THE OPEN HANDBOOK OF FORMAL EPISTEMOLOGY

RICHARD PETTIGREW & JONATHAN WEISBERG, EDS.

LIST OF CONTRIBUTORS

R. A. Briggs
Stanford University

Michael Caie
University of Pittsburgh

Kenny Easwaran
Texas A&M University

Franz Huber
University of Toronto

Jason Konek
University of Bristol

Hanti Lin
University of California, Davis

Anna Mahtani
London School of Economics

Konstantin Genin
University of Toronto

Johanna Thoma
London School of Economics

Michael G. Titelbaum
University of Wisconsin, Madison

Sylvia Wenmackers
Katholieke Universiteit Leuven

An epigraph: something pithy, and surprisingly apt to the context if you just stop and think about it a moment. — *Someone Famous*

For whosits

ACKNOWLEDGMENTS

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

CONTENTS

INFINITESIMAL PROBABILITIES

Sylvia Wenmackers

> *Suppose that a dart is thrown, using the unit interval as a target;*
> *then what is the probability of hitting a point?*
> *Clearly this probability cannot be a positive real number,*
> *yet to say that it is zero violates the intuitive feeling that,*
> *after all, there is some chance of hitting the point.*
>
> —Bernstein and Wattenberg (1969, p. 171)

> *It has been said that to assume that $0 + 0 + 0 + ... + 0 + ... = 1$ is absurd,*
> *whereas, if at all, this would be true if*
> *'actual infinitesimal' were substituted in place of zero.*
>
> —de Finetti (1974, p. 347)

Infinitesimals played an important role in the seventeenth century development of the calculus by Leibniz and—to a lesser extent—by Newton. In the twentieth century, calculus was applied to probability theory. By this time, however, Leibnizian infinitesimals had lost their prominence in mainstream calculus, such that "infinitesimal probability" did not become a central concept in mainstream probability theory either. Meanwhile, non-standard analysis (NSA) has been developed by Abraham Robinson, an alternative approach to the calculus, in which infinitesimals (in the sense of Equation 1 below) are given mathematically consistent foundations. This provides us with an interesting framework to investigate the notion of infinitesimal probabilities, as we will do in this chapter.

Even taken separately, both infinitesimals and probabilities constitute major topics in philosophy and related fields. Infinitesimals are numbers that are infinitely small or extremely minute. The history of non-zero infinitesimals is a troubled one: despite their crucial role in the development of the calculus, they were long believed to be based on an inconsistent concept. For probabilities, the interplay between objective and subjective aspects of the concept has led to many puzzles and paradoxes. Viewed in this way, considering infinitesimal probabilities combines two possible sources of complications.

This chapter aims to elucidate the concept of infinitesimal probabilities, covering philosophical discussions and mathematical developments (in as far as they are relevant for the former). The introduction first specifies what it means for a number to be infinitesimal or infinitely small and it addresses some key notions in the foundations of probability theory. The remainder of the chapter is devoted to interactions between these two notions. It is divided into three parts, dealing with the history, the

mathematical framework, and the philosophical discussion on this topic, followed by a brief epilogue on methodological pluralism. The appendix reviews the literature of 1870–1989 in more detail.

Infinitesimals

In an informal context, infinitesimal means extremely small. The word 'infinitesimal' is formed in analogy with 'decimal': decimal means one tenth part; likewise, infinitesimal means one infinith part. As such, the word 'infinitesimal' suggests that infinitesimal quantities are reciprocal to infinite ones, and that infinitely many of them constitute a unit. In Wenmackers ([2018]), I have introduced the term 'harmonious' as a property of number systems such that "each infinite number is the multiplicatory inverse of a particular infinitesimal number, and vice versa". In other words, an harmonious number system does justice to the etymology of 'infinitesimal'. Moreover, in such a number system, "neither the infinite nor the infinitesimal numbers are conceptually prior to or privileged over the other in any way."

These suggestions can be formalized in non-standard analysis (NSA), which allows us to work with so-called hyperreal numbers. The set of hyperreal numbers, $^*\mathbb{R}$, contains positive (and negative) infinite numbers, larger than any (standard) number, as well as their multiplicative inverses, which are strictly positive (or strictly negative, respectively) infinitesimal numbers, smaller than any positive real number yet greater than zero.[1] The hyperreals are harmonious in the sense just defined.

Let us now state the formal definition for infinitesimals that we consider in this chapter. A number x is infinitesimal if

$$\forall n \in \mathbb{N} : \ |x| < \frac{1}{n}.$$

According to this definition, *zero is an infinitesimal* and it is the only real-valued infinitesimal.[2] Number systems that do not contain strictly positive or strictly negative infinitesimals, such as $\mathbb{R}$, are called *Archimedean*; number systems that do contain non-zero infinitesimals, such as $^*\mathbb{R}$, are called *non-Archimedean*. NSA is certainly not the only framework for dealing with infinitesimals,[3] but currently it is the most common one for representing infinitesimal probabilities, so that is what this chapter focuses on.

---

1 Actually, it is more accurate to write '*a* set of hyperreal numbers', rather than 'the set', since the definition is not categoric (unlike that of $\mathbb{R}$) and there is no canonical choice among the $^*\mathbb{R}$'s. See section 1.2.1 of the appendix for details.

2 Some authors exclude zero in their definition of infinitesimals, but for the exposition in this chapter it will turn out to be beneficial to include it.

3 Section 11 mentions two alternative frameworks that deal with infinitesimal numbers.

What is an infinitesimal probability value? The answer depends on which number system you are using: we already observed that zero is *the* infinitesimal number within the real numbers, whereas the hyperreal numbers contain (infinitely many) strictly positive infinitesimals, which could serve as strictly positive infinitesimal probability values.

One way to obtain a new number system is by considering a suitable quotient space. In general, the definition of a quotient space relies on the definition of some equivalence relation on a collection of objects, which can be (generalized) sequences.[4] Informally, the equivalence relation expresses a condition for two objects to be "indistinguishable" from each other or for their difference to be "infinitesimal' or "negligible". In case of (generalized) sequences, this condition has to specify (i) *a criterion* to compare corresponding positions by and (ii) *a selection rule* that specifies at which collections of indices said criterion has to hold. Both the construction of the real numbers and that of the hyperreal numbers fits this general description, but the relevant equivalence relations impose different conditions for sequences to be indistinguishable from each other:

(1) The negligibility of a sequence can be formalized as "converging to zero": the sequence gets (i) *arbitrarily close to* the (rational) number zero (ii) *eventually*.

(2) Another way to define negligibility of a sequence is as being (i) *exactly equal to* the (real) number zero (ii) *except for a small index set*.

We will define the criteria and selection rules in italics later in this chapter (see section 8.5). For now, it suffices to know that two sequences can be defined to be equivalent if they differ only by a negligible sequence (in a well-defined sense). Using this equivalence relation, we can define equivalence classes of sequences; the structure of the collection of these equivalence classes is a quotient set. For some choices, this set may be isomorphic to that of the set of real or hyperreal numbers. In particular, the equivalence class of rational-valued Cauchy sequences that are negligible in the sense of (1) is the real number zero ($0_\mathbb{R}$) and the equivalence class of real-valued sequences that are negligible in the sense of (2) is the hyperreal number zero ($0_{*\mathbb{R}}$).

Since being exactly equal to zero implies being infinitely close to zero, but not vice versa, we may think of $0_\mathbb{R}$ as *the* infinitesimal in the set of the real numbers, which corresponds with an infinite equivalence class of sequences, many of which belong to that of non-zero infinitesimals in the hyperreal context. In this sense, the hyperreal numbers are capable of representing finer distinctions (among sequences) than the real numbers are.

---

4 For generalized sequences, see section 9.2.

After this brief introduction to infinitesimals, let us now give an even briefer intro to probabilities.

Probabilities

In an informal context, probable means plausible or likely to be true. Similar words were available in medieval Latin ('*probabilis*' for probable and '*verisimilis*' for likely). As such, probability can be seen as a shorthand for 'probability of truth' and likelihood is a measure of appearing to be true. This suggests that probability is a hybrid concept that combines *objective* chances and *subjective* degrees of belief (or credences). We may picture it as a two-layered concept with an objective ground layer, which represents the objective state of affairs (truth), and an epistemic cover layer, that deals with evidence presented to an agent and quantifying the possibility of it being misleading concerning what is underneath it (appearance).

Many authors have tried to capture this duality that is inherent in the probability concept. Hacking (1975) describes it very aptly as the Janus-faced nature of probability and Gaifman (1986) paints a colourful picture of probability as living on a spectrum from purely objective to purely epistemic forms. It may be helpful to imagine both layers as allowing for different degrees of opacity. For an agent with limited epistemic (cognitive and empirical) resources, the outer layer acts as a veil. First assume that the underlying system is purely deterministic, such that there are no probabilities "out there", or, put differently, they are all zero or one. However, the agent does not see things exactly as they are—only approximately so. Hence, the probabilities that are relevant to such an agent may be other than just zeros and ones.[5] If the underlying system is indeterministic, on the other hand, even an agent with unlimited epistemic resources (such as Laplace's demon), who could see right through the outer layer, would still need probabilities to describe the system.

Apart from its interpretation, the topic of this chapter also requires us to pay attention to the mathematical representation of probabilities. Probability is usually formalized as a function from the event space—a collection of subsets (often a sigma-algebra) of a given set, the sample space—to the unit interval of the real numbers or a non-standard extension thereof. A probability distribution is called *fair* or *uniform* if the same probability is assigned to any singleton from the domain. Depending on other background assumptions, this may imply slightly stronger properties, such as translation invariance.

---

5 This viewpoint helps us to understand that Laplace (1814) was strongly involved in the development and popularization of probability theory, while also popularizing the idea of a deterministic universe.

In this chapter, we will encounter infinitesimals both in the context of subjective probability (infinitesimal credences or degrees of belief) and in the context of objective probability (infinitesimal chances), as well as in contexts that are intermediate on this continuum.

PART I

HISTORICAL OVERVIEW

In this part, we review some essential mathematical developments that allow us to represent infinitely small probabilities as positive infinitesimals in a hyperreal field. We also review philosophical discussions of the topic. A much more detailed list of contributions from the period 1870–1989 can be found in the appendix. More recent contributions are discussed in Part IV.

The concept of infinitesimals was thought to be intrinsically problematic and inconsistent for most of European history. An important exception is the work of Archimedes, who allowed infinitesimals as a method to find new results, though he did not regard them sufficient for establishing rigorous proofs of those results. In the sixteenth century, a Latin translation of many of the works of Archimedes was published in Europe, which led to a revival of scholarly interest in infinitesimals, especially in Italy. (See Alexander 2014 for an overview of the seventeenth century response to infinitesimals in Europe.)

In the second half of the seventeenth century, infinitesimals played a crucial role in the development of the calculus, especially in the work of Gottfried Wilhelm Leibniz (see, *e.g.*, Katz & Sherry, 2012; Katz & Sherry, 2013). Whereas the guiding notion in Newton's calculus was the "fluxion" (the derivative of a continuous quantity), Leibniz developed his version of the calculus starting from infinite sums (integrals). Newton's and Leibniz's usage of infinitesimals was criticized early on, famously by Berkeley (1734), who called them "ghosts of departed quantities". Around the 1870s, the calculus received its formalization in terms of real numbers and standard limits, which do not allow non-zero infinitesimals. This further consolidated the general belief that infinitesimals do not live up to the rigour of modern mathematics, but we will see that a formalization of this concept was discovered later on, in the 1960's.

The current standard approach to calculus, which is used for instance in college physics, is based on the nineteenth century formalization, in which the epsilon-delta definition of the limit operation takes a central place (see appendix 1.1.1). As a result, our standard calculus differs from both the Newtonian and the Leibnizian version of it. The core idea of a limit operation is closer in spirit to the Newtonian version, while Leibnizian notation proved to be more enduring, with, for instance, $\mathrm{d}x / \mathrm{d}t$

for the derivative of x to t. (For Leibniz, this signified an actual ratio of infinitesimals, whereas our standard calculus defines it as the limit of a ratio of real numbers.)

As we will see below, measure and probability theory was developed based on the standard calculus. The non-standard approach, based on the alternative formalization of the calculus from the 1960's, is more recent. (Hence the unfortunate name 'non-standard'.) But, like infinitesimals in general, also the more specific notion of infinitesimal probability was in use long before its formal definition. For instance, in his famous wager argument (*Pensées* L418/S680), Pascal specifically excluded them from his argument.[6]

1   THE PRE-ROBINSONIAN ERA: 1880–1959

Around 1880, the current foundations of the real numbers and the standard calculus, with the epsilon-delta definition of the limit, were well in place. Non-standard analysis was not developed yet.

Standard measure theory was being developed by mathematicians such as Émile Borel, Henri Lebesgue, Johann Radon, Maurice Fréchet, Giuseppe Vitali, and many others. In response to the sixth problem of David Hilbert (1900), also the first axiomatization of probability theory was developed: Kolmogorov (1933) presented an approach that embedded probability theory into standard measure theory. (His axioms are included in section 7.)

After the foundational work by Kolmogorov, the measure-theoretic approach to probability became the standard formalism, which represents probabilities as real numbers. Strictly speaking, non-zero infinitesimal probabilities (defined as non-Archimedean quantities) are incompatible with this formalism. Nevertheless, informal usage of the term has remained in fashion in at least two ways. First, in some contexts it is used to discuss events that have zero probability but that are logically possible to occur. Second, the phrase 'infinitesimal probability' is also used in the context of continuous probability distributions, to refer to dp.[7]

At about the same time, Bruno de Finetti (1931) was developing a qualitative theory for ranking events in terms of their probability. He discovered that, in general, these rankings are non-Archimedean. His rankings can be said to be more fine-grained than what is expressible

---

6 In Krailsheimer's translation, the relevant sentence reads as follows (Pascal, 1670 / 1995, p. 151, my emphasis): "[W]herever there is infinity, and where there are not *infinite chances of losing against that of winning*, there is no room for hesitation, you must give everything."

7 The notation stems from Leibniz, for whom dp indicated an infinitesimal increment of a quantity p. In contemporary standard analysis, however, there are no non-zero infinitesimals and dp merely indicates that the variable of differentiation or integration is p.

by the real-valued probability functions in Kolmogorov's theory. Five years later, de Finetti (1936) specifically addressed logically possible events that receive probability zero in Kolmogorov's theory. Here, we see that de Finetti explicitly entertained the notion of infinitesimal probabilities, but he ultimately chose to stick to real-valued probabilities and to reject countable additivity.

Working on the subjective interpretation of probability, Frank P. Ramsey and Bruno de Finetti developed the notion of coherence: in order for an agent's degrees of belief to be rational (at a given point in time), they have to conform to Kolmogorov's axioms for probability. Abner Shimony (1955) aimed to strengthen this notion to strict coherence (now often called regularity): it requires that the degree of confirmation of an hypothesis h given a piece of evidence e is 1 if and only if h logically entails e. Shimony was aware that strict coherence required infinitesimal betting quotients—and thus was incompatible with Archimedean values—if the sample space was infinite. Inspired by this proposal, Rudolf Carnap (1980) set out to develop a theory for non-Archimedean credences. Although this interesting approach was written before Robinson's work, it was only published afterwards. As a result, it has not been very influential.

Meanwhile, Thoralf Skolem (1934) had discovered non-standard models of the natural numbers (Peano arithmetic), which we now call hypernatural numbers. By applying similar model-theoretic techniques to the real numbers, Robinson would be able to develop non-standard analysis. This brings us to the next period.

2 ROBINSON'S NON-STANDARD ANALYSIS: 1960'S

Abraham Robinson (1961, 1966) founded the field of NSA: he aplied earlier results from mathematical logic (such as that of Skolem) to real closed fields in order to develop an alternative framework for differential and integral calculus based on infinitesimals and infinitely large numbers. This allowed for a formal and consistent treatment of infinitesimal numbers and provided a harmonious number system (as defined in the introduction). Soon enough, NSA was applied to measure theory in general and to probability theory in particular.

