0003
B	0.3105 ± 0.0018	0.3607 ± 0.0039	0.4290 ± 0.0085	0.470 ± 0.013	1.197 ± 0.068	2.54 ± 0.58	0.668 ± 0.034
C	41.54 ± 0.15	32.86 ± 0.26	57.02 ± 0.83	162.1 ± 5.9	148E+2 ± 62E+2	44 ± 29	13.0 ± 1.0

F_{OP} at odd positions						
D	2	3	4	5	6	7	8
R^2	0.9983	0.9933	0.9883	0.9456	0.9917	0.8524	0.663
A	0.0108 ± 4E-3	0.0255 ± 0.0002	0.0601 ± 0.0005	0.1266 ± 0.0033	0.233 ± 0.022	0.0112 ± 0.0002	/
B	0.3086 ± 0.0018	0.3607 ± 0.0040	0.4284 ± 0.0085	0.469 ± 0.013	1.234 ± 0.067	1.074 ± 0.037	/
C	41.58 ± 0.14	32.93 ± 0.27	56.96 ± 0.83	161.3 ± 5.8	194E+2 ± 84E+2	206 ± 12	/

Chapter 6

Evaluation and Outlook

> *A good question is never answered.*
> *It is not a bolt to be tightened into place.*
> *But a seed to be planted and to bear more seed.*
> *Toward the hope of greening the landscape of idea.*
>
> John Ciardi (1972)

In Chapter 1, we have formulated two goals for this dissertation on the philosophy of probability. In section 6.1, we will evaluate to what extent these goals have been met. In section 6.2, we sketch a plan for future work.

6.1 Evaluation

> *[T]he point of philosophy is to start with something so simple as not to seem worth stating, and to end with something so paradoxical that no one will believe it.*
>
> Bertrand Russell (1986)

The first goal was to develop a mathematical basis for probability theory that allows us to deal with infinite sample spaces in a way that is satisfactory from the epistemological point of view. In Chapter 2, this goal has been met for the specific case of a fair lottery on the natural numbers, which is a situation that could not be modeled within Kolmogorov's axiomatization of probability. At the beginning of this investigation, we did not know about numerosity theory; we started from a construction based on free ultrafilters, which later on turned out to be equivalent to a model for numerosities. With the benefit of hindsight, we can summarize our solution in just two words: "normalized numerosities". The solution leads to a non-zero, infinitesimal probability of winning for all non-empty, finite subsets of tickets. This restores Regularity for this case. The generalized limiting frequency of the

associated sequence, which is necessarily a standard real number, can be obtained by taking the standard part of our non-standard probability function.

An additional part of the first goal was to check whether the obtained solution generalizes to smaller and larger cardinalities. In Chapter 3, we proposed a solution for a problem with rational beliefs related to probabilities on sample spaces of smaller—*i.e.* finite—cardinalities: the Lottery Paradox. The solution was inspired by that for the infinite lottery and is also based on non-standard analysis, in particular on the framework of relative or stratified analysis. This led to the definition of Stratified Belief, which is intended to overcome a number of unsatisfactory properties of the threshold-based model of rational belief: Stratified Belief formalizes the vague notion of probabilities 'sufficiently close to unity' without imposing a sharp boundary on the probability values and it is explicitly context-dependent. In Chapter 4, we focused on the analogies between large but finite (or indefinitely large) sample spaces and (countably) infinite sample spaces. The method of Chapter 2 generalizes to all countably infinite sample spaces, but not directly to the uncountable case. A natural follow-up study would be to develop a more general approach that is also suitable for higher cardinalities. We will come back to this in the outlook, given in section 6.2.

The second goal was to apply probability theory and an agent-based model to a problem in formal epistemology. In Chapter 5, we selected the problem of a group of agents whose opinions take the form of a theory. This study merges two research lines: that of opinion dynamics in sociophysics, and that of the discursive dilemma. Even when all agents start out with a consistent theory, they need not hold a consistent theory at a later point in time, if the agents update their opinion by averaging (in a specific way) each other's opinions. The goal was to find the probability that an agent arrives at an inconsistency after a single update. We regarded this probability as a function of three parameters: the number of atomic sentences in the theory, the number of agents, and the bound of confidence, which influences how many other agents' opinions are taken into account in a specific update. Although an analytical solution was found, it proved to be necessary to run computer simulations in order to evaluate it numerically.

Apart from providing an answer to a specific research question, Chapter 5 can also be regarded as a case study of what the application of probability theory and computer simulations can (and cannot) contribute to philosophy. Probability provides a formal framework and simulations provide data. Neither diminishes the need for critical thinking, thorough analysis, and other philosophical skills—on the contrary. What they contribute to philosophy are new opportunities as well as pitfalls: new challenges that philosophy cannot (and should not) refuse to take up. These new approaches deserve a place in philosophy, but their true merits will only become clear over time.

6.2 Outlook

> *Any change or reform you make is going to have consequences you don't like.*
>
> Morris K. Udall

Although a PhD dissertation is the end product of a doctoral study, it is really only the beginning of a research project. At this point, important questions have been left unanswered, but at least we now have some clues as to where we may find the answers.

As remarked in the evaluation of section 6.1, we have developed a method to describe a lottery on any countably sample space, but not for the uncountable case, yet. The remainder of this chapter presents a preliminary account of a general method of dealing with infinite sample spaces using probability functions that take values on a non-standard field. This will be developed further, in collaboration with Vieri Benci and Leon Horsten.

First, we will show that there are multiple, interrelated problems with Kolmogorov's axioms (as presented in section 1.4.2.1). To solve all of them, it does not suffice to make a minor adaptation of one of the axioms. Instead, we propose a new system of axioms and show that they are consistent. Kolmogorov (1933) presents probability theory as a part of measure theory, which is a branch of standard analysis or calculus. Because standard analysis does not allow for non-Archimedean quantities (*i.e.* infinitesimals), we may call Kolmogorov's approach an 'Archimedean probability theory'. We show that allowing non-Archimedean probability values may have considerable epistemological advantages in the infinite case.

6.2.1 Problems with Kolmogorov's axiomatization

Kolmogorov's probability theory works fine as a mathematical theory, but the *direct* interpretation of its language leads to counterintuitive results. Also, the fact that some seemingly simple situations cannot be described within Kolmogorov's system calls for some epistemological reflection. First, we will give an overview of known problems with Kolmogorov's axiomatization of probability theory. Then, we will lay out our solution strategy for these issues: instead of fighting the symptoms one-by-one, we will carefully select a new set of axioms. These axioms are stated in the next section.

6.2.1.1 Non-measurable sets

The elements of the domain of the probability function, $\mathfrak{A}$, are called events. A peculiarity of axiom (K0A) is that it allows $\mathfrak{A} \neq \mathcal{P}(\Omega)$. In fact, it is well known that there are (probability) measures (such as the Lebesgue measure on $[0,1]$) which cannot be defined for all the sets in $\mathcal{P}(\Omega)$. Thus, there are sets in $\mathcal{P}(\Omega)$ which are not events, which means that their probability value is undefined. This may even

occur for sets that are the union of elementary events in $\mathfrak{A}$.[1]

There does exist a type of function that is related to probability functions and that does always have the full set $\mathcal{P}(\Omega)$ as its domain: outer measures, but these functions are only countably subadditive, thus not necessarily CA. The way to turn an outer measure into a probability measure, which is necessarily CA, is by means of the Caratéodory extension theorem. However, the domain may have to be restricted in the process, and the resulting domain $\mathfrak{A}$ may be a proper subset of $\mathcal{P}(\Omega)$. This shows that there is a hidden connection between axiom (K0A) and axiom (K4): if we want to change the former, we will have to adapt the latter as well.

6.2.1.2   Problems with the interpretation of $P = 0$ and $P = 1$ events

It seems natural to interpret events that have zero probability as impossible events. This idea lies at the basis of the concept of 'Regularity', which can be stated formally as follows:

Regularity. For any event A:

$$P(A) = 0 \Leftrightarrow A = \varnothing \tag{6.1}$$

Observe that axiom (K1) does allow for events other than the empty set to have probability zero. In particular, when Ω is infinite, it may occur that possible events have probability zero. Thus, whereas Kolmogorov's axioms do guarantee the implication from right to left, they fail to secure the implication in the opposite direction. Hence, Kolmogorov's approach violates Regularity.

In particular, there are situations such that Ω is infinite and *all* the elementary events have probability zero. Then it seems like we have:

$$P(E_j) = 0, \quad j \in J \tag{6.2}$$

and

$$P\left(\bigcup_{j \in J} E_j\right) = 1. \tag{6.3}$$

This situation is very common when J is not denumerable. It looks as if eq. (6.2) states that each event E_j is impossible, but eq. (6.3) states that one of them will certainly occur. In this case, not only is Regularity violated, but CA, which is axiom (K4) of Kolmogorov's system, also fails. In other words, this situation cannot be described in the system at all.[2]

The interpretation of $P = 1$ events turns out to be just as problematic as the interpretation of $P = 0$ events: axiom (K2) does not forbid that a set $A \subsetneq \Omega$ has probability zero, and indeed, when Ω is infinite, the occurrence of an event with

---

[1]For example, if $\Omega = [0, 1]_{\mathbb{R}}$ and P is given by the Lebesgue measure, then all the singleton $\{x\}$ are measurable, but there are non-measurable sets; namely the union of events might not be an event.

[2]Hence the 'seems' in 'Then it seems like we have...'.

probability unity is not necessary. This observation can serve as an alternative motivation for Regularity. Indeed, Easwaran (2010) proposed to let 'Regularity' also refer to the 'twin' of eq. 6.1, which says that unit probability should be reserved for the certain event. This can be formalized as follows:

Regularity†. For any event A:

$$P(A) = 1 \Leftrightarrow A = \Omega \tag{6.4}$$

Again, Kolmogorov's axioms only ensure the implication in the right-to-left direction.

For those who regard Regularity (and its twin) merely as a convenience, there is an obvious solution at hand: interpret probability 0 as 'very unlikely' (rather than simply as 'impossible'), and interpret probability 1 as 'almost certain' (instead of 'absolutely certain'). Yet, there is a philosophical price to be paid to avoid these contradictions: the correspondence between mathematical formulas and reality is now quite vague—just how probable is 'very likely' or 'almost certain'?—and far from intuition.[3]

Moreover, this solution is not acceptable for authors who regard Regularity as a norm of rationality, such as Skyrms (1980) and Lewis (1986a). These authors have suggested that Regularity can be restored by allowing infinitesimals in the range of the probability function. In contrast to their optimistic viewpoint, Williamson (2007), Hájek (2010), and Easwaran (2010) have argued that switching to a non-Archimedean range may be of help in some cases, but does not restore Regularity in all circumstances.

We agree with the position of Easwaran (2010), who states that simply *claiming* that introducing infinitesimals restores Regularity is insufficient. We hope to convince the reader that it actually does, by *proving* it, as we will do in a paper that is in preparation at the time of this writing. In particular, we will show that Regularity can be obtained for a non-Archimedean probability function in the case of an infinite sequence of fair coin tosses.

6.2.1.3 Problems with conditional probability

The fact that in Kolmogorov's theory possible events may have probability zero also leads to problems with conditional probability, as defined by (D2). Popper (1938) developed his own basis for probability theory (see also the new appendix iv in the reprinted version of Popper, 1959), consisting of six axioms, in which conditional probability—denoted as $p(a, b)$—is fundamental. Within his system, conditionalizing on $P = 0$-events does not pose a particular difficulty. However, Popper limits his system to situations with, at most, denumerably many elements. Thus, his approach does not pose a general alternative to conditionalization on $P = 0$-events.

---

[3]The situation is somewhat similar to the study of probabilistic semantics for 'default rules' (rules to which there are exceptions). Adams (1966) interpreted the conditional statement 'if A then B' as a constraint on the conditional probability: $P(B|A) > 1 - \epsilon$, where ϵ is a positive real number that can be arbitrarily small. Later authors, such as Lehmann and Magidor (1992), have considered a similar semantics, but replacing the ϵ with a positive infinitesimal in the sense of NSA.

6.2.1.4 Fair lotteries

Fair odds is a very fundamental concept, not only in ethics (Stone, 2008b), but also in probability theory. In the classical interpretation of probability theory of Laplace (1814), equiprobability shows up in the important 'Principle of Insufficient reason' or the 'Principle of Indifference' (PI). PI states that if there is no reason to give a higher weight to any of all (n) of the elementary events, then we should assign exactly the same probability ($1/n$) to each of them. Although there are known problems with PI (*cf.* the paradoxes of Bertrand, 1888), the idea behind the principle is still appealing. After all, the Principle of Maximal Entropy can be regarded as a modern version of PI (Jaynes, 1957, 1973, Uffink, 1995).

Considering fair lotteries, we discover two things. First, we discover that the choice for $[0,1]_\mathbb{R}$ as the range of P is neither necessary to describe a fair lottery in the finite case nor sufficient to describe one in the infinite case:

- For a fair finite lottery, the unit interval of $\mathbb{R}$ is not necessary as the range of the probability function: the unit interval of $\mathbb{Q}$ suffices.

- In the case of a fair lottery on $\mathbb{N}$, $[0,1]_\mathbb{R}$ is not sufficient as the range: it violates our intuition that the probability of any set of tickets can be obtained by adding the probabilities of all individual tickets.[4]

Second, we discover that some very simple problems cannot be dealt with in Kolmogorov's formalism. In order to illustrate this, let us focus on the fair lottery on $\mathbb{N}$ (de Finetti, 1974). In this case, the sample space is $\Omega = \mathbb{N}$ and we expect $\mathfrak{D}$ to contain all the singletons of $\mathbb{N}$—otherwise there would be 'tickets' (individual, natural numbers) whose probability is undefined, which would be strange because we know they are equal in a fair lottery. Also, we expect to be able to assign a probability to any possible combination of tickets. This assumption implies that $\mathcal{E} = \mathcal{P}(X)$ (though not necessarily that $\mathfrak{D} = \mathcal{P}(\Omega)$). Moreover, we expect to be able to calculate the probability of an arbitrary event by some sort of summing over the individual tickets.

A critical inspection of this example leads us to two important observations:

A If we want to have a probability theory which describes a fair lottery on $\mathbb{N}$, assigns a probability to all singletons of $\mathbb{N}$, and follows a generalized additivity rule as well as the Normalization Axiom, the probability has to be non-zero but smaller than any finite, strictly positive real number. Hence, the range of P has to include infinitesimals. In other words, the range of P has to be a subset of a non-Archimedean field. Therefore, it cannot be $\mathbb{R}^+$, but it could be a non-standard set such as $\mathbb{Q}^{*,+}$ or $\mathbb{R}^{*,+}$, which are known from NSA.

B Our intuitions regarding infinite concepts are fed by our experience with their finite counterparts. So if we need to extrapolate the intuitions concerning finite lotteries to infinite ones, we need to introduce a sort of limit-operation which

---

[4] An investigation of our intuitions concerning finite and infinite lotteries can be found in Chapter 2.

transforms 'extrapolations' into 'limits'. Clearly, this operation cannot be the limit of classical analysis. Because the latter is used in Kolmogorov's Finite Additivity, axiom (K4) is suspect.

Motivated by the case study of a fair infinite lottery, we know at this point which elements in Kolmogorov's classical axiomatization we do not accept: the use of $[0,1]_\mathbb{R}$ as the range of the probability function and the application of classical limits in the Continuity Axiom. However, we have not yet offered an alternative to his approach: this is what we present in the next section.

6.2.2 Solution strategy

In order to solve one particular element of the problems mentioned, it may suffice to tinker with a single axiom. For instance, in order to be able to describe a fair lottery on the natural numbers, one seems to have the following options:

- Drop the axiom of Normalization (K2). This solution was explored by Rényi (1955).

- Drop the axiom of CA (K4). This solution was advocated by de Finetti (1974).

- Change (K0B), replacing $\mathbb{R}$ by $\mathbb{R}^*$, in order to allow infinitesimals in the range of the probability function. This was suggested by (Skyrms, 1980, Lewis, 1986a, Elga, 2004).[5]

Note that none of the above solutions generalizes to allow for the description of probabilistic problems on sample spaces of larger cardinalities as well (*e.g.* a fair lottery on $\mathbb{R}$).

Likewise, if some other of the previously discussed problems was our only concern, we could fix it by adjusting one of the axioms. However, from Figure 6.1, we see that *all* of Kolmogorov's axioms are involved in some of the epistemological problems mentioned. In particular, this chart shows that for cases dealing with infinite sample spaces, the main problem comes from the *combination* of the choice of the range of P with the axiom of CA (K4). Therefore, if we want to cure all symptoms simultaneously, we should formulate an entire, new set of axioms on which to base our theory of probability. Here, we list our main requirements for the new axioms:

- Whereas (K0A) allows for a domain which is strictly smaller than the full power set, it would be convenient if every set in $\mathcal{P}(\Omega)$ represents an event. For finite problems, the maximum number of different values taken by the function P depends on the problem (in particular, the size of the sample space Ω). Therefore, instead of fixing the range in advance, we will allow the range to depend on Ω for the finite as well as the infinite case.

---

[5]On closer inspection, we see that this approach also implies that (K4) has to be dropped, as we saw in Chapter 2.
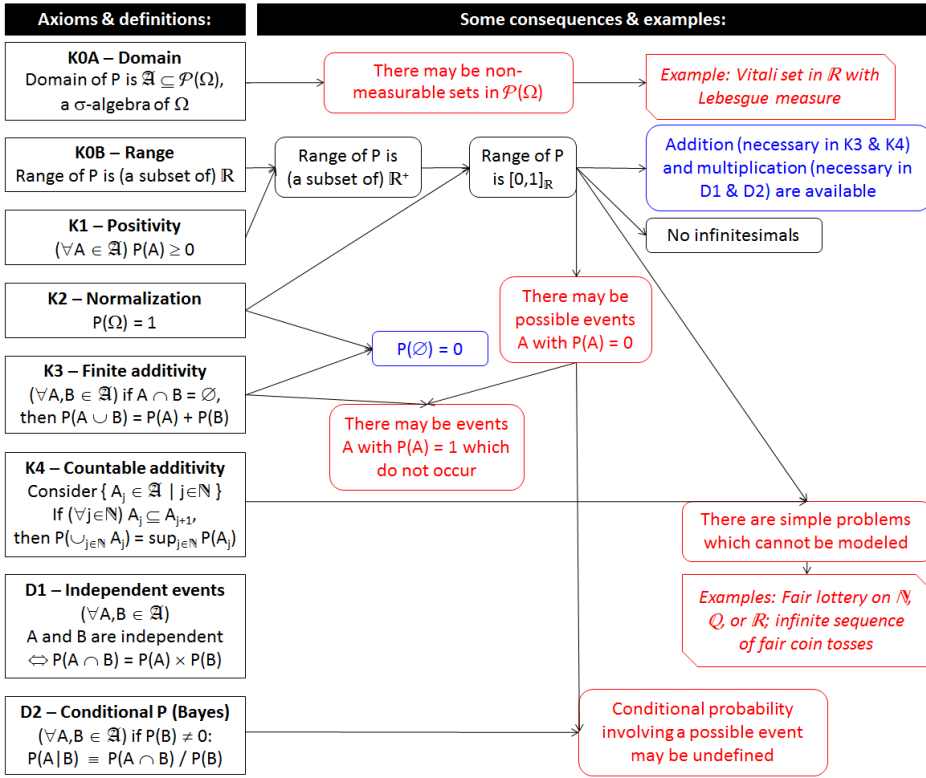
Figure 6.1: *Schematic representation of Kolmogorov's axioms and definitions on the left hand side and their consequences (with some examples) on the right hand side. Positive features are indicated in blue, negative properties in red.*

- We intend to replace (K1) by a formalization of Regularity. We should replace (K2) in a similar way, so as to ascertain that the interpretation of $P = 1$ events is also intuitively clear.