For our current purposes, it is good to be aware of two modes of operation of NSA: in one, the hyperreal numbers merely serve as a means to prove results about the real numbers, but in the other, obtaining a hyperreal-valued function or some other non-standard object is the final goal.[8] The first mode of operation represents the oldest and still the *most*

---

8 This situation is similar to that of the complex numbers. On the one hand, as Paul Painlevé (1967, pp. 1–2) writes: "*entre deux vérités du domaine réel, le chemin le plus facile et le plus court passe bien souvent par le domaine complexe*" ("between two truths of the real domain,

common application of NSA, which is to make proofs about standard analysis shorter, easier, or both—mainly by alleviating epsilon-delta management (Tao, 2007).[9] Although the most common one, this is *not the only* application of NSA. The second mode of operation allows us to investigate non-standard objects in their own right, including those that (roughly speaking) do not have standard counterparts.[10] In particular, if we are interested in developing a probability theory that allows us to assign non-zero infinitesimal probabilities to some events, we cannot achieve this if we move back to the real domain in the final step.

An early example of a non-standard measure was provided by Bernstein and Wattenberg (1969), who attempted to measure the infinitesimal probability of hitting a particular point when playing (infinitely precise) darts on the unit interval of the real numbers. This result was a very important first step in the development of probability theories in which the numerical values respect the non-Archimedean ordering of the events (as studied by de Finetti, 1936). Hence, Bernstein and Wattenberg (1969) have often been cited by philosophers who work on the foundations of probability theory. However, since they focused on a particular case, their result is not fully general: they did not present a non-standard probability theory, although their approach can be generalized and does in fact contain many of the essential ingredients present in later developments.

3 POST-ROBINSONIAN DEVELOPMENTS: 1970–1989

Seminal contributions to non-standard measure theory were obtained by Peter A. Loeb (1975). The dominant line of research in non-standard measure and integration theory is based on real-valued functions that have a non-standard domain and the main application (like for all of NSA) is finding new results in standard measure and integration theory. Although the well-developed theory of Loeb measures has proven fruitful in many applications, and therefore should not go unmentioned, it is not of immediate interest to the topic of this chapter (but see Herzberg, 2007, 2010). For, although infinitesimal probabilities do occur in the construction

---

the easiest and shortest route quite often passes through the complex domain"). On the other hand, complex numbers are also useful by themselves (for instance, to represent phasors in physics). This analogy is also employed by Bartha and Hitchcock (1999, p. 416), who write: "Just as imaginary numbers can be used to facilitate the proving of theorems that exclusively concern real numbers, our use of nonstandard analysis will be used to facilitate and motivate the construction of purely real-valued measures."

9 An early expression of this (prior to the development of NSA) can be found with Joseph-Louis Lagrange, as cited in Błaszczyk, Katz, and Sherry (2013, p. 63). Recent examples are given by Terence Tao in his blog posts (see, *e.g.*, Tao, 2007–2012).

10 These are "external" objects, as will be defined in section 4.

of Loeb measures, the end goal is to obtain real-valued measures, thereby eliminating all non-zero infinitesimal probabilities.

Although de Finetti lived long enough to see the advent of NSA and was aware of its existence, he never used it to continue his 1936 observations regarding infinitesimal probabilities and he did not show much interest in applying it in his own work on probability.[11]

To make the earlier, often technical, work accessible to a larger audience, including philosophers, it was important to summarize and interpret it. Brian Skyrms played an important role in this regard. For instance, in Skyrms (1980, appendix 4), he discussed the trade-off between four demands—additivity, translation invariance, everywhere-definedness, and regularity—for standard and non-standard measures. In the same year, David Lewis (1980) discussed infinitesimal credences, in the same spirit as Shimony and Carnap had done prior to Robinson's work. Later on, Lewis (1986a) also mentioned infinitesimal chances, in wordings very reminiscent of Bernstein and Wattenberg (1969).

Observe that at this point, there still was no non-Archimedean alternative to parallel Kolmogorov's Archimedean probability theory. It was Edward Nelson (1987), who provided the first axiomatic approach for a probability theory with infinitesimal values. His "radically elementary probability theory" is indeed very simple, but it requires an entirely different mind set than, for instance, Loeb's approach. In particular, Nelson's theory cannot be used to assign probability measures to any standard infinite set. Instead, one has to go one step back in the modelling process and represent the set of possibilities by an infinite hyperfinite set rather than a standard infinite set. We will introduce the notion of hyperfinite sets in section 4.3. Since hyperfinite sets are very similar to discrete finite ones, after that choice, everything resembles Kolmogorov's theory for finite sample spaces.

At this point, we end our historical overview. Some of the more recent approaches and debates will be discussed in sections 8, 9, and 14.

---

11 See section 1.3.1 of the appendix for details.

PART II

MATHEMATICAL PRELIMINARIES

In this part, we will briefly review some common non-standard tools and the dual notions of filters and ideals. We will apply these notions in the ultrafilter construction of the hyperreals. We also present the axioms of standard probability theory. After that, we will be properly equipped to address infinitesimal probabilities in the context of countable lotteries as well as other cases.

4   COMMON NON-STANDARD TOOLS

In this section, we review some common tools that appear in (nearly) all approaches to non-standard analysis.[12]

4.1   *Universe*

By a *universe*, we mean a non-empty collection of mathematical objects, such as numbers, sets, functions, relations, *etc.*—all of which can be defined as sets by working in Zermelo–Fraenkel set theory with the Axiom of Choice (ZFC). This collection is assumed to be closed under the following relations and operations on sets: $\subseteq$, $\cup$, $\cap$, $\setminus$, $(\cdot, \cdot)$, $\times$, $\mathcal{P}(\cdot)$, $\cdot$. Furthermore, we assume that the universe contains $\mathbb{R}$ and that it obeys transitivity (*i.e.,* elements of an element of the universe are themselves elements of the universe).

In particular, we are interested in the standard universe, which is the superstructure $V(\mathbb{R})$, and a non-standard universe, $^*V(\mathbb{R})$.

4.2   *Star-map*

The star-map (or hyperextension) is a function from the standard universe to the non-standard universe.

$$* : V(\mathbb{R}) \to {}^*V(\mathbb{R})$$
$$A \mapsto {}^*A$$

We assume that $\forall n \in \mathbb{N}$, $^*n = n$ and that $\mathbb{N} \neq {}^*\mathbb{N}$.

In the literature, two notations occur for the star map: before or after the standard object. In this chapter, I have opted for the former notation, because it allows us to read the *-symbol as the prefix 'hyper-'. For instance, $^*\mathbb{R}$ are the hyperreals.

---

12 For further information, see also Benci, Di Nasso, and Forti (2006, section 1) and Cutland (1983, section 1.2).

4.3 *Internal and external objects*

It is important to realize that the star-map does *not* produce all the objects in the superstructure of $^*\mathbb{R}$; it only maps to the internal objects, which live in $^*V(\mathbb{R}) \subsetneq V(^*\mathbb{R})$.

Some examples of internal objects ($\in {}^*V(\mathbb{R})$):

- any element of $^*\mathbb{R}$, so in particular any element of $\mathbb{N}$ or $\mathbb{R}$;

- any hyperfinite set, such as $\{1, \ldots, N\}$ with $N \in {}^*\mathbb{N}$ (which can be obtained via the hyperextension of a family of finite sets);

- the hyperextensions of standard sets, such as $^*\mathbb{N}$ and $^*\mathbb{R}$;

- the hyperpowerset of a standard set, A: $^*\mathcal{P}(A)$, which is the collection of all *internal* subsets of $^*A$.

Some examples of external objects ($\in V(^*\mathbb{R}) \setminus {}^*V(\mathbb{R})$):

- elementwise copies of standard, infinite sets (notation for the elementwise copy of A in the non-standard universe: $^\sigma A$), such as $^\sigma \mathbb{N}$ or $^\sigma \mathbb{R}$ (due to the embedding of $\mathbb{N}$ and $\mathbb{R}$ in $^*\mathbb{R}$, the $^\sigma$-prefix is often dropped);

- the complements of previous sets, such as $^*\mathbb{N} \setminus {}^\sigma \mathbb{N}$ and $^*\mathbb{R} \setminus {}^\sigma \mathbb{R}$;

- the *halo* or *monad* of any real number, r: $hal(r) = \{R \in {}^*\mathbb{R} \mid |r - R| \text{ is infinitesimal}\}$—in particular $hal(0)$, which is the set of all infinitesimals;

- the standard part function st (also known as the shadow), which maps a (bounded) hyperreal number to the unique real number that is infinitesimally close to it (Goldblatt, 1998, section 5.6);

- the full powerset of the hyperextension of a standard, infinite set, A: $\mathcal{P}(^*A)$, which is the collection of *all* subsets of $^*A$, both internal and external.

4.4 *Transfer principle*

Consider some standard objects $A_1, \ldots, A_n$ and consider a property of these objects expressed as an *elementary sentence* (a bounded quantifier formula in first-order logic): $P(A_1, \ldots, A_n)$. Then, the *Transfer* principle says:

$$P(A_1, \ldots, A_n) \text{ is true } \Leftrightarrow P(^*A_1, \ldots, {}^*A_n) \text{ is true.}$$

Observe: this is an implementation of Leibniz's "law of continuity" (or "*souverain principe*") in NSA (see Katz & Sherry, 2012, section 4.3). It may be helpful to consider two examples.

EXAMPLE 1: WELL-ORDERING OF $\mathbb{N}$ Consider the following sentence: "Every non-empty subset of $\mathbb{N}$ has a least element." Transfer does *not* apply to this, because the sentence is not elementary. Indeed, we can find a counterexample in $^*\mathbb{N}$: the set of infinite hypernatural numbers, $^*\mathbb{N} \setminus \mathbb{N}$, does not have a least element. (Of course, this is an external object.)

If we rephrase the well-ordering of $\mathbb{N}$ as follows: "Every non-empty element of $\mathcal{P}(\mathbb{N})$ has a least element", then we *can* apply Transfer to this. The crucial observation to make here is that $^*\mathcal{P}(\mathbb{N}) \subsetneq \mathcal{P}(^*\mathbb{N})$.

EXAMPLE 2: COMPLETENESS OF $\mathbb{R}$ Consider the following sentence: "Every non-empty subset of $\mathbb{R}$ which is bounded above has a least upper bound." Again, Transfer does not apply to this, for the same reason as in Example 1. A counterexample in $^*\mathbb{R}$ is $hal(0)$, the set of infinitesimals. (Again, an external object.)

If we rephrase the completeness property of $\mathbb{R}$ as follows: "Every non-empty element of $\mathcal{P}(\mathbb{R})$ which is bounded above has a least upper bound", then we can apply Transfer to it. Similarly as before, the crucial remark is that $^*\mathcal{P}(\mathbb{R}) \subsetneq \mathcal{P}(^*\mathbb{R})$.

5   FILTERS AND IDEALS

The introduction mentioned two ingredients for a new number system: the second one is a selection rule. This idea can be formalized using either filters or ideals. These are dual notions, and both are collections of subsets from an index set that fulfil additional criteria.

Intuitively, a filter on a set is a collection of its subsets that are "large enough", whereas an ideal is a collection of its subsets that are "small enough" or "negligible". The meanings of 'large enough' and 'small enough' are given by the formal definitions. The ultrapower construction of the hyperreal numbers crucially relies on the application of a particular kind of filter: a free ultrafilter. We review the relevant definitions here.[13]

$\mathcal{F}$ is a *proper, non-empty filter on X* if

$$\mathcal{F} \subseteq \mathcal{P}(X), \qquad \text{(collection of subsets)}$$

$$\varnothing \notin \mathcal{F}, \qquad \text{(proper)}$$

$$X \in \mathcal{F}, \qquad \text{(non-empty)}$$

$$A, B \in \mathcal{F} \Rightarrow A \cap B \in \mathcal{F}, \qquad \text{(closure under finite meets)}$$

---

13 Definitions are given, *e.g.*, in Schechter (1997, Ch. 5). For a further discussion of filters, including free ultrafilters, see, *e.g.*, Goldblatt (1998, p. 18–21) and Cutland (1983, section 1.1). For an introduction to the meaning and application of ultrafilters, see Komjáth and Totik (2008).

$$(A \in \mathcal{F} \wedge B \supseteq A) \Rightarrow B \in \mathcal{F}. \qquad\qquad \text{(upper set property)}$$

The smallest non-empty proper filter is simply $\{X\}$. A filter $\mathcal{F}$ is *principal* (or *fixed*) if $\exists x_0 \in X : \forall A \in \mathcal{F}, \ x_0 \in A$.

A filter $\mathcal{F}$ is *free* if it is *not* principal, or equivalently: if the intersection of all the sets in $\mathcal{F}$ is empty.

For an infinite set X, its *Fréchet filter* is the filter that consists of all the cofinite subsets of X. Such a filter is free, but it is not an ultrafilter. (For a finite set X, the Fréchet filter is not proper.)

$\mathcal{F}$ is an *ultrafilter* on X if $\mathcal{F}$ is a filter on X and

$$\forall A \subseteq X (A \notin \mathcal{F} \Rightarrow X \setminus A \in \mathcal{F}).$$

$\mathcal{F}$ is a *free ultrafilter* on X if $\mathcal{F}$ is an ultrafilter on X and $\mathcal{F}$ is free. This definition implies that a free ultrafilter contains no finite sets. Given the ultrafilter condition, it is equivalent to say that it does contain all cofinite sets. In other words: an ultrafilter is free if and only if it contains the Fréchet filter. Hence, free ultrafilters do not exist for finite X.

Given a (proper) filter on X, $\mathcal{F}$, the corresponding (proper) *ideal* in the Boolean algebra $\mathcal{P}(X)$, $\mathcal{I}$, is obtained as follows:

$$\mathcal{I} = \{X \setminus F \mid F \in \mathcal{F}\}.$$

The smallest proper ideal is simply $\{\varnothing\}$. The ideal corresponding to a free ultrafilter is called a Boolean prime ideal.

6 APPLICATION OF FREE ULTRAFILTERS: HYPERREAL NUMBERS

6.1 *Constructing the real and hyperreal numbers*

In the introduction, we indicated that both the standard real numbers and the hyperreal numbers can be defined as equivalence classes of sequences.[14] They differ in the collection of sequences on which they operate and in the equivalence relation that they impose.

The real numbers can be constructed based on rational-valued Cauchy sequences. The set of such functions is defined as follows:

$$\mathcal{C} = \{(q_n) \in \mathbb{Q}^{\mathbb{N}} \mid \forall \epsilon \in \mathbb{Q}_{>0}, \ \exists N \in \mathbb{N} : \ \forall n, m > N \left(|q_n - q_m| < \epsilon\right)\}.$$

Two sequences in this space are considered to be equivalent to each other if their difference (which is defined member-wise) is a sequence that gets *arbitrarily close to* (the rational number) zero, *eventually*. This means that for each target, from some position in the sequences onwards (*i.e.*, eventually

---

14 We will not consider Dedekind cuts or other constructions.

or cofinally), their member-wise difference is strictly smaller than the target. Symbolically, where $(q_n), (s_n) \in \mathcal{C}$:

$$(q_n) \sim (s_n) \Leftrightarrow \forall \epsilon \in \mathbb{Q}_{>0}, \ \exists N \in \mathbb{N} : \ \forall n > N \left(|q_n - s_n| < \epsilon \right).$$

The hyperreal numbers can be constructed based on real-valued sequences (all of $\mathbb{R}^{\mathbb{N}}$)—this is called the ultrapower construction of $^*\mathbb{R}$.[15] Two sequences in $\mathbb{R}^{\mathbb{N}}$ are considered to be equivalent to each other if their member-wise difference is *exactly equal to* (the real number) zero, *except for a small set of indices*. In this case, the first part of the condition is clear and all we are left to specify is what counts as a "small" set. If we choose to define small sets as finite sets, and thus large sets as cofinite ones, this coincides with the "eventuality" condition used in the construction of the real numbers. This is equivalent to imposing the Fréchet filter, consisting of the cofinite subsets of $\mathbb{N}$ (the complements of "small" sets, these are "large" sets), to the indices of the sequences. This setup does allow us to construct a non-standard model of the real numbers; in fact, it was the first one that was ever constructed and it is still of interest because it yields a constructive non-standard model.[16] However, such a system is rather weak (too weak for some of the questions we are interested in). According to the Fréchet filter, many sets (such as arithmetic progressions) are neither small nor large. Usually, small and large sets are defined by fixing a free ultrafilter on $\mathbb{N}$: a set is large if it is in the ultrafilter and small if it is not, and the ultra-condition guarantees that for each set either it is in the ultrafilter, or its complement is.