- The axiom of Finite Additivity (K3) is involved in one problematic issue, but does not seem to be the main culprit. So far, it seems as though we can keep this principle intact.

- Our replacement for (K4) will be the most drastic removal from Kolmogorov's approach. Because (K4) implies the use of classical limits, it is incompatible with our planned introduction of non-standard numbers in the range of P (*cf.* Chapter 2). Our new axiom will imply a limit operation of a different kind. In the case of a denumerably infinite sample space, this limit turns out to be the α-limit, as defined by Benci and Di Nasso (2003a) in the context of Alpha-Theory—an axiomatic approach to non-standard analysis (NSA). In the case of a non-denumerably infinite sample space, a generalization of the ideas

underlying the α-limit will have to be derived.

We will translate the above considerations into axioms in the next section.

6.2.3 Axioms for Non-Archimedean Probability (NAP)

Because the range of the probability function may contain infinitesimals in our approach, we call it Non-Archimedean Probability (NAP).

These are the Axioms of NAP, where Ω is a set called the sample space:

(NAP0) The probability function P has as its domain the full powerset of Ω (event space $= \mathcal{P}(\Omega)$) and as its range the unit interval of a suitable, ordered field F (range $= [0,1]_F$).

(NAP1) $P(A) = 0 \Leftrightarrow A = \varnothing$

(NAP2) $P(A) = 1 \Leftrightarrow A = \Omega$

(NAP3) For all events $A, B \in \mathcal{P}(\Omega)$ such that $A \cap B = \varnothing$:

$$P(A \cup B) = P(A) + P(B).$$

(NAP4) There exists a directed set $\langle \Lambda, \subseteq \rangle$[6] with $\Lambda \subseteq \mathcal{P}_{fin}(\Omega)$[7] such that:

$$\forall E \in P_{fin}(\Omega), \exists \lambda \in \Lambda : E \subset \lambda.\text{[8]}$$

When the sample space Ω is finite, the NAP-axioms have no advantage over Kolmogorov's axioms. This should not be surprising, as they were intended to overcome difficulties that arise only when considering problems on infinite sample spaces. We see that the NAP-axioms indeed fulfill the desiderata of the previous subsection:

- The axiom (NAP0) ensures that every set in $\mathcal{P}(\Omega)$ represents an event. Whereas Kolmogorov fixes the range of all probability functions (as $[0,1]_\mathbb{R}$), in our case, the range depends on the problem (in particular, the sample space Ω).

- The pair (NAP1) & (NAP2) makes it possible to interpret probability 0 and 1 events safely as impossible and necessary, respectively: they formalize Regularity and Regularity†, respectively.

- The axiom (NAP3) is exactly the same as Kolmogorov's addition-rule for events (finite additivity).

- The axiom (NAP4) implies the use of a generalized limit (direct limit), just as the Continuity Axiom implies the use of classical limits.

---

[6]$\langle \Lambda, \subseteq \rangle$ is a directed set if and only if Λ is a non-empty set and $\subseteq$ is a preorder such that every pair of elements of Λ has an upper bound: $(\forall A, B \in \Lambda) \exists C \in \Lambda : A \cup B \subseteq C$.

[7]$\mathcal{P}_{fin}(\Omega)$ denotes the family of *finite* subsets of Ω.

[8]This condition implies that $\bigcup_{\lambda \in \Lambda} \lambda = \Omega$.

Collectively, the axioms may force the range $[0,1]_F$ to be non-Archimedean. In particular, when the sample space Ω is countably infinite and the odds are fair, or when Ω is uncountably infinite, (NAP1) forces F to be a non-Archimedean field.[9]

In the case of a denumerably infinite sample space Ω, the limit implied by (NAP4) can be the axiomatically defined α-limit. A non-standard model for this concept can be obtained in terms of maximal ideals or free ultrafilters; the use of the directed set $\langle \Lambda, \subseteq \rangle$ has the advantage that unlike the other two options, Λ can be stated explicitly and can be chosen such as to model 'the physics' of the problem.

There is another analogy between Kolmogorov's axiom (K4) and the NAP-axiom (NAP4). Axiom (K4) allows you to extend the probability to a σ-algebra once you have defined the probability on a smaller family of sets. For example, if you want to define a uniform probability on $[0,1]_\mathbb{R}$, you start by defining the probability on the family of semi-open intervals $(\forall a > b \in [0,1]_\mathbb{R})\ P\big([a,b)\big) = b - a$. Then you extend the set of semi-open intervals to the σ-algebra generated by them. The Axiom (NAP4) has the same role: if you want to define a probability on $\mathcal{P}(\Omega)$, you start by defining the probability on a suitable family of finite sets first, and then you extend it to the whole power set.

The NAP axioms can be shown to be consistent by giving a model for them. The construction of such a model is a mathematical matter rather than a philosophical one, so it is not presented here. Once we have a working system, we can apply the new approach to all kinds of problems: countable lotteries (fair lottery on $\mathbb{N}$ or $\mathbb{Q}$), uncountable ones (fair lottery on $\mathbb{R}$), and infinite sequences of tosses with a fair coin ($2^\mathbb{N}$). This means that we can express the probability of a particular outcome of an infinite sequence of coin tosses as an infinitesimal, *pace* Williamson (2007). The development of these examples, however, goes beyond the scope of the outlook-section.

The further development of Non-Archimedean Probability and its application to examples known in the philosophy of probability are the topics of future work. Even in its preliminary form, however, the NAP system clearly illustrates that the analysis for a fair lottery on $\mathbb{N}$ given in Chapter 2 can be generalized to any lottery (fair or otherwise) on any sample space (countable or otherwise). Put differently, Non-Archimedean Probability shows that if we want to incorporate the advantages of qualitative approaches to probability into a quantitative framework, we can do so by allowing infinitesimals in the range of the probability function. As such, it is my hope that this dissertation can be regarded as a new step towards a better framework for the philosophy of probability.

---

[9]A fair lottery is the simplest but not the only example which leads to a non-Archimedean F in the case of a countably infinite Ω: the same is true if the winning odds of a ticket increase with the ticket number, and more advanced examples can be construed.

Bibliography

M. Abrams. Mechanistic probability. Accepted in *Synthese*, DOI: 10.1007/s11229-010-9830-3. 16

P. Achinstein. *The Book of Evidence*. Oxford University Press, Oxford, UK, 2003. 68, 69

D. Adams. *The Restaurant at the End of the Universe*. Pan Books, London, UK, 1980. Reprinted as second story in *The Ultimate Hitchhiker's guide* by Wings Books, New York, NY, 1996. 39

E. W. Adams. Aspects of inductive logic. In J. Hintikka and P. Suppes, editors, *Probability and the Logic of Conditionals*, pages 253–316. North-Holland, Amsterdam, The Netherlands, 1966. 81, 133

B. Allen, 2010. URL http://www.boallen.com/random-numbers.html. 19

I. Asimov. *Prelude to the Foundation*. Doubleday, New York, NY, 1988. 99

R. J. Aumann. Agreeing to disagree. *Annals of Statistics*, 4:1236–1239, 1976. 26

B. Barry. *Theories of Justice*. University of California Press, Los Angeles, CA, 1989. 26

P. Bartha. Countable additivity and the de Finetti lottery. *The British Journal for Philosophy of Science*, 55:301–321, 2004. 34, 35, 40

P. Bartha and C. Hitchcock. The shooting-room paradox and conditionalizing on measurably challenged sets. *Synthese*, 118:403–437, 1999. 35

P. Bartha and R. Johns. Probability and symmetry. *Philosophy of Science*, 68: S109–S122, 2001. 34

V. Benci and M. Di Nasso. Alpha-theory: an elementary axiomatic for nonstandard analysis. *Expositiones Mathematicae*, 21:355–386, 2003a. 30, 48, 56, 78, 136

V. Benci and M. Di Nasso. Numerosities of labelled sets: a new way of counting. *Advances in Mathematics*, 173:50–67, 2003b. 30, 31, 50, 51, 57, 95

V. Benci, M. Di Nasso, and M. Forti. The eightfold path to nonstandard analysis. In N. J. Cutland, M. Di Nasso, and D. A. Ross, editors, *Nonstandard Methods and Applications in Mathematics*, volume 25 of *Lecture Notes in Logic*, pages 3–44. Association for Symbolic Logic, AK Peters, Wellesley, MA, 2006a. 45, 77

V. Benci, M. Di Nasso, and M. Forti. An Aristotelian notion of size. *Annals of Pure and Applied Logic*, 143:43–53, 2006b. 31

V. Benci, S. Galatolo, and M. Ghimenti. An elementary approach to stochastic differential equations using infinitesimals. In V. Bergelson, A. Blass, M. Di Nasso, and R. Jin, editors, *Ultrafilters across Mathematics*, volume 530 of *Contemporary Mathematics*. American Mathematical Society, 2010. URL http://arxiv.org/abs/0807.3477. 58

A. T. Bernardes, U. M. S. Costa, A. D. Araujo, and D. Stauffer. Damage spreading, coarsening dynamics and distribution of political votes in Sznajd model on square lattice. *International Journal of Modern Physics C*, 12:159–167, 2001. 101

A. T. Bernardes, D. Stauffer, and J. Kertész. Election results and the Sznajd model on Barabasi network. *The European Physical Journal B*, 25:123–127, 2002. 103

J. L. F. Bertrand. *Calcul des Probabilités*. Gauthier-Villars et fils, Paris, France, 1888. 9, 10, 18, 134

E. Beth. Semantics of physical theories. In H. Freudenthal, editor, *The Concept and the Role of the Model in Mathematics and Natural and Social Sciences*. D. Reidel, Dordrecht, The Netherlands, 1961. 5

L. Bienvenu and W. Merkle. Effective randomness for computable probability measures. *Electronic Notes in Theoretical Computer Science*, 167:117–130, 2007. 20

L. Bienvenu, G. Shafer, and A. Shen. On the history of martingales in the study of randomness. *Electronic Journal for History of Probability and Statistics*, 5:1–40, 2009. URL http://www.jephs.net. 20

W. Blake. *The Marriage of Heaven and Hell*. 1790. Reprinted by Dover, Mineola, NY, 1994. 63

R. Booth and E. Richter. On revising fuzzy belief bases. *Studia Logica*, 80:29–61, 2005. 74

R. Brüggeman, G. Restrepo, and K. Voigt. Prescreening of chemicals by partial orders. In *ICCOMP'05 Proceedings of the 9th WSEAS International Conference on Computers*, 2005. 26

M. Burock. An outcome of the de Finetti infinite lottery is not finite. Unpublished manuscript, 2006. URL http://philsci-archive.pitt.edu/archive/00002605/01/infinite_lottery.pdf. 34

M. Buyse, S. L. George, S. Evans, N. L. Geller, J. Ranstam, B. Scherrer, E. Lesaffre, G. Murray, L. Edler, J. Hutton, T. Colton, P. Lachenbruch, and B. L. Verma. The role of biostatistics in the prevention, detection and treatment of fraud in clinical trials. *Statistics in Medicine*, 18:3435–3451, 1999. 23, 24

D. Byrne. Interpersonal-attraction and attitude similarity. *Journal of Abnormal and Social Psychology*, 62:713–715, 1961. 102

G. Cantor. Über eine Eigenschaft des Inbegriffes aller reellen algebraischen Zahlen. *Journal für die reine und angewandte Mathematik*, 77:258–262, 1874. 29

G. Cantor. Über eine elementare Frage der Mannigfaltigkeitslehre. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 1:75–78, 1891. 29

R. Carnap. *Logical Foundations of Probability*. University of Chicago Press, Chicago, IL, 1950. 7

L. Carroll. *Through the Looking-Glass, and What Alice Found There*. Macmillan, London, UK, 1871. 91

B. Carter. The anthropic principle and its implications for biological evolution. *Philosophical Transactions of the Royal Society of London A*, 310:347–363, 1983. 35

G. J. Chaitin. On the difficulty of computations. *IEEE Transactions on Information Theory*, IT-16:5–9, 1970. Reprinted in G. J. Chaitin *Information, Randomness & Incompleteness: Papers on Algorithmic Information Theory*, Series in Computer Science, Volume 8, pages 14–22, World Scientific, Singapore, 1987. 25

G. J. Chaitin. Randomness and mathematical proof. *Scientific American*, 232:47–52, 1975. Reprinted in G. J. Chaitin *Information, Randomness & Incompleteness: Papers on Algorithmic Information Theory*, Series in Computer Science, Volume 8, pages 1–13, World Scientific, Singapore, 1987. 20, 25

G. J. Chaitin. *Exploring Randomness*. Springer-Verlag, London, UK, 2001. URL http://www.cs.auckland.ac.nz/~chaitin/ait/index.html. 20

John Ciardi. *Manner of Speaking*. Rutgers University Press, New Brunswick, NJ, 1972. 129

R. D. Cousins. Why isn't every physicist a Bayesian? *American Journal of Physics*, 63:398–410, 1995. 13

R. Crawshay-Williams. *Methods and Criteria of Reasoning*. Routledge & P. Kegan, London, UK, 1957. Reprinted by The International Library of Philosophy, 2001. 86

N. Cutland. Nonstandard measure theory and its applications. *Bulletin of the London Mathematical Society*, 15:529–589, 1983. 45, 57, 61

N. Dalkey and O. Helmer. An experimental application of the Delphi method to the use of experts. *Management Science*, 9:458–467, 1963. 103

B. de Finetti. *Theory of Probability*. Wiley, London, UK, 1974. Translated by: A. Machí and A. Smith. 2, 10, 34, 35, 36, 40, 44, 45, 71, 92, 134, 135

R. Dedekind. *Was sind und was sollen die Zahlen?* Vieweg, Braunschweig, Germany, 1888. 29

G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch. Mixing beliefs among interacting agents. *Advances in Complex Systems*, 3:87–98, 2000. 101, 102, 108

M. H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69:118–121, 1974. 102

S. Dehaene, E. Spelke, P. Pinel, R. Stanescu, and S. Tsivkin. Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science*, 284:970–974, 1999. 84

S. Dehaene, V. Izard, E. Spelke, and P. Pica. Log or linear? Distinct intuitions of the number scale in Western and Amazonian Indigene cultures. *Science*, 320: 1217–1220, 2008. 84

A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics*, 38:325–339, 1967. 12, 74

K. DeRose. *The Case for Contextualism: Knowledge, Skepticism, and Context*. Oxford University Press, Oxford, UK, 2009. 86

R. Descartes. *Principia Philosophiae*. Elsevier, Amsterdam, The Netherlands, 1644. 31

M. Di Nasso and M. Forti. Numerosities of point sets over the real line. *Transactions of the American Mathematical Society*, 362:5355–5371, 2010. 31

P. Diaconis, S. Holmes, and R. Montgomery. Dynamical bias in the coin toss. *Society of Industrial and Applied Mathematics Review*, 49:211–235, 2007. 17, 23

Z. Domotor. Mathematical models in philosophy of science. In N. J. Smelser and P. B. Baltes, editors, *International Encyclopedia of the Social & Behavioral Sciences*, pages 9399–9405. Elsevier, Oxford, UK, 2001. 5

I. Douven. The lottery paradox and our epistemic goal. *Pacific Philosophical Quarterly*, 89:204–225, 2008. 36

I. Douven. Simulating peer disagreements. *Studies in History and Philosophy of Science*, 41:148–157, 2010. 102

I. Douven and L. Decock. Identity and similarity. *Philosophical Studies*, 151:59–78, 2009. 86

I. Douven and C. Kelp. Truth approximation, social epistemology, and opinion dynamics. *Erkenntnis*. In press. 100

I. Douven and A. Riegler. Extending the Hegselmann–Krause model I. *Logic Journal of the IGPL*, 18:323–335, 2010. 102

I. Douven and J. Uffink. The preface paradox revisited. *Erkenntnis*, 59:389–420, 2003. 37, 68

I. Douven and T. Williamson. Generalizing the lottery paradox. *British Journal for the Philosophy of Science*, 57:755–779, 2006. 36, 64

I. Douven, L. Horsten, and J.-W. Romeijn. Probabilist anti-realism. *Pacific Philosophical Quarterly*, 91:38–63, 2010. 43

R. M. Dudley. *Real Analysis and Probability*. Cambridge University Press, Cambridge, UK, 2004. 59

C. Durtshi, W. Hillison, and C. Pacini. The effective use of Benford's law to assist in detecting fraud in accounting data. *Journal of Forensic Accounting*, V:17–34, 2004. 23

K. Easwaran. Regularity and infinitesimal credences. Unpublished manuscript, 2010. 43, 133

W. Eckhardt. A shooting-room view of Doomsday. *Journal of Philosophy*, XCIII: 244–259, 1997. 35

P. Égré. Soritical series and Fisher series. In A. Hieke and H. Leitgeb, editors, *Reduction: Between the Mind and the Brain*, volume 12 of *Publications of the Austrian Ludwig Wittgenstein Society*, pages 91–115. Ontos Verlag, Frankfurt am Main, Germany, 2009. 68

A. Elga. Infinitesimal chances and the laws of nature. *Australasian Journal of Philosophy*, 82:67–76, 2004. 56, 135

H. Everett. Relative state formulation of quantum mechanics. *Reviews of Modern Physics*, 29:454–462, 1957. 16

R. Foley. Justified inconsistent beliefs. *American Philosophical Quarterly*, 16:247–257, 1979. 64

R. Foley. Beliefs, degrees of belief, and the Lockean thesis. In F. Huber and C. Schmidt-Petri, editors, *Degrees of Belief*, volume 342 of *Synthese Library*, pages 37–47. Springer, Dordrecht, The Netherlands, 2009. 65, 67

S. Fortunato, V. Latora, A. Pluchino, and A. Rapisarda. Vector opinion dynamics in a bounded confidence consensus model. *International Journal of Modern Physics C*, 10:1535–1551, 2005. 102