Informally, the sequence-based construction of the hyperreals can be thought of as follows. Consider the old equivalence class of the sequences that we have come to regard as the real number zero and define new equivalence classes on it, making distinctions among the infinitesimal sequences depending on their rate of convergence. As such, we dissect the single infinitesimal real number into infinitely many infinitesimal hyperreal numbers. In fact, we perform a similar dissection for each of the real numbers simultaneously. Does this give us old wine in new packages? Not quite: it is more like breaking the chemical bonds in the molecules

---

15 The ultraproduct construction is a general method in model theory: see Keisler (2010) (including the references in the introduction) for more information. To see how the ultrapower construction is related to the existence proof of non-standard models using the Compactness theorem (see appendix section 1.2.1), observe that one way to prove the Compactness theorem is based on the notion of an ultraproduct (*cf.* Goldblatt, 1998, p. 11).

16 Schmieden and Laugwitz (1958) were the first to give a construction in this style and they used a Fréchet filter on $\mathbb{N}$ rather than a free ultrafilter. Unlike a free ultrafilter, the existence of a Fréchet filter does not require any choice axiom. However, in strictly constructivist approaches, the framework of classical logic as used by Schmieden and Laugwitz (1958) also has to be replaced by intuitionist logic (Martin-Löf, 1990). More recently, Palmgren (1998) has investigated constructive approaches to NSA. For an accessible introduction to a weak system of NSA based on Fréchet filters, see also Tao (2012).

of the wine, and maybe even breaking the atoms—tearing apart the very fabric of what the original numbers are made of, and recombining the fragments in a novel way (with a completely different order structure): we get an entirely new set of numbers out of the operation. Observe that we still have infinitely many real-valued sequences in the equivalence class of the hyperreal number zero (those that differ from zero at only finitely many positions), but—in as far as they converge in the standard sense at all—only a strict subset of them converge to the real number zero.

6.2 *Remarks on the ultrapower construction*

When a free ultrafilter is applied in the ultrapower construction of the hyperreal numbers, its various properties affect the properties of the hyperreals in the following ways (see section 8.5):

- the upper set property of a filter is required to obtain an equivalence relation on $\mathbb{R}^{\mathbb{N}}$;

- the property of an ultrafilter, which ensures that each set is either large (in the filter) or small (in the corresponding ideal), is required to obtained trichotomy on $^*\mathbb{R}$ (*i.e.*, for each $r, s \in {}^*\mathbb{R}$ either $r < s$ or $r = s$ or $r > s$);

- the property of being free in combination with being ultra, which ensures that every finite set is small, is required to ensure that $\mathbb{R} \subsetneq {}^*\mathbb{R}$.

Although free ultrafilters can be proven to exist (given the usual set-theoretic assumptions), it can also be proven that no explicit example of them can be given; they are inherently non-constructible objects or "intangibles" (Schechter, 1997).

If we drop the condition of being free, and apply the Fréchet filter instead, we obtain a weaker but constructive model of the hypernatural numbers. Let us consider the implication for probability by considering the example of a fair lottery on $\mathbb{N}$. On the one hand, using a Fréchet filter would still allow us to obtain probability functions that take infinitesimal values for finite events. On the other hand, the system is too weak to obtain probability functions that are defined on all of $\mathcal{P}(\mathbb{N})$. For instance, the subset of odd numbers and the subset of even numbers are neither in the Fréchet filter nor in the corresponding ideal, so according to this filter and ideal they are neither large nor small, such that these events would not receive any probability value.

7 KOLMOGOROV'S AXIOMS FOR PROBABILITY THEORY

Since this theory does not contain actual infinitesimals, it may seem of less importance for the topic of this chapter. However, Kolmogorov's approach was very successful and influential: it lies at the basis of the contemporary presentation of probability theory as a special case of measure theory, which itself is a branch of real analysis (calculus).[17] Hence, any later proposal for a new theory of probability, possibly including infinitesimals, has to compete with it. Therefore, we do include Kolmogorov's axioms here, or at least an equivalent formulation thereof (taken from Benci, Horsten, & Wenmackers, 2013). P is the probability function and Ω is the sample space, a set whose elements represent elementary events:

(K0) DOMAIN AND RANGE. The events are the elements of $\mathfrak{A}$, a σ-algebra over Ω,[18] and P is a function $P : \mathfrak{A} \to \mathbb{R}$.

(K1) NON-NEGATIVITY. $\forall A \in \mathfrak{A}, P(A) \geq 0$.

(K2) NORMALIZATION. $P(\Omega) = 1$.

(K3) ADDITIVITY. $\forall A, B \in \mathfrak{A}$ such that $A \cap B = \varnothing$,

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

(K4) CONTINUITY. Let $A = \bigcup_{n \in \mathbb{N}} A_n$, where $\forall n \in \mathbb{N}, A_n \subseteq A_{n+1} \subseteq \mathfrak{A}$. Then

$$P(A) = \sup_{n \in \mathbb{N}} P(A_n).$$

The triple $(\Omega, \mathfrak{A}, P)$ is called a *probability space*.

---

17 For the incorporation of probability theory into measure theory, Kolmogorov's assumption of Countable Additivity was crucial. This move was motivated by mathematical convenience, rather than by philosophical reflection on the meaning of probability. Kolmogorov stated (with original italics):

> "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.*" (Kolmogorov, 1933, p. 15)

Later, de Finetti (1974, Vol. I, p. 119) would write about Countable Additivity:

> "it had, if not its origin, its systematization in Kolmogorov's axioms (1933). Its success owes much to the mathematical convenience of making the calculus of probability merely a translation of modern measure theory [...]. No-one has given a real justification of countable additivity (other than just taking it as a "natural extension" of finite additivity) [...]."

Compare to Schoenflies's reaction to Countable Additivity in Borel measure (footnote 57).

18 $\mathfrak{A}$ is a *σ-algebra* over Ω if $\mathfrak{A} \subseteq \mathcal{P}(\Omega)$ such that $\mathfrak{A}$ is closed under complementation, intersection, and countable unions. $\mathfrak{A}$ is called the *event algebra* or *event space*.

For our present purposes, the continuity axiom is the most important one, so let me briefly mention two aspects of it. First, (K4) uses a supremum, which is defined in terms of a standard limit; this limit is guaranteed to exist for real-valued functions, but not on the hyperreal numbers. Still, the general idea of this axiom can be phrased without reference to the specific limit operation. It can be regarded as a specific form of a more general idea: that is, to define the absolute probability of any event from an infinite domain as the limit (in some sense) of a sequence of conditional probabilities associated with that event, conditional on a suitable family of finite events. This more general principle was called the 'Conditional probability principle' in Benci et al. (2013, section 3.2) and Benci, Horsten, and Wenmackers (2018, section 3.2), where it was further shown how the same idea can be applied to hyperreal-valued probability functions (using a different kind of limit operation). Second, assuming the other axioms, (K4) is equivalent to requiring countable additivity, which is not compatible with hyperreal-valued probability functions (except in the trivial case of a finite domain).

PART III
AXIOMATIZATION OF INFINITESIMAL PROBABILITIES

In the historical overview, we have already encountered two approaches to probability theory that allow infinitesimal probabilities: the axiomatization of Nelson (1987) and the work of Loeb (1975). What is missing so far is an axiomatization of a theory that assigns probabilities to standard infinite sets (such as $\mathbb{N}$, on which Nelson's approach is silent) and that allows infinitesimal or other hyperreal values in the final result (unlike Loeb's approach, which is geared toward obtaining results in the standard domain). This is the purpose of the current part.

8   INFINITESIMAL PROBABILITIES AND COUNTABLE LOTTERIES

Within philosophy, infinitesimal probabilities have often been discussed in the context of the following example: a lottery on the natural numbers, $\mathbb{N}$, in particular a fair one (*i.e.*, a lottery in which each individual ticket receives the same probability as any other one). Since this example is so common, we discuss it first, before setting up a more general framework in the next section.[19] We start from a real-valued approach (in which zero is

---

19 In order to describe probability functions on infinite sample spaces, focusing on $\mathbb{N}$ as the sample space may seem like a very natural starting point, because $\mathbb{N}$ is the canonical example of a set with the smallest infinite cardinality. It will turn out that in some sense this problem is not the easiest one to describe, because it is in lockstep with other (less

the only infinitesimal) and investigate which modifications are required in order to allow for the assignment of non-zero infinitesimal probabilities.[20]

8.1 *Lotteries on initial segments of* $\mathbb{N}$

Ultimately, we want to describe a lottery, fair or weighted, on $\mathbb{N}$, but we start by considering a lottery, fair or weighted, on an arbitrary initial segment of $\mathbb{N}$: the sample space (set of atomic possible outcomes) is $\Omega_n = \{1, \ldots, n\}$. First, we introduce weights: a real number w_i for each of the elements i of Ω_n. Without loss of generality, we may assume these weights to be normalized, such that $\sum_{i=1}^{n} w_i = 1$ (*e.g.*, in a fair lottery $w_i = 1/n$ for all i). Then, we define the probability on Ω_n, P_n, of an arbitrary subset of $\mathbb{N}$, A, as follows:

$$P_n(A) = \sum_{i=1}^{n} w_i \times \#(A \cap \{i\}),$$

where # is the counting measure for finite sets. (This suffices: although A can be an infinite set, $A \cap \{i\}$ is empty or singleton.) In the case of a fair lottery, the probability $P_n(A)$ is just the relative frequency of A: the fraction of elements of A within Ω_n. That P_n is finitely additive follows directly from the counting measure being finitely additive.[21]

8.2 *Taking the limit*

Now, we want to consider a lottery on $\Omega = \mathbb{N}$, rather than on $\Omega_n = \{1, \ldots, n\}$. The idea is to consider the lottery on $\mathbb{N}$ as the limiting case of a sequence of finite lotteries. This idea seems apt, since we have $\Omega =$

---

obvious) occurrences of $\mathbb{N}$. Among the infinite sets, $\mathbb{N}$ is our usual benchmark, so we use it in and out of season. As a result, there are hidden symmetries in the problem of a (fair) lottery on $\mathbb{N}$, which make it harder to analyze it. To understand this statement, we first need to encounter the problems alluded to, so we will progress as planned, but I will return to this observation in the middle of section 8.3.

20 The current section presents some of the ideas originally developed in Wenmackers and Horsten (2013) in a more straightforward way.

21 For, consider a finite family of mutually disjoint subsets of $\mathbb{N}$, $\{A_k \mid k \in \{1, \ldots, m\}, A_k \subseteq \mathbb{N}\}$ (for some $m \in \mathbb{N}$) such that for each $i \neq j$, $A_i \cap A_j = \varnothing$. Defining the union of members of the family $A = \bigcup_{k=1}^{m} A_k$, we obtain for the probability of A:

$$
\begin{aligned}
P_n(A) &= \sum_{i=1}^{n} w_i \times \#(\bigcup_{k=1}^{m} A_k \cap \{i\}) \\
&= \sum_{i=1}^{n} w_i \times \sum_{k=1}^{m} \#(A_k \cap \{i\}) \\
&= \sum_{k=1}^{m} \sum_{i=1}^{n} w_i \times \#(A_k \cap \{i\}) \\
&= \sum_{k=1}^{m} P_n(A_k).
\end{aligned}
$$

$\lim_{n\to\infty} \cup_{i=1}^n \Omega_i$.[22] We will define the probability, P, for an arbitrary subset of $\mathbb{N}$, A, analogously to the limiting relative frequency:

$$P(A) = \lim_{n\to\infty} P_n(A).$$

Remarks:

- P is not defined for all subsets of $\mathbb{N}$.[23]

- Taking the limit of fair lotteries on Ω_n (where $P(\{i\}) = 1/n$ for any $i \in \Omega_n$) results in a fair lottery on $\mathbb{N}$, with $P(\{i\}) = 0$ for all $i \in \mathbb{N}$.

- For a fair lottery on $\mathbb{N}$, P is the *natural density* (also known as the *arithmetic density* or the *asymptotic density*).

- In a fair lottery, P is zero for all finite subsets as well as for some infinite ones (such as the set of squares and the set of primes),[24] unity for cofinite sets as well as for some infinite ones (such as the complements of the previous examples), and intermediate values for other infinite sets (such as arithmetic progressions[25] that receive probability $1/n$ for some n; *e.g.*, $1/2$ for the set of even numbers and for the set of odd numbers).

For those who have the intuition that the probability of a particular outcome in a fair lottery on the natural numbers ought to be *infinitesimal*, the above real-valued function P that assigns probability zero to such outcomes does fine: zero is *the* infinitesimal probability, the only one in the $[0, 1]$ interval of $\mathbb{R}$. Nevertheless, it may bother some that this function does not allow us to distinguish between the impossible event (represented by $A = \varnothing$) and some infinitely unlikely but possible events. The worry is that

---

22 On the other hand, the ordered set $(\mathbb{N}, <)$ is qualitatively different from any $(\Omega_n, <)$: unlike all of its initial segments, $\mathbb{N}$ does not have a last element. This observation is suggestive of taking a different kind of limit, which involves a hyperfinite set (which does have a last element) rather than a standard infinite one.

23 The collection of subsets for which P is defined does not form a σ-algebra. P can be extended to all of $\mathcal{P}(\mathbb{N})$ but the extension relies on Banach limits and is not unique. Whereas the usual limit relies on the notion of "eventuality" that can be captured by the Fréchet filter, which is a free filter that is constructively available, the Banach limit depends on a free ultrafilter on $\mathbb{N}$, which relies crucially on a non-constructive axiom (the ultrafilter principle, UF). See section 8.5 below for more details.

24 As such, this probability function can help us to make sense of Galileo's paradox, which revolves around the question of whether or not the set of perfect squares is smaller than the set of natural numbers (see Mancosu, 2009). As measured by the natural density, the answer to that question is affirmative: it assigns probability unity to the set of natural numbers and probability zero to the set of perfect squares. On the other hand, the function does not discriminate between a finite set, the set of perfect squares, and the set of primes.

25 Arithmetic progressions are sets of the form $a\mathbb{N} + b = \{n \in \mathbb{N} \mid n \mod a = b\}$ for some $a \in \mathbb{N}$ and some $b \in \{0, 1, \ldots, a - 1\}$.

the probabilities of these events are represented by the same infinitesimal, and since there can only be one zero (*i.e.*, neutral element under addition), this observation may motivate a search for *non-zero infinitesimals*. However, this worry may be partially addressed by considering a non-Archimedean ordering of the events, which is a question for qualitative probability theory[26] rather than for quantitative probability theory. Despite this, there is an underlying issue that cannot be addressed without considering numerical probabilities: it is that of additivity. We consider this in the next section.

8.3 *Additivity of P: finite, countable, or ultra*

We briefly mentioned (in Part 1) that Leibniz's approach to the calculus was based on infinite sums (integrals), unlike Newton's, for whom the notion of "fluxions" (derivatives) was more basic. Since infinitesimals were most prominent in Leibniz's approach, it should come as no surprise that the concept of infinitesimal probabilities is closely connected to foundational discussions concerning the additivity of probability values.

Skyrms (1983b) interprets the intuition that measures should be regular (that only the null set should receive measure zero) as a Zenonian intuition (*cf.* section 1.3.2 of the appendix): a whole of positive magnitude should not be made up of parts of measure zero. He argues that a principle of "ultra-additivity"[27] has been present, albeit often implicitly, in discussions concerning measures at least since the times of Zeno and Aristotle. Since the belief in ultra-additivity appears to be so deeply rooted in Western thinking about measures, it should not surprise us if it is present, whether presented as an explicit assumption or a tacit one, in many discussions about probability measures, too.

In fact, it was exactly such a principle that motivated my own search for a fair probability function on $\mathbb{N}$. My main motivation for wanting to assign non-zero probability to non-empty sets is that it should allow us to make arbitrary unions of events and obtain their probability by an addition rule for the individual probabilities (in the case of disjoint events, by taking the analogous arbitrary sum).[28]

---

26 Recall the work by de Finetti (1931) as discussed in section 1. See also Pedersen (2014), Easwaran (2014, p. 17), and Konek (this volume).

27 Ultra-additivity means additivity for arbitrary collections of disjoint events; it is sometimes called perfect additivity (see, *e.g.*, de Finetti, 1974, Vol. II, p. 118) or arbitrary additivity (Hofweber, 2014).