M. Friend. *Introducing Philosophy of Mathematics*. Acumen Publishing, Stockfield, UK, 2007. 27

P. Gács. *Komplexität und Zufälligkeit*. PhD thesis, Johann Wolfgang Goethe Universität, Frankfurt am Main, Germany, 1978. 19

S. Galam. Minority opinion spreading in a random geometry. *The European Physical Journal B*, 25:403–406, 2002. 100, 109

G. Gamow. *One, two, three . . . infinity: facts & speculations of science*. Viking Press, New York, NY, 1947. 30, 91

C. F. Gauss and H. C. Schumacher. *Briefwechsel zwischen C. F. Gauss und H. C. Schumacher*, volume 2. Esch, Altona, Germany, 1860. C. A. F. Peters, editor. 27

M. Gazalé. *Number; from Ahmes to Cantor*. Princeton University Press, Princeton, NJ, 2000. 29

A. Gelman and C. R. Shalizi. Philosophy and the practice of Bayesian statistics. 2010. URL http://arxiv.org/abs/arXiv:1006.3868. 13

E. Gibbon. *The History of the Decline and Fall of the Roman Empire*, volume VIII. W. Y. Birch & A. Small, Philadelphia, PA, 1805. 25

R. E. Goodin. Uncertainty as an excuse for cheating our children: The case of nuclear wastes. *Policy Sciences*, 10:25–43, 1978. 15, 24

M. Haahr, 1998–2010. URL http://www.random.org. 22

A. Hájek. What conditional probability could not be. *Synthese*, 137:273–323, 2003. 85

A. Hájek. Interpretations of probability. In N. Zalta, editor, *Stanford Encyclopedia of Philosophy*. 2007. URL http://plato.stanford.edu/entries/probability-interpret/. 9

A. Hájek. A philosopher's guide to probability. In G. Bammer and M. Smithson, editors, *Uncertainty and Risk: Multidisciplinary Perspectives*. Earthscan, London, UK, 2008. 9

A. Hájek. Staying regular. Unpublished manuscript, 2010. 43, 133

H. W. Hake and R. Hyman. Perception of the statistical structure of a random series of binary symbols. *Journal of Experimental Psychology*, 45:64–74, 1953. 19

J. D. Hamkins. Infinite time Turing machines. *Minds and Machines*, 12:521–539, 2002. 28

R. W. Hamming. Error detecting and error correcting codes. *The Bell System Technical Journal*, 29:147–160, 1950. 108

G. Harman. *Change in View: Principles of Reasoning.* MIT Press, Cambridge, MA, 1986. 85

S. Hartmann. Modeling in philosophy of science. In M. Frauchiger and W. K. Essler, editors, *Representation, Evidence, and Justification: Themes from Suppes*, volume 1 of *Lauener Library of Analytical Philosophy*, pages 95–121. Ontos Verlag, Frankfurt am Main, Germany, 2008. 7

S. Hartmann and J. Sprenger. Bayesian epistemology. In S. Bernecker and D. Pritchard, editors, *Routledge Companion to Epistemology.* Routledge, London, UK, 2010. 7

J. Hawthorne and L. Bovens. The preface, the lottery, and the logic of belief. *Mind*, 108:241–264, 1999. 81

B. Hayes. Randomness as a resource. *American Scientist*, 89:300–304, 2001. 22, 25

P. Hedström and P. Bearman, editors. *The Oxford Handbook of Analytical Sociology.* Oxford University Press, Oxford, UK, 2009. 100

R. Hegselmann and U. Krause. Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of Artifical Societies and Social Simulation*, 5, 2002. URL http://jasss.soc.surrey.ac.uk/5/3/2.html. 7, 101, 102, 103, 108

R. Hegselmann and U. Krause. Opinion dynamics driven by various ways of averaging. *Computational Economics*, 25:381–405, 2005. 102, 108

R. Hegselmann and U. Krause. Truth and cognitive division of labor: First steps towards a computer aided social epistemology. *Journal of Artificial Societies and Social Simulation*, 9, 2006. URL http://jasss.soc.surrey.ac.uk/9/3/10.html. 102, 109

R. Hegselmann and U. Krause. Deliberative exchange, truth, and cognitive division of labor: A low-resolution modeling approach. *Episteme*, 6:130–144, 2009. 102, 109

C. G. Hempel. Turns in the evolution of the problem of induction. *Synthese*, 46: 389–404, 1981. 12

D. Hilbert. Über das Unendliche. *Mathematische Annalen*, 95:161–190, 1926. 27

R. A. Hill and R. I. M. Dunbar. Social network size in humans. *Human Nature*, 14: 53–72, 2003. 104

L. Horsten and I. Douven. Formal methods in the philosophy of science. *Studia Logica*, 89:151–162, 2008. 5

K. Hrbacek. Stratified analysis? In I. van den Berg and N. Neves, editors, *The Strength of Nonstandard Analysis*, pages 47–63. Springer, Vienna, Austria, 2007. 64, 71, 74, 75, 96

K. Hrbacek, O. Lessmann, and R. O'Donovan. Analysis with ultrasmall numbers. *The American Mathematical Monthly*, 117:801–816, 2010. 64, 74, 76, 81

D. Hume. *A Treatise of Human Nature*. 1739–1740. Reprinted by Oxford University Press, Oxford, UK, 2000. 12

D. Jacobmeier. Multidimensional consensus model on a Barabási–Albert network. *International Journal of Modern Physics C*, 16:633–646, 2005. 102

E. T. Jaynes. Information theory and statistical mechanics. *Physical Review*, 106: 620–630, 1957. 10, 34, 134

E. T. Jaynes. The well-posed problem. *Foundations of Physics*, 3:477–493, 1973. 10, 18, 134

J. B. Kadane, M. J. Schervish, and T. Seidenfeld. Statistical implications of finitely additive probability. In P. K. Goel and A. Zellnder, editors, *Bayesian Inference and Decision Techniques*. Elsevier, Amsterdam, The Netherlands, 1986. 34

A. Katz. Process design for selection of hemodialysis and organ transplant recipients. *Buffalo Law Review*, 22:373–418, 1973. 26

S. Keeney, F. Hasson, and H. McKenna. *The Delphi Technique in Nursing and Health Care Research*. Wiley-Blackwell, West Sussex, UK, 2011. 103

K. T. Kelly. *The Logic of Reliable Inquiry*. Oxford University Press, Oxford, UK, 1996. 12, 34, 35, 92

J. M. Keynes. *A Treatise on Probability*. Macmillan, New York, NY, 1921. 9

J. F. Kilner. A moral allocation of scarce lifesaving medical resources. *Journal of Religious Ethics*, 9:245–285, 1981. 26

P. Klein. The virtues of inconsistency. *Monist*, 68:105–135, 1985. 64

A. N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitrechnung*. Ergebnisse der Mathematik. 1933. Translated by N. Morrison, *Foundations of probability*. Chelsea Publishing Company, 1956 (2nd ed.). 2, 31, 32, 40, 41, 59, 62, 71, 92, 131

P. Komjáth and V. Totik. Ultrafilters. *American Mathematical Monthly*, 115:33–44, 2008. 46

L. A. Kornhauser and L. G. Sager. Unpacking the court. *Yale Law Journal*, 96: 82–117, 1986. 109

J. Krhovják. Analysis, demands, and properties of pseudorandom number generators. In *Mikulášská Kryptobesídka*, pages 55–65, Prague, Czech Republic, 2006. 22

T. A. F. Kuipers, editor. *What is Closer-to-the-Truth?: A Parade of Approaches to Truthlikeness*, volume 10 of *Poznań Studies in the Philosophy of the Sciences and the Humanities*. Rodopi, Amsterdam, The Netherlands, 1987. 7

H. E. Kyburg, Jr. *Probability and the Logic of Rational Belief.* Wesleyan University Press, Middletown, CT, 1961. 36, 64, 90, 93

A. Lang. The limited capacity model of mediated message processing. *Journal of Communication*, 50:46–70, 2000. 68

P.-S. Laplace. *Essai Philosophique sur les Probabilités.* Paris, France, 1814. 3th edition printed by V. Courcier, Paris, France, 1816. 1, 9, 18, 19, 33, 134

B. Latour and S. Woolgar. *Laboratory Life: the Social Construction of Scientific Facts.* Sage Publications, Los Angeles, CA, 1979. 5

P. R. Laughlin, B. L. Bonner, and A. G. Miner. Groups perform better than the best individuals on letters-to-numbers problems. *Organizational Behavior and Human Decision Processes*, 88:605–620, 2002. 100

L. Lauwers. Purely finitely additive measures are non-constructible objects. Working paper, unpublished. Retrieved: May 29, 2010. URL http://www.econ.kuleuven.be/ces/discussionpapers/Dps10/DPS1010.pdf. 56

S. Lavine. Finite mathematics. *Synthese*, 103:389–420, 1995. 27, 39, 70, 75, 80, 94

D. Lehmann and M. Magidor. What does a conditional knowledge base entail? *Artificial Intelligence*, 55:1–60, 1992. 133

K. Lehrer and C. Wagner. *Rational Consensus in Science and Society.* D. Reidel, Dordrecht, The Netherlands, 1981. 102

H. Leitgeb. Reducing belief simpliciter to degrees of belief. Unpublished manuscript, 2010. 66

J. Leslie. *The End of the World: The Science and Ethics of Human Extinction.* Routledge, London, UK, 1998. 35

M. J. Levy, Jr. *Levy's Eleven Laws of the Disillusionment of the True Liberal.* Princeton University Press, Princeton, NJ, 1981. 5th ed. 39