28 Wenmackers (2011, p. 36): "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." Unaware of the work

Let us return to the probability functions of the previous sections. Finite additivity obtains for such a P, like it does for all the functions P_n. Since the function P is the limit of the sequence of functions (P_n), each member of which has the property of finite additivity (FA), one might suspect P to have the limiting property of FA: countable additivity (CA). However, this is not the case: limiting relative frequencies are not CA, because the relevant limiting operations (from the construction of P and from the condition of CA) do not commute. To illustrate this, consider a countably infinite family of mutually disjoint subsets of $\mathbb{N}$, $\{A_k \mid k \in \mathbb{N}, A_k \subseteq \mathbb{N}\}$ such that for each $i \neq j$, $A_i \cap A_j = \varnothing$, and define the union of members of the family, $A = \bigcup_{k \in \mathbb{N}} A_k$. We say that CA holds for a function p if the following equality holds:

$$p(A) = \lim_{n \to \infty} \sum_{i=1}^{n} p(A_i). \tag{1}$$

In the case of P, we find for the lefthand-side of Equation (1):

$$P(A) = \lim_{n \to \infty} P_n(A)$$
$$= \lim_{n \to \infty} \sum_{i=1}^{n} w_i \times \lim_{m \to \infty} \sum_{k=1}^{m} \#(A_k \cap \{i\}).$$

Let us now consider a fair lottery (substituting $w_i = 1/n$) with $A_k = \{k\}$ such that $A = \mathbb{N}$; we find:

$$P(A) = \lim_{n \to \infty} (n \times 1/n)$$
$$= 1.$$

Then, we consider the righthand-side of Equation (1), applying it to P in the fair case, where $P(A_i) = 0$ for all i:

$$\lim_{n \to \infty} \sum_{i=1}^{n} P(A_i) = \lim_{n \to \infty} \sum_{i=1}^{n} 0$$
$$= 0.$$

Clearly, 0 is not equal to 1, so CA does not obtain for P, the real-valued probability function for a fair lottery on the natural numbers.

---

by Skyrms (1983b), Wenmackers and Horsten (2013, p. 40) clumsily referred to a "SUM" intuition: "SUM [is the intuition that] [t]he probability of a combination of tickets can be found by summing the individual probabilities. [...] 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 this case. In other words, we will argue that the implementation of SUM is not as straightforward an affair as is commonly thought."

The righthand-side requires us to consider the function P and thus to take the limit of n to infinity of $P_n(\{i\}) = 1/n$ first, which is zero; taking the limit of a sum of zeros is zero. The lefthand-side requires us to consider P_n. Sure, as n increases, $P_n(\{i\})$ tends to zero for any $i \in \Omega_n$ (like $1/n$), but the sum of all singleton probabilities is in lock-step with this decrease: $n \times 1/n = 1$, such that the sum of probabilities of all singletons equals the probability of the entire sample space (total number of tickets times probability of each ticket), which is unity. This is just FA and it holds for any n, no matter how large. It also holds that $\lim_{n\to\infty}(n \times 1/n) = 1$, but this cannot be read as "the number of tickets times the probability of each ticket". It is no additivity principle and it does not suggest an alternative way of obtaining a real-valued probability function either.[29] Yet, it does suggest the following: that the singleton probability in a fair lottery on the natural numbers ought to be a non-zero infinitesimal, such that some sort of infinite sum over them can result in a non-zero (and non-infinitesimal) value corresponding to the probability of the corresponding union of events. In particular, the sum can be unity if we sum the probabilities of all point events.[30]

There is another strange aspect to setting $P(\{n\}) = 0$ for all $n \in \mathbb{N}$: it is not so much that it can be used to represent a fair lottery on $\mathbb{N}$, but rather that it can also represent the limit of many kinds of non-fair probability distributions. Consider, for instance, finite lotteries in which (i) the set of even numbers is double as likely as the set of odd numbers, (ii) all even numbers are equally likely and (iii) all odd numbers are equally likely. For the limit of such weighted lotteries, too, we would have to assign probability zero to all singleton events (and thus obtain a fair distribution in the limit).[31]

8.4 *Diagnosis*

Within the context of standard probability theory, we have a single infinitesimal probability at our disposal: zero. Even for a lottery on a sample space that is countably infinite, the lowest infinite cardinality, this turns out to be too little for three reasons.

1. Across lotteries, it does not allow us to obtain different singleton probabilities for limits of sequences of qualitatively different finite

---

29 Although this idea is suggestive of a *procedure* for assigning probabilities in such a way that we can make sense of infinite sums, it does not allow us to define a probability *function*.

30 Recall the quote on p. 1 by de Finetti (1974, p. 347) concerning the absurdity of $0 + 0 + 0 + \ldots + 0 + \ldots = 1$. It turns out that this idea is false if the sum represents the usual, countably infinite sum: such a sum is not defined for infinitesimal terms.

31 As far as I know, this worry has not yet appeared in the literature.

lotteries (*e.g.*, finite lotteries that assign equal probability to even and odd versus finite lotteries that do not).

2. Within a fair lottery, it does not allow us to discriminate between the probability of many events that are strict subsets of each other (*e.g.*, all perfect squares versus a single perfect square).

3. Within a fair lottery, it does not allow us to define an adequate infinite additivity principle; alternatively, if we insist on countable additivity, it does not allow us to describe a fair lottery on the natural numbers.

The first reason is related to a more general observation: like any real number, zero is the limit of qualitatively different sequences (of rational or real numbers). In particular, sequences may differ in their speed of convergence. The rate at which the corresponding sequence of relative frequencies tends to zero is smaller for a singleton event (the convergence of $1/n$ is linear) than it is for the set of perfect squares (the convergence of $1/n^2$ is quadratic) or for the powers of two (the convergence of $1/2^n$ is exponential). This suggests that within the collection of sequences that are considered to be infinitesimal, and thus to converge to zero, some are smaller than others (even though their limits are all defined to be zero when working within the real numbers). This brings us to reconsider what the real number zero is, continuing along the lines set out in the introduction, and to define an alternative limit operation on sequences. One way to achieve this is found in the construction of a non-standard model of a real closed field as was shown in section 6.

8.5 *Alternative approach with non-zero infinitesimal probabilities*

We apply the equivalence relation discussed that is used to construct the hyperreals (section 6) to the sequence of relative frequencies belonging to initial segments of $\mathbb{N}$. This results in a different kind of probability function, which takes its values in the $[0,1]$ interval of the hyperreal numbers.[32]

Wenmackers and Horsten ([2013]) assumed all of NSA as given, whereas we mainly needed this alternative equivalence relation on the sequences of relative frequencies in order to obtain a hyperreal-valued probability value on $\mathbb{N}$ that allows for an infinite additivity principle.

Now that we know the outlines of our labyrinth, we can drastically reduce the length of our escape route. With the benefit of hindsight, we

---

32 Actually, it is more accurate to say: *a* set of hyperreal numbers (*cf.* footnote 1), because the result of the construction depends on the free ultrafilter and there are uncountably many. We do not dwell on the issue of non-uniqueness now, but we will come back to it in section 14.

see ways to obtain our results with much less baggage. One way, which is suitable only for fair lotteries and which is alluded to in the 2013 paper, is to assume a numerosity function on $\mathbb{N}$ and to normalize it. Numerosity theory has been developed to address some of the very same problems that are also discussed in the literature on a fair lottery on $\mathbb{N}$ (Benci & Di Nasso, 2003; Mancosu, 2009). The main difference is that it is not a probability function but a measure of set size that should coincide with the usual counting measure for finite sets, so it is not normalized and assigns unity to singletons rather than to $\mathbb{N}$. However, because of its nice algebraic properties, normalizing the numerosity function, in order to obtain a fair probability measure, does not cause any complications at all.

Alternatively and more elegantly, one could set up an axiomatic system that states the existence of probability functions on $\mathbb{N}$ that may assign non-zero values to singleton outcomes (possibly all equal) and repurpose the previous results in order to prove its consistency.

For instance, consider this proposal for the axioms governing P.

EVERYWHERE DEFINED. P is defined on all subsets of $\mathbb{N}$: its domain is the powerset of $\mathbb{N}$, $\mathcal{P}(\mathbb{N})$.

HYPERREAL-VALUED. The range of P is the unit interval of some suitable field $\mathcal{R}$.

REGULAR. $P(A) = 0$ iff $A = \varnothing$.

NORMALIZED. $P(\mathbb{N}) = 1$.

FINITELY ADDITIVE. $\forall A, B \in \mathcal{P}(\mathbb{N})$ if $A \cap B = \varnothing$, then $P(A \cup B) = P(A) + P(B)$.

ULTRA-ADDITIVE. For any collection of mutually disjoint subsets of $\mathbb{N}$[33] an analogous additivity property holds.

We do not prove the joint consistency of the proposed axioms here: it is a consequence of what preceded and can be viewed as a special case of the proof in Benci et al. (2013).

8.6 *Examples*

Now that we have seen that there exists a hyperreal measure that captures the idea of a uniform probability distribution over the natural numbers, let's illustrate some consequences. In this section, P always refers to such a distribution. (For proofs, see Benci et al. 2013.)

---

33 The collection can have an arbitrary cardinality, although, of course, at most countably many of its members can be non-empty.

By assumption, P assigns the same infinitesimal probability to any singleton outcome of the lottery. If we regard P as a normalized numerosity function, we see that $\forall n \in \mathbb{N}$, $P(\{n\}) = 1/\alpha$, where $\alpha \in {}^*\mathbb{N} \setminus \mathbb{N}$ is the numerosity of $\mathbb{N}$.

For any finite set $A \subset \mathbb{N}$, the numerosity equals the finite cardinality (#), so: $P(A) = \#(A)/\alpha$, which is an infinitesimal. For example, $P(\{1, 2, 4, 8, 16, 32\}) = 6/\alpha$.

For an infinite subset B, $P(B)$ differs by at most an infinitesimal from the natural density of B (if the latter exists). For example, if B is the set of even numbers, the natural density is $1/2$ and either $P(B) = 1/2$ (if the even numbers are in the free ultrafilter used to construct P) or $P(B) = (1 - 1/\alpha)/2$.

For a set that lacks a natural density, P is infinitesimally close to some Banach limit. Different Banach limits of the same set and Ps constructed by a different free ultrafilter can differ by more than an infinitesimal amount. (See Kerkvliet and Meester 2016 for an example.) In particular, there are subsets of $\mathbb{N}$ for which the possible P-values range from an infinitesimal to one minus an infinitesimal. This range can be regarded as a measure of how pathological a set is.

9   MORE SCENARIOS INVOLVING INFINITESIMAL PROBABILITIES

In the previous section, we discussed one particular scenario that involves infinitesimal probabilities: a lottery on the set of natural numbers. In this section, we give a more comprehensive overview of common examples that feature in discussions of infinitesimal probabilities. Then we show how we can generalize the approach of the previous section to an all encompassing theory that is able to assign infinitesimal probabilities to all of these scenarios.

9.1   *Common examples*

We list the examples involving infinitesimal probabilities below, sorted by increasing cardinality of the sample space: finite, countably infinite or uncountably infinite.

First, there are some examples with **finite sample spaces** that allow for infinitely small differences in probability among the possible outcomes. The simplest such case is that of an *almost* fair coin toss, in which there is an infinitesimal advantage to one of the sides.

Second, there are examples with **countably infinite sample spaces**, in particular with uniform probability distributions. We already discussed the most common example of this kind: a lottery on the set of natural numbers,

in particular a fair one. A fair lottery on $\mathbb{N}$ is also known as the de Finetti lottery (Bartha, 2004) or God's lottery (McCall & Armstrong, 1989). In this category, there are also fair lotteries on other countable sets, such as $\mathbb{Z}$, $\mathbb{Q}$ and the unit interval of the rational numbers: $[0,1]_{\mathbb{Q}}$. Discussions of non-uniform probability distributions on countable domains are less common, but they do exist, especially in the context of discussions of the incompatibility between CA and uniform probability distributions on countable domains.[34]

Third, there are examples with **uncountable sample spaces**, with uniform and arbitrary probability distributions. Two popular ways of presenting this is as throwing darts uniformly at the unit interval of the real numbers ($[0,1]_{\mathbb{R}}$) (*e.g.*, Bernstein & Wattenberg, 1969) or as a fair spinner with unit circumference ($[0,1[_{\mathbb{R}}$) (*e.g.*, Skyrms, 1995; Barrett, 2010).[35] Three-dimensional variations on this theme include the uniform probability on a unit sphere and the associated Borel–Kolmogorov paradox of a meridian versus the equator. A different way of obtaining an uncountable domain is by considering a countably infinite sequence of stochastic processes, each with a countable number of possible outcomes. The most common example of this kind is an infinite sequence of tosses with a fair coin (in which the outcomes of the tosses are taken to be statistically independent: an infinite Bernoulli process) (*e.g.*, Skyrms, 1980; Williamson, 2007; Weintraub, 2008).[36]

Categorizing a probabilistic problem by one of these three labels need not be final. Once we have a method of representing probability distributions on uncountable domains, we may arrive back at the finite and countably infinite case by conditionalization (assuming the relevant events are measurable) (*cf.* Skyrms, 1983b). It may also happen that we want to replace a finite sample space by an infinite refinement of it (for instance, a suitable product space of the initial sample space). For instance, Pedersen (2014, p. 827) mentions a case in which "an agent's state of belief cannot rule out arbitrarily deep[ly] nested subdecompositions of a finite decomposition of a dartboard".

---

34 For instance, Kelly (1996) has reflected on the consequences of denying the existence of a fair infinite lottery: this would have the strange implication 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.

35 This example was also mentioned in Lewis (1980), as well as many others.

36 It should be noted that Skyrms (1980) refers to the work of Bernstein and Wattenberg (1969), but they only described a hyperreal-valued probability measure on subsets of $[0,1]$. However, for assigning infinitesimal probabilities to infinite sequences of coin tosses, a hyperreal-valued probability measure on subsets of $\{0,1\}^{\mathbb{N}}$ would be needed instead. Yet, the informal account given by Skyrms (1980, pp. 30–31) is consistent with later developments of hyperreal probability functions on $\{0,1\}^{\mathbb{N}}$ (see, *e.g.*, Benci et al., 2013).

Some of these scenarios cannot be described by standard probability theory, whereas others—it has been argued—cannot be described adequately by it, or would benefit from an alternative treatment involving infinitesimal probabilities. So far, we have seen isolated recipes for hyperreal-valued probability functions: Bernstein and Wattenberg (1969) gave a recipe to assign uniform probabilities to subsets of the unit interval of the real numbers. And, in the previous section, we discussed a recipe for assigning regular probabilities to the canonical countably infinite sample space, $\mathbb{N}$. In the end, we would like to have a method that is fully general, which can be applied to all the examples above, and more. We describe such a method below.

9.2   *Non-Archimedean probability (NAP) theory*

In this section, we will review some crucial elements that allow us to generalize the approach from section 8.[37] In section 8.5, we replaced the standard limit operation that associates at most one real number with a sequence of (possibly weighted) relative frequencies by a non-standard limit that associates a hyperreal number with each of these sequences. Sequences can be thought of as functions from $\mathbb{N}$ (the index set) to some set, X. In the case of relative frequencies $X = \mathbb{Q}$, but in general we allow real-valued weights, so then $X = \mathbb{R}$. Both the standard and the non-standard limit operation can be understood such as to involve a filter on the index set (the Fréchet filter on $\mathbb{N}$ and a free ultrafilter on $\mathbb{N}$, respectively).

A probability function has to assign values to sets in $\mathcal{P}(\mathbb{N})$, not to $\mathbb{N}$ itself, so the appropriateness of using countable sequences and filters on $\mathbb{N}$ to set up such a function is not immediate, even in cases in which the sample space is countable. Observe that we used the countable indices to correspond to the relative frequencies of initial segments of $\mathbb{N}$. Since the usual ordering of the natural numbers induces a natural ordering on this collection of initial segments, we are able to work with sequences of the corresponding relative frequencies and with filters on $\mathbb{N}$.

Our choice for the collection of initial segments may seem self-evident, because we are familiar with it from the context of natural density, but it is not canonical: we could have considered $\mathcal{P}_{\text{fin}}(\mathbb{N})$, the collection of all finite subsets of $\mathbb{N}$ (or those except the empty set, $\mathcal{P}_{\text{fin}}(\mathbb{N}) \setminus \varnothing$). In that case, we can slightly generalize the approach: $\mathcal{P}_{\text{fin}}(\mathbb{N})$ with the subset

---

37 The information given here suffices to get a rough idea of the approach. Further details (for instance, restrictions on the free ultrafilter to secure certain properties of the resulting probability functions) can be found in Benci et al. (2013).

ordering forms a directed set.[38] We can use this directed set as an index set, instead of $\mathbb{N}$, obtaining a generalized sequence, also called a 'net' (see, *e.g.*, Schechter, 1997, pp. 157–158): a function from a directed set, which serves as the index set, to a set, X. Filters on $\mathbb{N}$ are a special case of this more general set-up, since they are collections of subsets of $\mathbb{N}$ that can be directed by the subset relation.

If we want to assign probability functions to subsets of some sample space Ω other than $\mathbb{N}$, we can follow a similar approach: change the relevant index set to $\mathcal{P}_{\mathrm{fin}}(\Omega) \setminus \varnothing$. In this case, we also have to consider free ultrafilters on Ω.

These are the axioms for Non-Archimedean Probability (NAP) theory from Benci et al. (2013), where the triple (Ω, P, J) is called a 'NAP-space:

- (NAP0) **Domain and range.** The events are all the elements of $\mathcal{P}(\Omega)$ and the probability is a function

$$P : \mathcal{P}(\Omega) \to \mathcal{R}$$

  where $\mathcal{R}$ is a superreal field

- (NAP1) **Non-negativity.** $\forall A \in \mathcal{P}(\Omega)$,

$$P(A) \geq 0$$

- (NAP2) **Normalization.** $\forall A \in \mathcal{P}(\Omega)$,

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

- (NAP3) **Additivity.** $\forall A, B \in \mathcal{P}(\Omega)$ such that $A \cap B = \varnothing$,

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

- (NAP4) **Non-Archimedean continuity.** $\forall A, B \in \mathcal{P}(\Omega)$, with $B \neq \varnothing$, *let $P(A|B)$ denote the conditional probability, namely*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{2}$$

  Then

---

38 A directed set $(X, \preccurlyeq)$ is a special case of a preordered set (see, *e.g.*, Schechter, 1997, p. 52). A preordered set is a pair $(X, \preccurlyeq)$ consisting of a set X and a preorder $\preccurlyeq$ on X, *i.e.*, a relation on X that is transitive (for all $x, y, z \in X$, if $x \preccurlyeq y$ and $y \preccurlyeq z$ then $x \preccurlyeq z$) and reflexive (for all $x \in X$, $x \preccurlyeq x$). For a directed set, there is an additional condition on the preorder:

$$\forall x_1, x_2 \in X, \ \exists y \in X : (x_1 \preccurlyeq y \wedge x_2 \preccurlyeq y).$$

- $\forall \lambda \in \mathcal{P}_{fin}^0(\Omega),\ P(A|\lambda) \in \mathbb{R}^+$;
- there exists an algebra homomorphism

$$J : \mathfrak{F}\left(\mathcal{P}_{fin}^0(\Omega), \mathbb{R}\right) \to \mathcal{R}$$

such that $\forall A \in \mathcal{P}(\Omega)$

$$P(A) = J\left(\varphi_A\right)$$

where
$$\varphi_A(\lambda) = P(A|\lambda) \text{ for any } \lambda \in \mathcal{P}_{fin}^0(\Omega).$$

Axiom (NAP4) specifies P for an infinite sample space Ω as a non-standard limit of probability functions restricted to (or conditionalized on) finite subsets of Ω.

Some properties of NAP theory:

- NAP theory produces regular probability functions. Hence, they allow us to conditionalize on any possible event by a ratio formula (*i.e.*, any subset of the sample space, except the empty set).

- Within NAP theory, the domain of the probability function can be the full powerset of any standard set from applied mathematics (*i.e.*, of any cardinality), whereas the general range is a non-Archimedean field. Hence, there are no non-measurable sets.

- Kolmogorov's countable additivity (which is a consequence of the use of standard limits) is replaced by a different type of infinite additivity (due to the use of a non-Archimedean limit concept).

- For fair lotteries, the probability assigned to an event by NAP theory is directly proportional to the numerosity of the subset representing that event.

- NAP functions are external objects: they cannot be obtained by taking a standard object (such as a family of standard sets) and applying the star-map to it.

A price one has to pay for all this is that certain symmetries, which hold for standard measures, do not hold for NAP theory. This theory is closely related to numerosity and has a similar Euclidean property: a strict subset has a smaller probability, as is necessary by regularity. Hence, for infinite sample spaces, NAP is bound to violate the Humean principle of one-to-one correspondence. This principle requires that if the elements of a given set can be put in a one-to-one correspondence with the elements of another set, then there 'size'—or in this case, probability—be equal.

Translation symmetries require that $P(A) = P(A + t)$ (with $A, A + t \subseteq \Omega$ and $A + t = \{a + t \mid a \in A\}$). Since this amounts to a particular type of one-to-one correspondence, these symmetries are not guaranteed to hold in NAP (*cf.* Williamson (2007), Parker (2013), and section 14.1), although they can hold up to an infinitesimal (Bernstein & Wattenberg, 1969). Bartha (2004) and Weintraub (2008) have pointed out before that these measures are strongly label-dependent, but it is probably more accurate to say that once events have been embedded in a sample space (*i.e.*, each event is described as a particular subset of a particular sample space Ω), this embedding needs to be applied in a consistent way henceforth (Hofweber, 2014; Benci et al., 2018).

For more details and proofs, see Benci et al. (2013). In the next part, we elaborate on the motivation for and the philosophical discussion of these results.

PART IV

PHILOSOPHICAL DISCUSSION

10 MOTIVATIONS FOR INFINITESIMAL PROBABILITIES

In the foregoing parts, we have encountered motivations for introducing infinitesimal probabilities as given by various authors. Most of these motivations occurred in the context of a particular interpretation of probability, with some arguing for the relevance of infinitesimal chances and others advocating for the introduction of infinitesimal credences. In this section, we search for the leitmotifs that arise from this polyphony.

Let us first revisit Bernstein and Wattenberg (1969): although they gave a probabilistic scenario as the motivation of their paper, the technical details of their results do not depend on the interpretation in terms of probability. If we want a measure of length that allows us to represent the length of countable collections of points as a non-zero infinitesimal, we can use the result of Bernstein and Wattenberg (1969) without modification. On the one and, it may fit even better in such a context, since the Lebesgue measure was originally motivated as an idealization of length measurements. Hence, obtaining a non-standard measure that is infinitely close to Lebesgue measure (at least, where the latter is defined) can be regarded as an alternative idealization of length measurements. On the other hand, the request for representing the measure of non-null countable sets as an infinitesimal may seem especially pressing when this measure is a measure of probability (rather than length). This motivation may be formulated as follows: probability measure should be maximally sensitive to distinguish

possibility from impossibility. Indeed, we have encountered this motivation for infinitesimal probabilities via regularity at various instances throughout this chapter.

Depending on the context, this motivation is related to a different kind of modality:

- objective probability: some chance (quantifying an ontic possibility)

- subjective probability: open-mindedness (quantifying an epistemic possibility)

We have encountered the epistemic motivation under the name 'strict coherence' and 'regularity'. Alan Hájek (2012b, p. 1 of draft) "canvass[ed] the fluctuating fortunes of a much-touted constraint, so-called *regularity*", which "starts out as an intuitive and seemingly innocuous constraint that bridges modality and probability, although it quickly runs into difficulties in its exact formulation". He takes "to be its most compelling version: a constraint that bridges doxastic modality and doxastic (subjective) probability". Kenny Easwaran (2014) presented regularity as a normative constraint on rational credences, which are related to doxastic modality, but he adds that other authors allow for various transmodal connections. Dennis Lindley called this demand, that prior probabilities of zero or one should only be assigned to logical truths or falsehoods, "Cromwell's rule".[39] Regarding the ontic motivation, Thomas Hofweber (2014) introduced a minimal constraint (MC) on the proper measurement of chances, which is akin to but not quite the same as regularity, which can be phrased in relation to various modalities. He concludes that (p. 6): "In the regularity principle, modality is best understood as epistemic, and chance is best understood as credence. In (MC) chance should be understood as objective chance."

At the root of this common motivation for infinitesimal chances and infinitesimal credences, there may be an even more basic motivation or implicit assumption, which Skyrms (1983b) called the principle of 'ultra-additivity' (and which also constituted my main motivation for starting a research project on infinitesimal probabilities). We discussed this in section 8.3 (see also appendix section 1.3.2). Thus, the motivation for

---

39 This is a reference to the following phrase from a 1650 letter by Oliver Cromwell: "I beseech you, in the bowel of Christ, think it possible you may be mistaken" reprinted in, Carlyle (1845). Like strict coherence, Cromwell's rule is clearly intended as a criterion for open-mindedness: even a well-confirmed theory like Einstein's is not as certain as a logical truth. Lindley (1991, p. 104) asked us to "leave a little probability for the moon being made of green cheese; it can be as small as 1 in a million, but have it there since otherwise an army of astronauts returning with samples of the said cheese will leave you unmoved". And Lindley (2006, p. 91) linked this open-mindedness criterion also to the Jain maxim "It is wrong to assert absolutely". (This was probably influenced by statistician Kantilal Mardia, who practised Jainism.)

introducing infinitesimal probabilities can be summarized by the following slogan:[40]

> Without infinitesimals, probabilities just don't add up.

11 ALTERNATIVES TO HYPERREAL PROBABILITIES

11.1 *Other ways to introduce infinitesimal probabilities*

There do exist ways to formalize infinitesimals other than Robinson's hyperreal numbers. One of them is smooth-infinitesimal analysis (SIA), which describes nilpotent infinitesimals: non-zero numbers whose square is zero. This system relies on intuitionistic logic. However, I am not aware of any proposals for smooth-infinitesimal probabilities.

Then there is the class of Conway numbers, which includes the infinitesimals from any non-standard field. This option has been suggested for application to probability theory, for instance, by Hájek (2003) (see section 12 below) and by Easwaran (2014). I, too, believe this can be a fertile approach. A first proposal has been offered by Chen and Rubio (2018), but it is too early to evaluate it here.

11.2 *Related approaches without infinitesimals*

Besides the possibility of introducing infinitesimals within a different framework, there are also relations between hyperreal infinitesimals and systems that do not include any infinitesimal numbers at all. For instance, one may combine an Archimedean quantitative probability theory (in particular, the orthodox approach with real-valued probability functions), with a non-Archimedean qualitative probability theory.[41] Moreover, Joseph Y. Halpern (2010) revealed some deep connections between hyperreal-valued probability functions, conditional probabilities (including Popper functions; see also Vann McGee, 1994), and lexicographic probabilities. Recently, Brickhill and Horsten (2018) have given a representation theorem that relates NAP functions and Popper functions; they also gave a lexicographic representation.

Skyrms (1983a) considered three ways of giving probability assignments a memory. One of his proposals was to "utilize orders of infinitesimals to implement long term-memory", such that "[s]uccessive updatings do not destroy information, but instead push it down to smaller orders of

---

40 Benci et al. (2018) listed perfect additivity as one among four desiderata for their theory, the others being: regularity, totality, and weak Laplacianism.

41 This was suggested by de Finetti, *cf.* section 1. See also the discussion of the 'numerical fallacy' by Easwaran (2014).

infinitesimals" (p. 158). He evaluated this proposal as having a certain theoretical simplicity, but lacking practical feasibility. However, given that the proposal essentially boils down to introducing lexicographical probabilities, it may turn out that this judgment was too harsh.

11.3 *Yet another point of view*

Introducing non-standard probabilities amounts to changing the range of the probability function. Skyrms (1995) considered an alternative way to achieve strict coherence, which involves changing the domain, such that the events to which infinitesimal probabilities are assigned in the previous approach are no longer in the event space at all. In this context, he cites (Jeffrey's translation of) Kolmogorov (1948):

> "The notion of an elementary event is an artificial superstruc-
> ture imposed on the concrete notion of an event. In reality,
> events are not composed of elementary events, but elementary
> events originate in the dismemberment of composite events."

Let me unpack this. In Kolmogorov (1933)'s approach, the sample space was assumed to contain all fully specific possible outcomes: the elements of the sample space are called 'elementary events'. On the other hand, we have the informal notion of concrete events or possible outcomes, which does not presuppose infinite precision. Here we see that Kolmogorov (1948) rejected his former approach in favour of a more realistic one: if we take into account the limited precision of any physical measurement, we can distinguish outcomes only with limited precision, too. With increasing precision, we can decompose events into more fine-grained ones, but not up to elementary precision.

Although no infinitesimal probabilities occur in the second approach, it is still relevant in the context of the current chapter, because of an interesting analogy: in both cases, starting from the orthodox approach, a symmetry is quotiented out to arrive at the new structure (*cf.* the reference to quotient spaces in the introduction).

12 INTERPLAY BETWEEN INFINITESIMAL PROBABILITIES AND INFINITE UTILITIES: PASCAL'S WAGER

We have seen in section 3, that discussions of rational degrees of belief often proceed via a betting interpretation (*e.g.*, motivating adherence to the axioms of probability theory by the avoidance of a sure loss). As such, they involve considerations of monetary loss or gain. However, the subjective value of money need not be linear. Therefore, it is useful to introduce 'utility' as a more abstract measure that represents subjective

worth directly. Utility is usually taken to be a real-valued (interval scale) measure.

However, non-Archimedean probabilities do not mix well with real-valued utilities. Hence, to deal adequately with infinitesimal probabilities in the context of decision theory, a non-Archimedean utility theory is needed, such as the one developed by Pivato (2014).

We consider the famous example of Pascal's wager. With this argument, found in his *Pensées*, Pascal purported to show that it is rational to wager for God's existence. In modern terminology, we have to consider all combinations of the existence or non-existence of God, on the one hand, and an agent's belief or disbelief in God, on the other hand. This leads to four cases each with its own expected utility. In the case that God exists, it is assumed that there are everlasting heavenly rewards for those who believe (positive infinite expected utility) and everlasting infernal punishments for those who disbelieve (negative infinite expected utility). In the case that God does not exist, there are a lifetime of earthly burdens for those who believe (negative finite expected utility) and a lifetime of earthly pleasures for those who disbelieve (positive finite expected utility). If the agent is maximally uncertain about the existence of God (assigning 50% probability to the possibility of existence and 50% probability to the possibility of non-existence), the expected utility of believing is infinitely better than that of disbelieving. So, according to this argument, if one has to wager, it is better to wager for God's existence.

In the context of a discussion of Pascal's wager, Oppy (1990, p. 163) considered the epistemic possibility "that the probability that God exists is infinitesimal", in which case "the calculation of the expected return of a bet on [the existence of] God is no longer as straightforward as the initial argument suggested".

Following up on this suggestion, Hájek (2003) considered whether salvation has an infinite utility. He mentioned two formal approaches that allow us to tell apart various infinite expectation values that occur in Pascal's wager and related problems. Hájek mentions NSA as one possibility of dealing with infinitesimal probabilities and infinite utilities, but he favours Conway's numbers, citing their ingenuity and user-friendliness. He speculates that such a formal approach can illuminate a whole scala of problems involving infinitesimal probabilities (such as the two envelope paradox).

On p. 38, Hájek wrote that "the infinitesimal probability can 'cancel' the infinite utility so as to yield a finite expectation for wagering for God". The idea of cancelling is indeed what NSA allows us to formalize: each infinitesimal is the reciprocal of an infinite number and vice versa. Multiplying an infinite hyperreal number and its multiplicative inverse, a particular infinitesimal, yields unity. So, on the one hand, we may obtain finite (non-infinite) and non-infinitesimal values by multiplying infinite

and infinitesimal numbers. On the other hand, there are also combinations of infinite and infinitesimal numbers whose product is an infinitesimal or an infinite number. More details can be found in Wenmackers (2018). For a treatment with surreal probabilities and utilities, see Chen and Rubio (2018): there approach also allows them to treat the St. Petersburg paradox.

13 THE LOCKEAN THESIS AND RELATIVE INFINITESIMALS

Whereas standard probability measures may seem too coarse-grained for some applications, where we would like to distinguish between possible and impossible events, they may not seem coarse-grained enough for other applications, as we will see in this section.

Suppose that you have detailed knowledge of the probabilities in a given situation. It has been argued that it may still be beneficial to hold some full (dis-)beliefs (Foley, 2009). But when is it rational to believe something in this case? The Lockean thesis suggests that it is rational to believe a statement if the probability of that statement is sufficiently close to unity.[42] This is usually modelled by means of a probability threshold. As is demonstrated by the Lottery Paradox (Kyburg, 1961), the threshold-based model is incompatible with the Conjunction Principle. Moreover, it can be objected that the actual probabilities are too vague to put a sharp threshold on them, and that a threshold should be context-dependent.