D. K. Lewis. A subjectivist's guide to objective chance. In R. C. Jeffrey, editor, *Studies in Inductive Logic and Probability.* University of California Press, Berkeley, CA. 43, 65, 92

D. K. Lewis. *Convention: a Philosophical Study.* Harvard University Press, Cambridge, MA, 1969. Reprinted by Wiley-Blackwell, Oxford, UK, 2002. 26

D. K. Lewis. *Philosophical Papers*, volume II. Oxford University Press, Oxford, UK, 1986a. 16, 61, 133, 135

D. K. Lewis. *Philosophical Papers*, volume II, chapter A Subjectivist's Guide to Objective Chance. Oxford University Press, Oxford, UK, 1986b. 10

H. A. Linstone and M. Turoff, editors. *The Delphi Method: Techniques and Applications*. Addison-Wesley, Reading, MA, 1975. 108

C. List. The probability of inconsistencies in complex collective decisions. *Social Choice and Welfare*, 24:3–32, 2005. 110, 112

C. List and C. Puppe. Judgment aggregation: A survey. In P. Anand, P. Pattaniak, and C. Puppe, editors, *The Handbook of Rational and Social Choice*, pages 457–482. Oxford University Press, Oxford, UK, 2007. 109, 110

J. Lorenz. A stabilization theorem for dynamics of continuous opinions. *Physica A: Statistical Mechanics and its Applications*, 355:217–223, 2005. 7, 100

J. Lorenz. Fostering consensus in multidimensional continuous opinion dynamics under bounded confidence. In D. Helbing, editor, *Managing Complexity: Insights, Concepts, Applications*, pages 321–334. Springer, Berlin, Germany, 2008. 102, 104

M. Macy and A. Flache. Social dynamics from the bottom up. Agent-based models of social interaction. In P. Hedström and P. Bearman, editors, *The Oxford Handbook of Analytical Sociology*, pages 245–268. Oxford University Press, Oxford, UK, 2009. 100

M. Macy and R. Willer. From factors to actors: Computational sociology and agent-based modeling. *Annual Review of Sociology*, 28:143–166, 2002. 100

D. Makinson. The paradox of the preface. *Analysis*, 25:205–207, 1965. 37, 70

P. Mancosu. Measuring the size of infinite collections of natural numbers. *The Review of Symbolic Logic*, 2:612–646, 2009. 28, 30, 31, 57

P. Martin-Löf. The definition of random sequences. *Information and Control*, 9: 602–619, 1966. 20

M. Mäs. *The Diversity Puzzle. Explaining Clustering and Polarization of Opinions*. PhD thesis, University of Groningen, Faculty of Behavioural and Social Sciences, Department of Sociology, Groningen, The Netherlands, 2010. 100, 103

M. Matsumoto and T. Nishimura. Mersenne Twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation*, 8:3–30, 1998. 22

S. McCall and D. M. Armstrong. God's lottery. *Analysis*, 49:223–224, 1989. 40

H. L. Mencken. *Prejudices: First Series*. 1919. Reprinted by Wildside Press LLC, Rockville, MD, 2005. 67

N. Metropolis. Beginning of the Monte Carlo method. *Los Alamos Science*, pages 125–130, 1987. 22

H. Milbrodt. *Wahrscheinlichkeitstheorie – Eine Einführung mit Anwendungen und Beispielen aus der Versicherungs- und Finanzmathematik.* Verlag Versicherungswirtsch, Karlsruhe, Germany, 2010. 17

B. Nagle, V. Rödl, and M. Schacht. The counting lemma for regular k-uniform hypergraphs. *Random Structures & Algorithms*, 28:113–179, 2006. 21

D. K. Nelkin. The lottery paradox, knowledge, and rationality. *The Philosophical Review*, 109:373–409, 2000. 64

E. Nelson. Internal set theory: a new approach to nonstandard analysis. *Bulletin of the American Mathematical Society*, 83:1165–1198, 1977. 71

A. Niés. *Computability and Randomness.* Clarendon Press, Oxford, UK, 2008. 21

M. Nigrini. A taxpayer compliance application of Benford's law. *Journal of the American Tax Association*, 18:72–91, 1996. 23

G. R. Oppy. *Philosophical Perspectives on Infinity.* Cambridge University Press, Cambridge, UK, 2006. 30, 36

G. Owen and B. Grofman. To vote or not to vote: the paradox of nonvoting. *Public Choice*, 42:311–325, 1984. 89

P. Pagin. Vagueness and domain restriction. 2010. URL http://people.su.se/~ppagin/papers/vagdom4.pdf. 96

A. Pedeferri and M. Friend. Formalism or pluralism; which better describes mathematical practice? Abstract presented at 'Logic Colloquium 2010' at Paris, France, 2010. URL http://www.logique.jussieu.fr/~malod/submissions. 5

P. Péraire and G. Wallet. Une théorie relative des ensembles internes. *Comptes rendus de l'Académie des sciences. Série 1, Mathématique*, 308:301–304, 1989. 75

P. Pettit. Deliberative democracy and the discursive dilemma. *Philosophical Issues*, 11:268–299, 2001. 109, 110

D. Phan and F. Varenne. Agent-based models and simulations in economics and social sciences: from conceptual exploration to distinct ways of experimenting. *Journal of Artificial Societies and Social Simulation*, 13, 2010. URL http://jasss.soc.surrey.ac.uk/13/1/5.html. 7

A. Pluchino, V. Latora, and A. Rapisarda. Compromise and synchronization in opinion dynamics. *The European Physical Journal B*, 50:169–176, 2006. 102

K. R. Popper. A set of independent axioms for probability. *Mind*, 47:275–277, 1938. 133

K. R. Popper. *The Logic of Scientific Discovery.* Hutchinson, London, UK, 1959. Reprinted by Routledge Classics, London, UK, 2002. 12, 133

L. Price. *Dialogues of Alfred North Whitehead.* Little - Brown, Boston, MA, 1954. Reprinted by David R. Godine Publisher, Boston, MA, 2001. 63, 91

G. Priest. What is so bad about contradictions? *The Journal of Philosophy*, 95: 410–426, 1998. 64

H. Putnam. *Reason, Truth, and History.* Cambridge University Press, Cambridge, UK, 1981. 85

D. Ramirez-Cano and J. Pitt. Follow the leader: Profiling agents in an opinion formation model of dynamic confidence and individual mind-sets. In T. Nishida *et al.*, editor, *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT-06)*, pages 660–667. IEEE Computer Society Press, 2006. 102

F. P. Ramsey. Truth and probability. In R. B. Braithwaite, editor, *The Foundations of Mathematics and Other Logical Essays*, volume 5 of *International library of psychology, philosophy, and scientific method*, pages 156–198. Routledge & P. Kegan, London, UK, 1931. (Original paper from 1926). 40, 65

Corporation 'RAND'. *A Million Random Digits with 100,000 Normal Deviates.* Free Press, Glencou, IL, 1955. Reprinted by Rand, 2001. 23

J. Rawls. *A Theory of Justice.* Belknap Press, Cambridge, MA, 1999. Revised edition. 26

A. Rényi. On a new axiomatic theory of probability. *Acta Mathematica Hungarica*, 6:285–335, 1955. 35, 92, 135

A. Robinson. *Non-Standard Analysis.* North-Holland, Amsterdam, The Netherlands, 1966. 34, 45, 71, 95

K. H. Rosen. *Handbook of Discrete and Combinatorial Mathematics.* CRC Press, Boca Raton, FL, 2000. 124, 125

J. E. Rubio. *Optimization and Nonstandard Analysis.* Marcel Dekker, New York, NY, 1994. 53

R. Rucker. *Infinity and the Mind: The Science and Philosophy of the Infinite.* Princeton University Press, Princeton, NJ, 1982. 27

B. Russell. The philosophy of logical atomism. In J. G. Slater, editor, *Collected Papers of Bertrand Russell*, volume 8. Allen & Unwin, London, UK, 1986. (Original essay from 1918). 129

G. Schurz and H. Leitgeb. Finitistic and frequentistic approximation of probability measures with or without σ-additivity. *Studia Logica*, 89:257–283, 2008. 44, 45, 61, 95, 97

T. Seidenfeld. Why I am not an objective Bayesian; some reflections prompted by Rozenkrantz. *Theory and Decision*, 11:413–440, 1979. 10, 18

L. G. Shapiro and G. C. Stockman. *Computer Vision*. Prentice-Hall, Upper Saddle River, NJ, 2002. 68

M. Shermer. Patternicity. *Scientific American*, 299:48, 2008. 19

B. Skyrms. *Causal Necessity*. Yale University Press, New Haven, CT, 1980. 43, 92, 133, 135

W. Spohn. A survey of ranking theory. In F. Huber and C. Schmidt-Petri, editors, *Degrees of Belief*, volume 342 of *Synthese Library*. Springer, Dordrecht, The Netherlands, 2009. 12

J. Stanley and Z. G. Szabó. On quantifier domain restriction. *Mind & Language*, 15: 219–261, 2000. 96

D. Stauffer. The Sznajd model of consensus building with limited persuasion. *International Journal of Modern Physics C*, 13:315–317, 2002. 101, 102

W. Stegmüller. *The Structure and Dynamics of Theories*. Springer, New York, NY, 1976. 5

P. Stone. What can lotteries do for education? *Theory and Research in Education*, 6:267–282, 2008a. 26

P. Stone. What is a fair lottery? Working paper, 2008b. URL http://www.stanford.edu/~pstone/documents/What_Is_a_Fair_Lottery.pdf. 14, 18, 24, 26, 134