Based on certain analogies between large and infinite lotteries, Wenmackers (2012) suggested the use of NSA to introduce a form of vagueness or coarse-graining and context-dependence in the formal model of the Lockean thesis.[43] Hrbáček (2007) developed relative or stratified analysis, an alterative approach to NSA that contains 'levels' as a formalization of the intuitive scales-of-magnitude concept. Applying Hrbáček's framework, Wenmackers (2013) introduced 'Stratified Belief' as an alternative formalization of the Lockean Thesis.[44]

The basic idea is to interpret the Lockean thesis as follows: it is rational to believe a statement if the probability of that statement is indistinguishable from unity (in a given context). The context-dependent indistinguishability relation is then modelled using the notion of differences up to a level-dependent, ultrasmall number. These ultrasmall numbers, also called 'relative infinitesimals', are ordinary real numbers, which are merely unobservable, or do not have a unique name, in a given context. The aggregation rule for this model is the 'Stratified conjunction principle', which entails that the conjunction of a standard number of rational beliefs is rational,

---

42 This is reminiscent of the concept of 'moral certainty'; see also footnote 78.
43 An earlier version can be found in Wenmackers (2011, Ch. 4).
44 An earlier version can be found in Wenmackers (2011, Ch. 3).

whereas the conjunction of an ultralarge number of rational beliefs is not necessarily rational.[45]

14   RECENT OBJECTIONS AND OPEN QUESTIONS

In this section, we give a brief overview of developments from the two last decades in which new objections against and defences for infinitesimal probabilities have been added to the literature. It may be too early to evaluate the most recent collection of attempted refutations and acclaims for infinitesimal probabilities. Still, we briefly mention some here. More discussion can be found in , Benci et al. (2018)

14.1   *Symmetry constraints and label invariance*

In a number of publications, Paul Bartha applied ideas from non-standard measure theory to problems in the philosophy of probability. Bartha and Hitchcock (1999) used NSA in the usual way, *i.e.*, in order to obtain a real-valued probability function. Bartha and Johns (2001) also considered the application of NSA to a probabilistic setting, but they favoured a simpler appeal to symmetry in order to obtain the conditional probabilities relevant to their problem. (Later, Bartha (2004) discussed de Finetti's lottery and used infinitesimal probabilities as one way to escape the conclusion that CA is mandatory, since they exhibit hyperfinite additivity instead.)

Considering the case of an ω-sequence of coin tosses, Timothy Williamson (2007) demonstrated the incompatibility between infinitesimal probabilities and requiring the equiprobability of what he called "isomorphic events", which are "events of exactly the same qualitative type" (p. 175). In particular, for ω-sequence of coin tosses, he argues that the probability assigned to the event should not depend on when exactly the tossing started. Williamson contrasted his finding with that of Elga: whereas Elga (2004) found regularity to lead to too many eligible non-standard distributions, Williamson found regularity in combination with what he called "non-arbitrary constraints" to rule out all candidate distributions.

Ruth Weintraub (2008) attempted to demonstrate that Williamson's argument depends on the assumption of label-independence, which is itself incompatible with infinitesimal probabilities. More recently, Benci et al. (2018) have analyzed Williamson's argument in the light of NAP theory. They, too, concluded that isomorphic events cannot be assigned equal hyperreal-valued probabilities without contradicting the assumptions on which this theory relies. Simultaneously, Howson (2017) argued—without

---

45 Although this model is intended to describe beliefs that are almost certain, it can be used for weaker forms of belief by substituting a lower number instead of unity.

using any details of NAP theory—that "it is not regularity which fails in the non-standard setting but a fundamental property of shifts in Bernoulli processes". Recently, Matthew W. Parker (2018) has argued that these objections to the argument of Williamson (2007) fail.

14.2   *Non-uniqueness of hyperreal probabilities*

Adam Elga (2004) considered the zero-fit problem of the 'best system' analysis of laws: if all systems of laws assign probability zero to the actual history up to now, then one cannot identify the best system based on a measure of goodness-of-fit. He entertained the option of applying non-standard probability functions and thus to assign a non-zero infinitesimal probability to the actual history, thereby escaping a zero fit. Ultimately, however, he rejected this proposal:

> "We have required our nonstandard probability function to be regular, and to approximate given standard probability functions. But those requirements only very weakly constrain the probabilities those functions assign to any individual outcome. [...] And the fit of a system associated with such a function is just the chance it assigns to actual history. So the fit of such a system indicates nothing about how well its chances accord with actual history."

The relevant construction of a non-standard probability function was given in an appendix, where Elga phrased the conclusion as such: "[T]he probabilities that these approximating functions ascribe to actual history span the entire range of infinitesimals [...]. So by picking an appropriate approximating function, we can get any such system to have any (infinitesimal) fit we'd like." In other words, Elga concluded that there are too many ways of assigning different infinitesimal probabilities to the same history and that there is no principled way to prefer one over the others.

Frederik Herzberg (2007) contrasted Elga's viewpoint, in which all hyperreal-valued functions that differ from a particular real-valued function by at most an infinitesimal (where the latter is defined) are to be treated on a par, with the praxis of NSA. As Herzberg pointed out, applications of NSA typically involve the construction of a *particular* non-standard object, usually some hyperfinite combinatorial object, leading to a particular internal probability measure. In order to appreciate how Herzberg's viewpoint differs from Elga's, it is helpful to consider an example:[46] Anderson (1976) has presented an internal representation of Brownian motion, which makes it possible to treat Brownian motion in terms of (infinite) combina-

---

46 I am grateful to Frederik Herzberg for this suggestion.

torics.[47] In order to be scientifically relevant, however, such an alternative description has to fulfil two criteria: (1) it has to approximate the standard probability function associated with the process (in this case, the Wiener measure)[48] and (2) it has to promote further research (as is indeed the case for Anderson's work; consider, for instance, Perkins ([1981])'s work on Brownian local time). Although many non-standard measures fulfil the first condition, the vast majority of them do not fulfil the second one.

Many worries and some open questions about infinitesimal probabilities arise due to the non-uniqueness and associated arbitrariness of hyperreal-valued probability measures (also discussed, *e.g.*, by Hofweber, [2014]).[49] When comparing the situation to that of real-valued probability functions that are CA, there is a trade-off between definiteness of the domain and definiteness of the range. In the case of an infinite sample space, CA functions have many non-measurable events in the powerset of that sample space. Which subsets of the sample space are measurable and which are not is to a certain extent arbitrary. If we settle for FA, we can extend the real-valued function to the entire powerset (by considering Banach limits, see for instance, Schurz & Leitgeb, [2008]), but then we introduce a lot of arbitrariness. Again in the case of an infinite sample space, NAP functions allow for the same kind of variation in their standard part as the FA functions do, and more given that also the infinitesimal part may vary (see for instance Kremer, [2014]). Given that it reappears in slightly different guises across different frameworks, we cannot set aside this arbitrariness as a flaw of one particular theory. Rather, it reminds us that the powerset of an infinite sample space contains a lot of uncharted territory.[50]

At least some of the worries related to arbitrariness are alleviated if we take into account the distinction between the ontology of infinitesimal probabilities and the deductive procedures they encourage: very similar modes of reasoning can be applied in related frameworks that suggest a different ontology (recall section [11]).[51]

---

47 See also Albeverio, Fenstad, Hoegh-Krøhn, and Lindstrøm ([1986], section 3.3).

48 Since internal probability functions differ from standard ones both in terms of domain and of range, this approximation can be thought of as a two-step procedure, the second of which involves the standard part function.

49 As mentioned in section [5], free ultrafilters are intangible objects. As a result, non-standard probability functions that rely on these filters are intangibles, too.

50 In particular, even if the sample space is just countably infinite, its powerset (which contains the events to which we want to assign probabilities) is uncountably large. Among the uncountably many sets that are neither finite nor co-finite, there is a wild variety (for instance, in terms of Turing degrees or other complexity measures) and it is here that we should take heed of Feferman's reservations about considering the totality of all arbitrary subsets of $\mathbb{N}$, $\mathcal{P}(\mathbb{N})$, as a well-defined notion; see, *e.g.*, Feferman ([1979], p. 166) and Feferman ([1999]). I am grateful to Paolo Mancosu for suggesting this connection.

51 Following a distinction introduced by Benacerraf ([1965]), a similar remark has been made by Katz ([2014], section 2.3) regarding interpretations of the work of Euler (and also that of Leibniz) in the context of standard or non-standard analysis.

More generally, various authors have argued that hyperreal numbers are not quite right for the task at hand (*e.g.*, that the infinitesimals are too small) (Easwaran, 2014; Pruss, 2014). Easwaran (2014, pp. 34–35) argued that "the structure of the hyperreals goes beyond the physical structure of credences" and that they "can't provide a faithful model of credences of the sort wanted by defenders of Regularity". On the other hand, Hofweber (2014) tried to defend infinitesimal chances and outlined some additional principles (non-locality, flexibility, and arbitrary additivity) that are required for a theory to capture our concept of chance. Also Benci et al. (2018) are optimistic that NAP theory can be defended against many of the previously raised objections.

14.3   *Cardinality considerations*

Hájek (2012b) argued that regularity is an untenable constraint on credences, even if we allow probability functions to take hyperreal values. He invited us to "imagine a spinner whose possible landing points are randomly selected from the $[0, 1)$ interval of the hyperreals", concluding that regularity fails if we apply the same interval of hyperreals as the range of a function that assigns probabilities to events associated with this hyperreal spinner. He envisaged

> "a kind of arms race: we scotched regularity for real-valued probability functions by canvassing sufficiently large domains: making them uncountable. The friends of regularity fought back, enriching their ranges: making them hyperreal-valued. I counter with a still larger domain: making its values hyperreal-valued"

and so on. Following up on Hájek's informal suggestion of an arms race, Alexander Pruss (2013) proves that for each set of probability values, possibly including hyperreal values, there exists a domain on which regularity fails.

However, as NAP theory illustrates, the defender of regularity need not participate in this race at all and Hájek considers this option, too: "Perhaps we could tailor the range of the probability function to the domain, for each particular application?" However, he worried "that in a Kolmogorov-style axiomatization the commitment to the range of *P* comes *first*". He continued by saying that "[i]t is not enough to say something unspecific, like 'some non-Archimedean closed ordered field ...' Among other things, we need to know what the additivity axiom is supposed to be." Of course, NAP theory does exactly this: by requiring ultra-additivity, for any sample space a range can be constructed that ensures regularity. However, one

cannot switch the quantifiers: in agreement with Pruss (2013), there is no universal range that can ensure regularity for all sample spaces.[52]

14.4 *Non-conglomerability*

Before we can address this worry, we first have to introduce the notion of conglomerability.

We will call a (hyper-)real-valued probability function P finitely, countably, or uncountably conglomerable if and only if for any finite, countable, or uncountable (resp.) partition $\{A_1, A_2, \ldots\}$ of the sample space (whose members are measurable according to P) and for any event A that is measurable according to P, the following conditional statement holds. If a and b are (hyper-)real numbers such that $\forall A_n \in \{A_1, A_2, \ldots\}, a \leq P(A|A_n) \leq b$, then $a \leq P(A) \leq b$.

In standard probability theory, both finite and countable conglomerability are guaranteed to hold. The proof of this relies crucially on the axiom of normalization and on the axiom of finite or countable additivity (resp.). Even in the standard approach, uncountable conglomerability does not hold in general.

Theories that lack normalization or countable additivity, are not guaranteed to be countably conglomerable. In particular, both de Finetti's proposal for FA probability theory and NAP theory are finitely but not countably conglomerable.[53]

This has been raised as an objection to theories that allow infinitesimal probabilities by Pruss (2012) and 2014. In recent work, Nicholas DiBella (2018, p. 1200) has shown that the failure of countable conglomerability already arises in qualitative probability theories that are non-Archimedean and that this carries over to any quantitative theory that is non-Archimedean (of which NAP theory is an example). Since it is such a general feature of the underlying probability ordering, he suggested that non-conglomerability is not suitable as a criticism of non-Archimedean theories.

---

52 Hájek (2012b) also stated that "[i]f we don't know exactly what the range is, we don't know what its notion of additivity will look like." Maybe prolonged exposure to real-valued measures, in which ultra-additivity is clearly unattainable, makes us overlook this very natural notion of additivity that does not depend on any further parameters?

53 The failure of countable conglomerability can be seen by considering a uniform distribution over the sample space $\mathbb{N} \times \mathbb{N}$ and two countable partitions: $A_i = \{(i, n)|n \in \mathbb{N}\}$ and $B_i = \{(n, i)|n \in \mathbb{N}\}$. For the demonstration in the case of FA probability, see **DeFinetti:1972**

15   EPILOGUE: ON THE VALUE OF METHODOLOGICAL PLURALISM

I would like to end this chapter with some remarks that may apply to formal epistemology (and related endeavours) more generally. Only by comparing different methodologies may one obtain some indication of their strengths and limitations and how they distort the results.

We tend not to notice what is always present. An atmosphere was present before our ancestors developed eyes and to this day the air between us remains invisible to us. By experimenting with other gas mixtures, we learn, not only about those new substances, but also about the air that surrounds us. We became aware of its weight, its oxygen content, and its capacity to carry our voice. And although we keep living in air for most of the time, for particular purposes, we may prefer other mixtures over air (*e.g.*, increasing the oxygen content to help someone breathe or decreasing the oxygen content to avoid oxidation).

Like air in our biosphere, the real numbers are equally pervasive in our current mathematical practice. It appears to me that we are subjected to methodological adaptation to an extent no less than we are to sensory adaptation. The study of infinitesimal probabilities involves a departure from the standard formalism of real-valued probability functions. By changing our methodological environment, we may start to notice certain assumptions in the usual approach. Dealing with a familiar problem in an unfamiliar way thus presents a unique opportunity: it allows us to distinguish elements that are essential to its solution from aspects that are merely artifacts due to the method that has been applied.

Investigating in a formal way a rich concept such as probability cannot be carried out within the bounds of any single formalization, but challenges us to combine perspectives from an equally rich selection of frameworks. In particular, I believe that methods involving hyperreal probability values, while detracting nothing from the merits of the monometric standard approach, have much to add to this polymetric selection.

1 APPENDIX: HISTORICAL SOURCES CONCERNING INFINITESIMAL PROBABILITIES (1870–1989)

This part does not contain an overarching story arc, but it can be used as an annotated bibliography or to look up specific details.

Despite its length, this appendix does not pretend to be exhaustive; some developments—especially the early ones—are merely sketched. The subdivision into decades is indicative rather than strict. Usually, the publication date is taken as the decisive factor for the chronology, except for Carnap's work from 1960: this work was only published in 1980, but it is included in an earlier section, for thematic reasons.

1.1 Before 1960: pre-Robinsonian era

1.1.1 The 1870s: The real numbers and the standard limit

The modern approach to standard analysis was developed by "the great triumvirate" (Boyer, 1949, p. 298): Georg Cantor, Richard Dedekind, and Karl Weierstrass. First, Cantor gave a construction of the real numbers via Cauchy sequences (recall section 8.5). Then, Dedekind gave an alternative construction of the real numbers via Dedekind cuts (which we will not discuss). Weierstrass introduced the modern epsilon-delta definition of the limit (which builds on earlier work by Bernard Bolzano in the 1810s and by Augustin-Louis Cauchy in the 1820s).

As an example, we consider the derivative as a limit of the quotient of differences and express this limit in terms of an epsilon-delta definition:

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}$$
$$= \lim_{\Delta x \to 0} \frac{y(x + \Delta x) - y(x)}{\Delta x},$$

where:

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = L \qquad \Leftrightarrow$$

$$\forall \epsilon > 0 \in \mathbb{R}, \ \exists \delta > 0 \in \mathbb{R} : \forall \Delta x \in \mathbb{R} \left(0 < |\Delta x| < \delta \Rightarrow |\frac{\Delta y}{\Delta x} - L| < \epsilon \right).$$

1.1.2 The 1880s: The Archimedean axiom

In the introduction, we encountered the criterion to decide whether a number system is Archimedean or non-Archimedean (condition (1)). In particular, hyperreal fields are non-Archimedean and those can be employed to represent infinitesimal probabilities. Here, we investigate the origins of this sense of the word 'Archimedean'.

Around 225 BC, Archimedes of Syracuse published two volumes known in English as "On the Sphere and Cylinder". At the beginning of the first book, Archimedes stated five assumptions. The fifth assumption is that,[54] starting from any quantity, one may exceed any larger quantity by adding the former quantity to itself sufficiently many times.[55] In a paper on ancient Greek geometry, Otto Stolz (1883) discussed this postulate, which he calls "*das Axiom des Archimedes*" for ease of reference. Although Stolz was well aware that Archimedes himself attributed an application of this axiom to earlier geometers, apparently he did not notice that the axiom also appeared in Euclid's Elements (Bair et al., 2013, p. 888). In his textbook on arithmetic, which was very influential according to Ehrlich (2006, p. 5), Stolz (1885) presented examples of *Grössensysteme* (systems of magnitudes) that fail to satisfy this Archimedean axiom, whereas systems that are continuous in the sense of Dedekind do satisfy it.

1.1.3 *The 1890s: Infinitesimal probabilities in a geometric context*

In 1891, Giulio Vivanti and Rodolfo Bettazzi discussed infinitesimal line segments in the context of probability (see Ehrlich, 2006). In these early discussions, infinitesimal probabilities are considered in the context of a geometric interpretation of probability. As such, this provides an interesting contrast to the more recent literature, in which infinitesimal probabilities are usually introduced in the context of subjective interpretations of probability (related to a criterion of open-mindedness).

Later on, in the 1910s, Federigo Enriques discussed the (impossibility of) infinitesimal probabilities on two occasions, again in a geometric context.[56]

1.1.4 *The 1900s: Measurability and non-measurability*

Building on Émile Borel's countably additive measure from the 1890s, Henri Lebesgue introduced his translation invariant and countably additive measure in 1902. In 1905, Giuseppe Vitali gave the first example of a non-Lebesgue measurable set. See for instance Skyrms (1983b) for some discussion.[57]

---

54 Heath (1897, p. 4) translated the assumption as follows: "Further, of unequal lines, unequal surfaces, and unequal solids, the greater exceeds the less by such a magnitude as, when added to itself, can be made to exceed any assigned magnitude among those which are comparable with [it and with] one another."

55 This formulation suggests a strong relation between Archimedean quantities and addition. Additivity also plays an important role in intuitions concerning infinitesimal quantities, including infinitesimal probabilities, even though these are non-Archimedean probabilities: recall the discussion on ultra-additivity (section 8.3 and appendix section 1.3.2).

56 Thanks to Philip Ehrlich for this addition. He is planning an article on the work of Enriques; meanwhile, Ehrlich (2006) contains the relevant references.

57 Skyrms (1983b) argued that the Peano-Jordan measure (which preceded the Borel measure) only employs ideas that were available in Plato's time, whereas Borel measure crucially

1.1.5 *The 1930s: Kolmogorov, Skolem and de Finetti*

KOLMOGOROV'S PROBABILITY MEASURES   Andrey Kolmogorov (1933) introduced probability as a one-place function with as the domain a field of sets over a given sample space and as the range the unit interval of the real numbers. In the first chapter of his book, he laid out an elementary theory of probability "in which we have to deal with only a finite number of events". The axioms for the elementary case stipulate non-negativity, normalization, and the addition theorem (now called 'finite additivity', FA). In the second chapter, dealing with the case of "an infinite number of random events", Kolmogorov introduced an additional axiom: the Axiom of Continuity. Together with the axioms and theorems for the finite case (in particular, FA), this leads to the generalized addition theorem, called 'σ-additivity' or 'countable additivity' (CA) in the case where the event space is a Borel field (or σ-algebra, in modern terminology). We reviewed his axiomatization in section 7.

SKOLEM'S NON-STANDARD MODELS OF PEANO ARITHMETIC   The second-order axioms for arithmetic are categoric: all models are isomorphic to the intended model $\langle \mathbb{N}, 0, +1 \rangle$, a triple consisting of the domain of discourse (infinite set of natural numbers), a constant element (zero), and the successor function (unary addition). Dedekind (1888) was the first to prove this. His 'rules' for arithmetic were turned into axioms by Giuseppe Peano (1889), giving rise to what we now call 'Peano Arithmetic' (PA).

The first-order axioms for arithmetic are non-categoric: there exist non-standard models $\langle {}^*\mathbb{N}, {}^*0, {}^*+{}^*1 \rangle$ that are not isomorphic to $\langle \mathbb{N}, 0, +, 1 \rangle$. Thoralf Skolem (1934) was the first who proved this.[58] With the Löwenheim-Skolem theorem, it can be proven that there exist models

---

relies on distinctions among infinite cardinalities only introduced by Cantor. Peano-Jordan measure is finitely additive, which follows from its definition, and it lacks the stronger property of countable additivity (CA). Borel measure is CA, but this has to be specified in the definition by hand. Skyrms observes that this approach was contested, for instance by Schoenflies in 1900, who objected that the matter of extending additivity into the infinite cannot be settled by positing it. Lebesgue measure is CA, too, and it is translation invariant, which is appealing to our intuitions.

58 See Stillwell (1977, section 3) and Kanovei, Katz, and Mormann (2013, section 3.2) for some comments on the direct construction given by Skolem (1934). In contrast to Skolem's result, the proof given in modern presentations usually relies on the Compactness property of first-order logic. First, consider a first-order language for arithmetic, $\mathcal{L}_{PA}$, which has a name for each natural number. Call PA the set of sentences in $\mathcal{L}_{PA}$ that are true about arithmetic. Then, add a new constant, c, to the language and consider PA', which is the union of the PA and $\{c > 0, c, > 1, c > 2, \ldots\}$. Since each finite subset of PA' has a model (in which c is a natural number that is larger than any of the other natural numbers that are named in the the finite subset), it follows from the Compactness of first-order logic that PA' has a model (which contains a copy of the natural numbers and in which c is an infinite hypernatural number).

of any cardinality. $^*\mathbb{N}$ contains finite numbers as well as infinite numbers. We now call $^*\mathbb{N}$ a set of hypernatural numbers.[59]

DE FINETTI ON NON-ARCHIMEDEAN PROBABILITY RANKINGS   In 1931, Bruno de Finetti addressed the relation between qualitative and quantitative probability. Qualitative probability deals with ordering or ranking events by a partial order relation, $\succeq$, interpreted as 'at least as likely as'. Quantitative probability deals with probability functions that assign numerical values—usually real numbers—to events.

On pp. 313–314, de Finetti (1931, section 13) presented four postulates for the probability ordering.[60] In particular, the second postulate states that every event that is merely possible (rather than impossible or certain) is strictly more likely than the impossible event and strictly less likely than the certain event. He considers the question whether such a ranking is compatible with the usual way of measuring probabilities by real numbers. De Finetti observed that such a probability ranking has a non-Archimedean structure, whereas real-valued probability functions are Archimedean. Related to this point, de Finetti (1931, p. 316) wrote:

> "However, it is anyway possible to satisfactorily measure probabilities by numbers, that is by making such a structure Archimedean by neglecting the infinitely small probabilities [...]"

Since this was written well before the development of NSA, we should be careful not to interpret "infinitely small probabilities" as the values of a hyperreal-valued probability function, which can subsequently be truncated by the standard part function. On the other hand, de Finetti was not merely referring to infinitesimal probabilities in an informal sense, either. In the continuation of the sentence quoted above, he stated, concerning infinitely small probabilities:

> "[...] that, when multiplied [...] by a number n, however large, they never tend to certainty, that is in other words, they are always less than the probability $1/n$ of one among n incompatible, identically probable events forming a complete class."

As a result, the partial order on the probability of events (which is just the order relation on the real numbers, $\geq$) does not coincide with the partial order on events ($\succeq$): taking A and B to be events, $P(A) \geq P(B)$ implies

---

59 For a discussion of the order-type of countable non-standard models of arithmetic, see *e.g.* Boolos, Burgess, and Jeffrey (2007, Ch. 25, p. 302–318) and McGee (2002). More advanced topics can be found in the book by Kossak and Schmerl (2006).

60 Thanks to Paul Pedersen for some pointers to de Finetti's early work on non-Archimedean probability rankings.

$A \succeq B$, but not vice versa, and $A \succeq B$ together with $B \succeq A$ implies $P(A) = P(B)$, but not vice versa. (Counterexamples to the inverse implications can be obtained by considering A to be the impossible event, $\varnothing$, and B a possible event with $P(B) = 0$.) The non-Archimedean partial ordering of events can be said to be more fine-grained than the Archimedean partial ordering of probabilities of those events, since the former leads to more equivalence classes (sets of events $\{B \mid B \succeq A \land A \succeq B\}$ for some event A) than the latter (with equivalence classes of events of the form $\{B \mid P(A) = P(B)\}$ for some event A).

In 1936, de Finetti reflected on the meaning of possible events (*i.e.*, events represented by non-empty sets) that have probability zero. He agrees with Borel and Lévy[61] that these are merely theoretical constructs: they do not represent events that are practically observable, but are merely defined as limiting cases thereof. They would require information from infinitely many experiments or an experiment involving an absolutely exact measurement, both of which exceed what is practically achievable.[62] In this context, and unlike the 1931 article, de Finetti did consider the option of infinitesimal probability values and even an infinite hierarchy thereof ("*chacune infiniment petite par rapport á la précédente*" p. 583). Ultimately, however, he advocated sticking to real numbers as probabilities and dropping the assumption of countable additivity (p. 584), which is a position he stood by throughout all of his later work (see appendix section 1.3.1).

1.1.6 *The 1950s: From weak to strict coherence*

In the context of Bayesianism and decision theory, infinitesimal probabilities have been discussed in relation to 'strict coherence'[63] and 'regularity'. This discussion started in the 1950s, with the Ph.D. dissertation of Abner Shimony followed by the article publication of Shimony (1955).

Earlier, both Frank P. Ramsey (1931) and de Finetti (1937) had combined a subjective interpretation of probability with an important rationality constraint, imposed on the set of an agent's degrees of belief: in order to be considered rational, a person's set of beliefs must meet the condition of 'coherence'. This condition can be regarded as a probabilistic extension of the consistency condition from classical logic. In particular, an agent's

---

61 See also footnote 78 for the relation to Cournot's principle.

62 This is the relevant quote in French de Finetti (1936, p. 577): "*Il n'y a pas de doute, ainsi que l'a remarqué M. Borel, et comme cela se trouve très clairement expliqué dans le traité de M. Lévy, que la notion d'événement possible et de probabilite nulle est purement théorique, car il s'agit en géneral d'événements définis comme des cas limites d'événements pratiquement observables, et leur vérification exigerait par conséquent une* infinité *d'expériences ou une expérience comportant une mensuration absolument* exacte."

63 In the early literature, there circulated other names for this criterion as well: 'strict fairness' (Kemeny) and [strong] 'rationality' (Lehman, Adams). See Carnap (1971a, p. 114) for a helpful overview of the terminology in the early literature.

degrees of belief are coherent just in case no Dutch book can be made against the agent: no finite combination of bets, of which the prizes are set in accordance with the agent's degrees of belief, should lead to a sure loss. De Finetti (1937) showed that an agent's degrees of belief are coherent (and thus that no Dutch Book can be made against him) just in case his degrees of belief are such that they respect the axioms for finitely additive probability functions.

SHIMONY'S STRICT COHERENCE   Shimony (1955) strengthened the earlier notion of coherence (now called 'weak coherence') to that of coherence 'in the strong sense' (now 'strict coherence'): no finite combination of bets, of which the prizes are set according to the agent's degrees of belief, should lead to a sure loss (as before) or a possible net loss without the possibility of a net profit (stronger condition). To obtain strong coherence, Shimony had to strengthen one of the probability axioms accordingly. The original axiom says that the degree of confirmation (or conditional credence) of some hypothesis h given a piece of evidence e is 1 if e entails h, whereas the stronger version reads: the degree of confirmation of h given e is 1 if and only if e entails h.

Initially, Shimony (1955) only defined (strict) coherence for finite sets of beliefs, but in a later section he did discuss "[t]he difficulty of extending the notion of coherence so as to apply to infinite sets" (p. 11). In this context, he wrote (p. 20):

> "An appropriate betting quotient would be an 'infinitesimal', which is neither 0 nor finite; but this is impossible because of the Archimedean property of the positive real numbers".

Shimony also remarked that strong coherence on infinite sets of belief cannot be used to justify CA (which he calls 'the Principle of Complete Additivity' on p. 18).

STRICT COHERENCE WITHOUT INFINITESIMALS   The work on strict coherence initiated by Shimony was soon picked up by others. Some of the ensuing publications were related to the notion of 'regularity'. In the context of finite sample spaces, Rudolf Carnap (1950, Ch. 5) had introduced regularity as the condition that a function should assign positive values to state descriptions that sum to unity. In particular, he applied this condition to credence functions (probability functions in the sense of rational degrees of belief) associated with a finite set of state descriptions (finite sample space).[64]

---

64 For infinite sample spaces, Carnap (1950) considers limits of unconditional and conditional probability functions; although those limit functions may assign zero to state descriptions, Carnap calls them 'regular', too. This usage should be contrasted with that in contemporary

Combining the earlier result of Shimony ([1955]) on the one hand and that of John G. Kemeny ([1981]) and R. Sherman Lehman ([1955]) on the other hand, we have that a probability function on a finite sample space is strictly coherent if and only if it is 'regular' (*cf.* Carnap, [1971b], p. 15).

Ernest W. Adams ([1959], [1962–63], [1964]) was interested in the case of infinite sample spaces: he focused on the issue of additivity. Walter Oberschelp ([1962–63]) wrote on a similar topic in German: he looked for a similar, but weaker constraint for the infinite case than Adams's.

So, none of these authors did follow up on Shimony's remark regarding infinitesimal probabilities. An important exception was Carnap: in 1960, he explicitly considered the option of non-real-valued degrees of belief that admit infinitesimal values. (Although this work was published posthumously, in 1980, we do discuss it already at this point.)

CARNAP'S QUEST FOR NON-ARCHIMEDEAN CREDENCES   Inspired by Shimony's work on strict coherence, Carnap ([1980]) considered a language with real-valued functions, $\mathcal{L}$, and a credence function with non-Archimedean range, $\mathcal{C}$. He wrote (p. 146):

> "we could regard these axioms as axioms of regularity for $\mathcal{L}$; and we would call $\mathcal{C}$ regular iff it fulfilled all these axioms. However, to carry out this program would be a task beset with great difficulties."

The first problem he considered is that of finding axioms for the binary relations *IS* (to be read as: 'is Infinitely Small compared to') and *SEq* (to be read as: 'is Smaller or Equal in size to'), both defined on the class of all subsets of the set of real numbers.[65] Further on, Carnap considered the problem of constructing a measure function π that is defined on all subsets of the set of real numbers. He stated (p. 154, italics in the original):

> "The values of π are not real numbers but numbers of a *non-Archimedean number system Ω to be constructed.*"

1.2   *The 1960s: Robinson's NSA and Bernstein & Wattenberg's non-standard probability*

The development of non-standard analysis by Abraham Robinson in the 1960s allowed for a formal and consistent treatment of infinitesimal num-

---

writings on infinitesimal credences, where regularity is (equivalent to) the condition that a probability function should assign strictly positive values to singleton events, even for infinite sample spaces.

65 Upon publication of these notes, Hoover ([1980]) remarked that one of the axioms Carnap had proposed for *SEq* was in contradiction with the others (axiom 3f on p. 147 amounted to countable additivity, which is incompatible with a non-Archimedean range); also one of the proposed axioms for *IS* was in contradiction with the others (axiom 7p on p. 148).

bers. Soon enough, this work was applied to measure theory in general and to probability theory in particular. Beyond this point, some technical notions from NSA appear: please consult sections 4 and 5 for the meaning of unfamiliar terms.

1.2.1 *Non-standard models of real closed fields and Robinson's NSA*

Robinson (1961, 1966) founded the field of NSA: he combined some earlier results from mathematical logic[66] in order to develop an alternative framework for differential and integral calculus based on infinitesimals and infinitely large numbers.

Robinson's hyperreal numbers are a special case of a real closed field (RCF). In general, a RCF is any field that has the same first-order properties as $\mathbb{R}$. The second-order axioms for the ordered field of real numbers are categoric: all models are isomorphic to the intended model $\langle \mathbb{R}, +, \times, \leq \rangle$, a quadruple consisting of the set of real numbers, the binary operations of addition and multiplication, and the order relation. Skolem's existence proof of non-standard models of arithmetic (section 1.1.5) can be applied to RCFs, too.[67] The axioms for RCFs (always in first-order logic) are non-categoric: there exist non-standard models $\langle {}^*\mathbb{R}, {}^*+, {}^*\times, {}^*\leq \rangle$ that are not isomorphic to $\langle \mathbb{R}, +, \times, \leq \rangle$.