P. Suppes. *Introduction to Logic*. Van Nostrand, Princeton, NJ, 1957. 5

J. Surowiecki. *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. Doubleday, New York, NY, 2004. 100

E. Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arithmetica*, 27:199–245, 1975. 21

K. Sznajd-Weron and J. Sznajd. Opinion evolution in closed community. *International Journal of Modern Physics C*, 11:1157–1166, 2000. 100, 101

K. Sznajd-Weron and J. Sznajd. Who is left, who is right? *Physica A: Statistical Mechanics and its Applications*, 351:593–604, 2005. 101, 102

N. Taleb. *The Black Swan: The Impact of the Highly Improbable*. Random House, New York, NY, 2007. 12

C. Tavris and E. Aronson. *Mistakes Were Made (But Not by Me)*. Harcourt, Orlando, FL, 2007. 109

G. Tenenbaum. *Introduction to Analytic and Probabilistic Number Theory.* Cambridge University Press, Cambridge, UK, 1995. 40, 44

P. R. Thagard. *Computational Philosophy of Science.* MIT-press - Bradford Books, 1988. 6

P. R. Thagard. Societies of minds: science as distributed computing. *Studies in History and Philosophy of Science*, 24:49–67, 1993. 6

P. R. Thagard. Computation and the philosophy of science. In T. W. Bynum and J. H. Moor, editors, *The Digital Phoenix: How Computers are Changing Philosophy*, pages 48–61. Blackwell, Oxford, UK, 1998. 4, 6

J. Thomson. Tasks and super-tasks. *Analysis*, XV:1–13, 1954–55. Reprinted in W. Salmon (editor), *Zeno's paradoxes.* Bobbs-Merril, Indianapolis, IN, 1970. 28

J. R. R. Tolkien. *The Fellowship of the Ring: being the first part of The Lord of the Rings.* George Allen & Unwin, London, UK, 1954. Reprinted by HarperCollins, New York, NY, 2nd edition, 2001. 99

J. Truss. *Foundations of mathematical analysis.* Clarendon Press, Oxford, UK, 1997. 43

J. Uffink. Can the maximum entropy principle be explained as a consistency requirement? *Studies in History and Philosophy of Modern Physics*, 26:223–261, 1995. 134

J. Uffink. Compendium of the foundations of classical statistical physics. In J. Butterfield and J. Earman, editors, *Philosophy of Physics.* North-Holland - Elsevier, Amsterdam, The Netherlands, 2007. 4

S. Ulam, R. D. Richtmyer, and J. von Neumann. Statistical methods in neutron diffusion. *Los Alamos Scientific Laboratory Report LAMS*, 551, 1947. 22

B. C. van Fraassen. *Laws and Symmetry.* Clarendon Press, Oxford, UK, 1989. 13, 14

M. Väth. *Nonstandard Analysis.* Birkhäuser Verlag, Basel, Switserland, 2007. 48

A. Vilenkin. *Many worlds in One: The Search for other Universes.* Hill and Wang, New York, NY, 2006. 17

J. Ville. *Etude critique de la notion de collectif.* Gauthiers-Villars, Paris, France, 1939. 20

R. von Mises. Grundlagen der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 5:52–99, 1919. 20

R. von Mises. *Probability, Statistics and Truth.* George Allen & Unwin, London, UK, 1928. Reprinted by Dover, Mineola, NY, 1981. 20

J. von Neumann. Various techniques used in connection with random digits. In A. S. Householder, G. E. Forsythe, and H. H. Germond, editors, *Monte Carlo Method*, volume 12 of *National Bureau of Standards Applied Mathematics Series*, pages 36–38. U.S. Government Printing Office, Washington, D.C., 1951. 21

V. G. Vovk and G. R. Shafer. Kolmogorovs contributions to the foundations of probability. *Problems of Information Transmission*, 39:21–31, 2003. 21

G. Weisbuch, G. Deffuant, F. Amblard, and J. P. Nadal. Meet, discuss and segregate! *Complexity*, 7:55–63, 2002. 102

C. Werndl. Are deterministic descriptions and indeterministic descriptions observationally equivalent? *Studies In History and Philosophy of Modern Physics*, 40: 232–242, 2009. 11

G. Wheeler. A review of the lottery paradox. In W. L. Harper and G. Wheeler, editors, *Probability and Inference: Essays in Honour of Henry E. Kyburg, Jr*, pages 1–31. King's College Publications, London, UK, 2007. 90, 93

A. N. Whitehead. *The Concept of Nature.* Cambridge University Press, Cambridge, UK, 1920. 9

E. Wigner. The unreasonable effectiveness of mathematics in the natural sciences. *Communications on Pure and Applied Mathematics*, 13:1–14, 1960. 8

J. Williamson. Countable additivity and subjective probability. *The British Journal for Philosophy of Science*, 50:401–416, 1999. 34

T. Williamson. *Vagueness.* Routledge, London, UK, 1994. 87

T. Williamson. Imagination, stipulation and vagueness. *Philosophical Issues*, 8: 215–228, 1997. 87

T. Williamson. *Knowledge and its Limits.* Oxford University Press, Oxford, UK, 2002. 12, 15

T. Williamson. How probable is an infinite sequence of heads? *Analysis*, 67:173–180, 2007. 43, 133, 138

W. Wood. Attitude change: Persuasion and social influence. *Annual Review of Psychology*, 51:539–570, 2000. 100

Summary

Any philosophy that can be put 'in a nutshell' belongs there.
Sydney J. Harris—'Leaving the Surface'

In **Chapter 1**, we give a motivation for this study and formulate two main goals. We introduce the branch of philosophy to which this work is a contribution: formal and computational philosophy. We give a review of the foundations of probability and randomness, which includes mathematical and philosophical aspects. Special attention goes out to the case of infinite sample spaces in probability.

In **Chapter 2**, we discuss how the concept of a fair finite lottery can best be extended to denumerably infinite lotteries. Techniques and ideas from non-standard analysis are brought to bear on the problem.

Chapter 3 analyzes rational belief as 'almost certainty' and proposes a solution to the Lottery Paradox. A popular way to relate probabilistic information to binary rational beliefs is the Lockean Thesis, which is usually formalized in terms of thresholds. This approach seems far from satisfactory: the value of the thresholds is not well-specified and the Lottery Paradox shows that the model violates the Conjunction Principle. We argue that the Lottery Paradox is a symptom of a more fundamental and general problem, shared by all threshold-models that attempt to put an exact border on something that is intrinsically vague. We propose application of the language of relative analysis—a type of non-standard analysis—to formulate a new model for rational belief, called Stratified Belief. This contextualist model seems well-suited to deal with a concept of beliefs based on probabilities 'sufficiently close to unity' and satisfies a moderately weakened form of the Conjunction Principle. We also propose an adaptation of the model that is able to deal with beliefs that are less firm than 'almost certainty'. The adapted version is also of interest for the epistemicist account of vagueness.

Chapter 4 gives a summary of the findings of Chapters 2 and 3. It shows that by exploiting the parallels between large, yet finite lotteries on the one hand and countably infinite lotteries on the other, we gain insights in the foundations of probability theory as well as in epistemology. We solve the 'adding problems' that occur in these two contexts using a similar strategy, based on non-standard analysis. The new element in this chapter is the development of the epistemology of an infinite lottery.

In **Chapter 5**, we present a model for studying communities of epistemically interacting agents who update their belief states by averaging (in a specified way) the belief states of other agents in the community. Our main goal is to calculate the probability for an agent to end up in an inconsistent belief state due to updating (in the given way). To that end, an analytical expression is given and evaluated numerically, both exactly and using statistical sampling. It is shown that, under the assumptions of our model, an agent always has a probability of less than 2 % of ending up in an inconsistent belief state. Moreover, this probability can be made arbitrarily small by increasing the number of atomic sentences in the agents' language or by increasing the size of the community of agents.

In **Chapter 6**, we evaluate the goals set out in Chapter 1. We draw some conclusions from the philosophy of probability to the philosophy of science in general. In the outlook, we present a preliminary version of a paper on Non-Archimedean Probability (NAP) theory, which extends the solution of Chapter 2 to uncountable sample spaces.

Samenvatting

Een filosofie die 'in een notendop' verteld kan worden, hoort daar thuis.

Sydney J. Harris—'Leaving the Surface'

In **Hoofdstuk 1** geven we een motivatie voor dit onderzoek en formuleren we twee hoofddoelstellingen. We geven een inleiding over die takken van de filosofie waaraan dit werk een bijdrage wil leveren: formele en computationele filosofie. We geven een overzicht van de grondslagen van waarschijnlijkheid en willekeur, waarbij zowel wiskundige als filosofische aspecten aan bod komen. Speciale aandacht gaat uit naar situaties met oneindige kansruimten.

In **Hoofdstuk 2** bespreken we hoe het concept van een eerlijke, eindige loterij het beste uitgebreid kan worden naar aftelbaar oneindige loterijen, waarbij we gebruik maken van technieken en ideeën uit de niet-standaard analyse.

Hoofdstuk 3 analyseert rationeel geloof in termen van 'bijna zekerheid' en stelt een oplossing voor van de Lotterijparadox. Een populaire manier om probabilistische informatie te relateren aan binaire, rationele overtuigingen is de 'lockeaanse stelling', die gewoonlijk geformaliseerd wordt in termen van drempelwaarden. Deze aanpak lijkt verre van bevredigend: de waarde van de drempel wordt niet duidelijk gedefinieerd en de Lotterijparadox toont aan dat dit model niet aan het Conjunctieprincipe voldoet. Wij voeren aan dat de Lotterijparadox een symptoom is van een meer fundamenteel en algemeen probleem, dat gedeeld wordt door alle modellen met drempelwaarden, die trachten een exacte grens te plakken op iets dat inherent vaag is. We stellen voor om een nieuw model voor rationeel geloof te formuleren, in termen van relatieve analyse—een soort van niet-standaard analyse—en noemen dit model 'Gelaagd Geloof'. Dit contextualistisch model lijkt geknipt te zijn om overtuigingen die gebaseerd zijn op kansen 'dicht genoeg bij n' te beschrijven en voldoet aan het Conjunctieprincipe, zij het in een iets afgezwakte vorm. We stellen ook een aangepaste versie van het model voor, dat in staat is om geloof dat minder sterk onderbouwd is dan 'bijna zekerheid' te beschrijven Deze aangepaste versie is ook relevant voor de epistemische aanpak van vaagheid.

Hoofdstuk 4 vat de bevindingen uit Hoofdstuk 2 en 3 samen. Het toont aan dat we door parallellen te trekken tussen grote, maar eindige loterijen enerzijds en oneindige loterijen anderzijds, inzicht kunnen verwerven zowel in de grondslagen van de kansrekening als in de epistemologie. De 'optel-problemen' die in deze twee

contexten opduiken, kunnen via een gelijklopende strategie opgelost worden, namelijk met behulp van niet-standaard analyse. Een nieuw element in dit hoofdstuk is dat we nu ook de epistemologie van een oneindige loterij ontwikkelen.

In **Hoofdstuk 5** stellen we een model voor om een gemeenschap van epistemisch interagerende individuen te bestuderen. Deze individuen herzien hun geloofsovertuigingen aan de hand van een middeling (op een specifieke manier) over de geloofsovertuigingen van andere individuen in hun gemeenschap. Onze doelstelling is om de kans te berekenen dat een individu in een inconsistente geloofstoestand terechtkomt door deze manier van updaten. Hiertoe stellen we een analytische uitdrukking op en evalueren we deze numeriek, zowel op een exacte manier als via een statistische sampling-methode. We tonen aan dat, onder de aannames van ons model, de kans dat een individu in een inconsistente geloofstoestand belandt altijd lager is dan 2 %. Deze waarschijnlijkheid kan bovendien willekeurig klein gemaakt worden door het aantal atomaire zinnen in de taal van de individuen uit te breiden, of door het aantal individuen in de gemeenschap te verhogen.

In **Hoofdstuk 6** evalueren we de doelstellingen die we in Hoofdstuk 1 vooropgesteld hebben. We trekken een aantal besluiten uit de filosofie van de waarschijnlijkheid voor de wetenschapsfilosofie in het algemeen. In de outlook presenteren we een voorlopige versie van een artikel over niet-Archimedische kansrekening, die de oplossing uit Hoofdstuk 2 uitbreidt naar niet-aftelbare kansruimten.

Acknowledgements / Dankwoord

fellow assistants Myriam Delagrense and Anthony Parmentier for the nice time. I am also grateful to the mathematicians of Ghent University who helped to initiate me in the wondrous world of non-standard analysis: Michiel De Smet, Chris Impens, and Sam Sanders. I thank Karel Hrbacek for the engaging e-mail correspondence on relative analysis and vagueness.

Tot slot een woord van dank aan mijn ouders, Irène Das en Luc Wenmackers. Ik besef dat het voor jullie niet altijd even duidelijk is waar ik mee bezig ben: is het levenslang leren of toch echt werken? En waarom lijkt dat zo op het reizend volkstheater? Toch twijfelen jullie niet dat het allemaal goed zal komen. Bedankt voor het vertouwen, de steun en het feit dat ik altijd welkom ben. Mama, nog steeds hoop ik op een dag alles te kunnen wat jij kunt!

Men zegt dat samen op reis gaan een goede relatietest is. Samen doctoreren is dat natuurlijk ook. Dat het ons gelukt is en zelfs voor herhaling vatbaar bleek, schrijf ik grotendeels toe aan mijn vriend, Danny Vanpoucke. Danny, zonder jou was ik al lang verhongerd, verdwaald in een wereld van gedachten. Er waren elke dag wel goede redenen om je te bedanken: omdat je mijn dans- en levenspartner bent, mijn kok en co-auteur, mijn Luke Danes en Peter Bishop ... maar vandaag bedank ik je in het bijzonder voor je steun en liefde.

Sylvia

List of Publications

This is a list only of those articles that are related to a chapter of this dissertation.

[1] *Fair Infinite Lotteries*
S. Wenmackers and L. Horsten
Forthcoming in *Synthese* (2010); DOI: 10.1007/s11229-010-9836-x.
(**Chapter 2** is based on this paper.)

[2] *Ultralarge Lotteries and Stratified Belief*
S. Wenmackers
Submitted as two separate articles (2011).
(**Chapter 3** is based on these papers.)

[3] *Ultralarge and Infinite Lotteries*
S. Wenmackers
Forthcoming in B. Van Kerkhove, T. Libert, G. Vanpaemel, and P. Marage (eds.)
'Logic, Philosophy and History of Science in Belgium II', Proceedings of the Young
Researchers Days 2010, Koninklijke Vlaamse Academie van België voor Wetenschappen en Kunsten, Brussels (2011).
(**Chapter 4** is based on this paper.)

[4] *Determining the Probability of Inconsistencies in Theory Updating under Bounded Confidence*
S. Wenmackers, D. E. P. Vanpoucke, and I. Douven
Submitted to *Journal of Mathematical Sociology* (2010).
(**Chapter 5** is based on this paper.)

[5] *Axioms for Non-Archimedean Probability (NAP)*
V. Benci, L. Hosten, and S. Wenmackers
Forthcoming in proceedings of 'PhDs in Logic III' (2011).
(The outlook-section of **Chapter 6** is based on this paper.)

List of Conference Contributions

This is a list only of those oral and poster contributions at conferences that are directly related to the subject of this dissertation. I am grateful to the audiences at these meetings for valuable feedback. I also thank the London School of Economics for funding my stay during the Graduate Conference *POPIII*.

In 2011:

Models in material science: two cases without error bars (with D. E. P. Vanpoucke), Oral contribution at: All models are wrong. . .—Model uncertainty & selection in complex models,
March 14–16, 2011, University of Groningen, Groningen, The Netherlands.

Axioms for non-Archimedean probability (NAP) (with V. Benci and L. Horsten), Oral contribution at: PhDs in Logic III,
February 17–18, 2011, Palais des Académies, Brussels, Belgium.

Sociophysics and opinion dynamics (with D. E. P. Vanpoucke and I. Douven), Oral contribution at: Vastestofwisselingen—Research seminar,
February 11, 2011, Ghent University: Solid State Sciences, Ghent, Belgium.

In 2010:

Indefinitely large and infinite lotteries (with L. Horsten),
Oral contribution at: Second Young Researchers Days, *YRDII*—Workshop on the Relations between Logic, Philosophy and History of Science,
September 6–7, 2010, Palais des Académies, Brussels, Belgium.

Finite and infinite, denumerable lotteries (with L. Horsten),
Oral contribution at: Logic Colloquium 2010, *Logic2010*—Annual European conference on logic, organized under the auspices of the Association for Symbolic Logic,
July 25–31, 2010, Paris Diderot University, Paris, France.

Probability and Epistemology of Denumerable Lotteries Using Non-Standard Analysis,
Oral contribution at: Philosophy of Probability III, *POPIII*—Graduate Conference,
June 25–26, 2010, London School of Economics, London, UK.

Probability and epistemology of denumerable lotteries,
Poster contribution at: Fourth Workshop in Decisions, Games & Logic '10, *DGL10*,
June 9–11, 2010, ENS Paris, France.

Ultralarge lotteries,
Oral contribution at: Philosophy of Science in a Forest, *PSF2010* - Vijfde Vlaams-
Nederlands Congres voor Algemene en Speciale Wetenschapsfilosofie,
May 14–15, 2010, Leusden, The Netherlands.

Curriculum Vitae

Name: Sylvia Wenmackers
Date of birth: May 12, 1980
Place of birth: Maasmechelen, Belgium

Work experience

Dec. 2010–now PhD-student on the Formal Epistemology Project
Faculty of Philosophy, University of Groningen, The Netherlands

Jan. 2010–Nov. 2010 Practical assistant
Department of Solid State Science, Ghent University, Belgium

Dec. 2009–Nov. 2010 Postdoctoral researcher on the Formal Epistemology Project
Faculty of Philosophy, Catholic University of Leuven, Belgium

Oct. 2008–Oct. 2009 Postdoctoral researcher of the Flemish Research Fund (FWO)
Institute for Materials Research, Hasselt University, Belgium

Sept. 2002–Sept. 2008 PhD-student and teaching assistant
Institute for Materials Research, Hasselt University, Belgium

Degrees

May 26, 2008 PhD in Science: Physics—Hasselt University, Belgium
Dissertation: "Morphology, functionality and molecular conformation study of CVD diamond surfaces functionalised with organic linkers and DNA"
Promoter: Prof. dr. P. Wagner
Copromoter: Prof. dr. M. Nesládek

July 3, 2002 Master in Science: Theoretical Physics—Ghent University, Belgium
Thesis: "Interpretaties van de kwantummechanica: overzicht en synthese"
Promoter: Prof. dr. W. De Baere

July 10, 2002 Academic teaching degree: Physics—Ghent University, Belgium

July, 2000 Bachelor in Science: Physics—Limburgs Universitair Centrum, Belgium