Applying the Löwenheim-Skolem theorem, it can be proven that there exist models of any cardinality; in particular, there are countable models (*cf.* the 'paradox' of Skolem, 1923). In the context of hyperreal numbers, however, only uncountable models are considered. First of all, in this context the uncountable set of real numbers is assumed to be embedded in the non-standard model. Moreover, in the context of NSA also functions are transferred, which requires uncountably many symbols, thereby blocking the construction of a countable model.

The standard real numbers are Archimedean, *i.e.*, they contain no non-zero infinitesimals in the sense of condition (1):

$$\forall a \in \mathbb{R} \setminus \{0\}, \ \exists n \in \mathbb{N} : \frac{1}{n} < |a|. \tag{3}$$

In particular, $\langle \mathbb{R}, +, \times, \leq \rangle$ is the only complete Archimedean field.[68] In contrast, non-standard models do not have such a property: $\langle {}^*\mathbb{R}, {}^*+, {}^*\times, {}^* \leq \rangle$

---

66 See Robinson (1966, p. 48) for some references. In particular, Hewitt (1948) had constructed hyperreal fields using an ultrapower construction and Łoś (1955) had proven a transfer theorem for these fields.

67 Applying the idea of footnote 58 to RCF instead of PA, c will represent an infinite hyperreal number and its multiplicative inverse will represent an infinitesimal number.

68 Here, 'complete' can refer both to Cauchy or limit completeness (meaning that each Cauchy sequence of real numbers is guaranteed to converge in the real numbers) and to Dedekind or order completeness (meaning that each non-empty set of real number that has an upper bound is guaranteed to have a least upper bound), because Cauchy completeness together with the Archimedean property implies Dedekind completeness.

is a non-Archimedean ordered field and it is not complete. Saying that $^*\mathbb{R}$ is non-Archimedean means that it does contain non-zero infinitesimals in the sense of condition (1):

$$\exists a \in {}^*\mathbb{R} \setminus \{0\}, \ \forall n \in \mathbb{N} : \ \frac{1}{n} \leq |a|.$$

In other words: $^*\mathbb{R}$ contains infinitesimals. As a consequence, for any such a hyperreal infinitesimal a it holds that

$$\forall n \in \mathbb{N} : \ \sum_{i=1}^{n} |a| < 1.$$

$^*\mathbb{R}$ contains finite, infinite and infinitesimal numbers; we call $^*\mathbb{R}$ a set of hyperreal numbers.

1.2.2   *Bernstein & Wattenberg's non-standard probability function*

The infinitesimal numbers contained in the unit interval of a non-standard model of a RCF can be used to represent infinitesimal probabilities. Allen R. Bernstein and Frank Wattenberg (1969) were the first to apply Robinson's NSA in a probabilistic setting and thus to describe infinitesimal probabilities in a mathematically rigorous framework. On p. 171, they stated the following goal:

> "Suppose that a dart is thrown, using the unit interval as a target; then what is the probability of hitting a point?"

They followed up this question with an informal answer:

> "Clearly this probability cannot be a positive real number, yet to say that it is zero violates the intuitive feeling that, after all, there is some chance of hitting the point."

In their paper, Bernstein and Wattenberg formalized this intuitive answer using positive infinitesimals from Robinson's NSA.[69] Their measure is based on a hyperfinite counting measure of a hyperfinite subset of the hyperextension of the sample space.[70] The non-standard result for any Lebesgue-measurable set is infinitely close to its Lebesgue measure:[71]

---

69 Observe that, in order to assign non-zero infinitesimals to point events, they have to depart from the usual application of NSA. Moreover, the function that they obtain is an external object, which means (roughly) that it does not have a counterpart within standard analysis (*cf.* section 4). On the other hand, it is possible to take the standard part of the function's output, which yields the unique real value that is closest to the hyperreal value.

70 Recall section 4 for the meaning of 'hyperfinite' and 'hyperextension'.

71 One may object against the use of measure theory to represent probability, since measures are motivated by a desire to idealize the notions of physical length, area, and volume, and not probability *per se*. Hence, the usual reservations of representing probability by measure functions, be they standard or non-standard, may apply here.

> "In particular, nonempty sets of Lebesgue measure zero will
> have positive infinitesimal measure."

They stated that:

> "Thus, for example, it is now possible to say that 'the probabil-
> ity of hitting a rational number in the interval $[0, \frac{1}{4})$ is exactly
> half that of hitting a rational number in the interval $[0, \frac{1}{2})$, de-
> spite the fact that both sets in question have Lebesgue measure
> zero."

Of course, the former probability being half that of the latter also applies
if both probabilities are zero, rather than infinitesimals.[72] This observation
is only relevant if an additional assumption is made, for instance that the
probabilities are non-zero or that the former should be smaller than the
latter.

1.3 *After 1969: Further developments and philosophical discussions*

1.3.1 *The 1970s: Further mathematical developments*

PARIKH & PARNES'S CONDITIONAL PROBABILITY FUNCTIONS   Start-
ing from a standard absolute probability function, the ratio formula does
not always suffice to define a conditional probability function. This may fail
in two ways: the probabilities may be undefined (non-measurable events)
or the conditioning event may have probability zero. The non-standard
absolute probability function obtained by Bernstein and Wattenberg (1969)
does allow us to define a non-standard absolute probability function for all
pairs of subsets of the real numbers by the usual ratio formula, provided
that the conditioning event is non-empty. By taking the standard part, we
obtain a real-valued function defined for all pairs of subsets of the real
numbers (as long as the conditioning event is non-empty). However, Rohit
Parikh and Milton Parnes (1974) remarked that the conditional probability
function so obtained does not necessarily exhibit translation invariance in
the following sense:

$$\forall A, B \subseteq \mathbb{R} \text{ such that } B \neq \varnothing, \ \forall x \in \mathbb{R}, \ P(A + x, B + x) = P(A, B),$$

where $A + x$ is the set obtained by adding x to all elements of A and P is the
standard conditional probability function obtained by applying the ratio
formula to a non-standard absolute probability function as constructed by
Bernstein and Wattenberg (1969) and then taking the standard part.

Parikh and Parnes (1974) did not consider non-standard conditional
probability functions. Instead, they merely used NSA as a means of ob-
taining standard functions. Using techniques from NSA (in particular,

---

72 This observation is due to Alan Hájek, whose copy of the article I was allowed to copy.

hyperfinite sets), Parikh and Parnes ([1974](#)) construct standard conditional probability functions, each fulfilling a number of algebraic conditions that correspond with our intuitions. Apart from a condition that entails the above criterion of translation invariance, they also obtain: (i) $P(B, B) = 1$ for all B, (ii) if $B = [0, 1]_{\mathbb{Q}}$ (the unit interval of $\mathbb{Q}$ with endpoints included) and $0 \leq a < b \leq 1$, then $P([a, b], B) = b - a$, and (iii) $P(A, B) = 0$ whenever A is finite and B is not.[73] It requires a bit more effort (choosing a suitable ideal on $\mathbb{R}$, *cf.* section [5](#)) to obtain a function P such that the following stronger version of (iii) also holds: $P(A, B) = 0$ whenever A is countable and B is not. After proving the relevant existence theorems, they show that the cardinality of the set of standard conditional probability functions satisfying the various combinations of properties is 2^c, with c the cardinality of the continuum.

HENSON'S REPRESENTATION THEOREM   Meanwhile, C. Ward Henson ([1972](#)) had shown that for every standard, finitely additive probability measure that assigns zero to finite sets there exists a non-standard representation. Once again, the proof relies on a hyperfinite counting measure on a hyperfinite subset of the hyperextension of the sample space of the standard function. He also considered the special case in which the standard measure is countably additive. As is typical in the context of NSA, Henson showed how to apply his result in order to obtain a shorter proof of a standard result (in section 2 of his paper).[74]

LOEB MEASURE   Seminal contributions to non-standard measure theory were obtained by Peter A. Loeb ([1975](#)). A good overview of this topic (up to the early 1980s) can be found in Cutland ([1983](#)). Loeb measures require more advanced technical knowledge than any of the other approaches covered in this chapter. In particular, they require non-standard models with a saturation beyond countable saturation.[75]

DE FINETTI'S RESPONSE   As indicated in section [1.1.5](#), de Finetti wrote on the topic of non-Archimedean probability rankings well before the development of NSA. Although he lived long enough and was aware of the development of NSA, he never showed much interest in applying it

---

73 Observe that these conditional probability functions violate regularity, but this should not be surprising since they are real-valued.

74 See also Hofweber and Schindler ([2016](#)) for "a new and completely elementary proof of this fact".

75 In the construction of $^*\mathbb{R}$, we used a free ultrafilter on $\mathbb{N}$ (see part [3](#)). This is sufficient to obtain a model with countable saturation. It is possible to fix a free ultrafilter on an infinite index set of higher cardinality. In particular, by choosing 'good' ultrafilters, it is possible to arrive at the desired level of saturation in a single step (Keisler, [2010](#), section 10). See (Hurd & Loeb, [1985](#), p. 104–108) for more on saturation.

to his own work on probability. This can be seen by inspecting his work from the 1970s.

In the second volume of his 1974 book, de Finetti famously returned to the discussion of possible events with zero probability—a topic already on his mind (and in his publications) in the 1930s. In particular, he wondered whether it is "possible to compare the zero probabilities of possible events" and whether "a union of events with zero probabilities [can] have a positive probability" (de Finetti, 1974, Vol. II, p. 117). On p. 118, he remarks that the latter question can be rephrased in terms of additivity and he distinguishes three cases: finite additivity, countable additivity, and perfect additivity "if the additivity always holds".[76] On p. 119, he discusses weak and strong coherence; of the latter he writes "This means that 'zero probability' is equivalent to 'impossibility'." However, he warns us that besides "these serious authors" who have written on this topic, there are others "who refer to zero probability as impossibility, either to simplify matters in elementary treatments, or because of confusion, or because of metaphysical prejudices". So, according to de Finetti, if we are careful enough not to interpret zero probability as impossibility, we do not need infinitesimal probabilities at this point—in fact, he does not mention them on these pages.

Elsewhere in his book, however, de Finetti does consider non-zero infinitesimal probabilities in relation to additivity. De Finetti (1974, p. 347) writes:

> "Let us just mention that the consideration of probability as a non-Archimedean quantity would permit us to say, if we wished, that 'zero probabilities' are in fact 'infinitely small' (actual infinitesimals), and only that of the impossible event is zero. Nothing is really altered by this change in terminology, but it might sometimes be useful as a way of overcoming preconceived ideas. It has been said that to assume that $0 + 0 + 0 + ... + 0 + ... = 1$ is absurd, whereas, if at all, this would be true if 'actual infinitesimal' were substituted in place of zero. There is nothing to prevent one from expressing things in this way, [...]"

This seems to be a welcoming invitation to adopt techniques from NSA in order to deal with infinitesimal probabilities and associated puzzles concerning their additivity. However, de Finetti continues his sentence less enthusiastically: "[...] apart from the fact that it is a useless complication of language, and leads one to puzzle over '*les infiniment petits*'."[77]

---

76 *Cf.* ultra-additivity in the terminology of Skyrms (1983b): see section 1.3.2.
77 The French expression '*les infiniment petits*' was in use since the development and popularization of the calculus; consider, for instance, the title of de l'Hôpital's 1696 book

Moreover, in 1979 (as transcribed in de Finetti, 2008, Ch. 12, p. 122), a graduate student asked de Finetti about his thoughts concerning NSA. The student (referred to as 'Alpha' in the transcript) asked: "do you consider it plausible that this hierarchy of zero probabilities could be replaced by a hierarchy of actual infinitesimals in the sense of non-standard analysis?" To which de Finetti responded:

> "I only attended a few talks on non-standard analysis and I have to say that I am not sure about its usefulness. On the face of it, it does not persuade me, but I think I have not delved enough into this topic in order to be able [to] give [a] well thought-out judgment. [...] I made those speculations on infinitely small probabilities to see the extent to which the idea of a comparison between zero probabilities is plausible. However, I did not attach much importance to it and I am not sure whether one needs sophisticated theories, such as non-standard analysis, for that goal."

1.3.2 *The 1980s: Skyrms, Lewis, and Nelson*

SKYRMS ON INFINITESIMAL CHANCES   Skyrms (1980) argued that propensity (for instance, the bias parameter in a binomial distribution) does not equal the limiting relative frequency (for instance, of an infinite Bernoulli process). He did so by appealing to infinitesimal probabilities (pp. 30–31):

> "If we extend our language so that we can talk in it about limiting relative frequencies in an infinite sequence of trials and make a few assumptions about limiting probabilities, we can state what appears to be a more powerful version of the law of large numbers: the probability that, in a given sequence of independent and identically distributed trials, the limiting relative frequency will either fail to exist or diverge by some positive real number from the probability of the outcome is infinitesimal. Then, if our coin is flipped an infinite number of times, the probability that the limiting relative frequency fails to be one-half is infinitesimal."

"*Analyse des Infiniment Petits pour l'Intelligence des Lignes Courbes*". The use of infinitesimals in calculus was discredited in subsequent years (in favour of epsilon-delta constructions developed in the work of Weierstrass, *cf.* section 1.1.1). Although NSA did much to re-instate them, this process of rehabilitation of infinitesimals was neither immediate nor uniform (and remains incomplete, even today). So, it seems that de Finetti held on to the post-Weierstrassian and pre-Robinsonian viewpoint of infinitesimals as a suspect concept, to be avoided when possible.

He then went on to show that this viewpoint is not compatible with the idea "that infinitesimal propensity implies impossibility". The stance that Skyrms is refuting here is sometimes called the 'principle of Cournot'.[78]

> "[T]he assumptions that get the striking version of the strong law of large numbers give us infinitesimal probability not only for the outcome sequence All Heads, but for each other definite sequence of outcomes as well. But the coin has to do something! There is nothing more probable than that something improbable will happen, but it is impossible that something impossible should happen. Small probability, even infinitesimally small probability, does not mean impossibility. Then even if, for each process, the propensity for a divergence between propensity and relative frequency is infinitesimal, it hardly follows that the propensity for a divergence for some process, somewhere in the world, is infinitesimal. But this is just what those who wish to turn the law of large numbers into a philosophical analysis of propensity must assume."

Here, Skyrms used infinitesimal probabilities to illustrate the qualitative difference between possible events and the impossible event. In particular, in cases of equiprobability it may be certain that a highly unlikely event will occur. This seems to be diametrically opposed to Cournot's principle and similar ideas such as the Lockean thesis (but see also section 13).

LEWIS ON INFINITESIMAL CHANCES AND CREDENCES   David Lewis (1980) introduced his 'Principal Principle' as a way to connect subjective credences to objective chances. In this context, he discussed how infinitesimal chances lead to the introduction of infinitesimal credences (p. 269):

> "The Principal Principle may be applied as follows: you are sure that some spinner is fair, hence that it has infinitesimal chance of coming to rest at any particular point; therefore (if your total evidence is admissible) you should believe only to an infinitesimal degree that it will come to rest at any particular point."

On pp. 267–268, Lewis (1980) discussed infinitesimal credences in the context of regularity (*cf.* section 1.1.6) and a "condition of reasonableness":

---

78 The principle of Cournot is named after Augustin Cournot, because of his writings on the notion of 'physical impossibility' (of events corresponding to infinitesimal probabilities in a geometric context). The roots of the concept go back to that of 'moral certainty' (practical certainty) in the work of Jacob Bernoulli. Similar ideas also arose in the work of Paul Lévy and Émile Borel (which inspired de Finetti's speculations on hierarchies of infinitesimals). The name for the principle was introduced by Maurice Fréchet. For more details, see, *e.g.*, Shafer (2008).

"I should like to assume that it makes sense to conditionalize
on any but the empty proposition. Therefore I require that [any
reasonable initial credence function] *C* is *regular*: *C*(*B*) is zero,
and *C*(*A*/*B*) is undefined, only if *B* is the empty proposition,
true at no worlds. You may protest that there are too many
alternative possible worlds to permit regularity. But that is so
only if we suppose, as I do not, that the values of the func-
tion *C* are restricted to the standard reals. Many propositions
must have infinitesimal *C*-values, and *C*(*A* | *B*) often will be
defined as a quotient of infinitesimals, each infinitely close but
not equal to zero. (See Bernstein and Wattenberg [1969].) The
assumption that *C* is regular will prove convenient, but it is
not justified only as a convenience. Also it is required as a
condition of reasonableness: one who started out with an irreg-
ular credence function (and who then learned from experience
by conditionalizing) would stubbornly refuse to believe some
propositions no matter what the evidence in their favor."

SKYRMS ON REGULARITY AND ULTRA-ADDITIVITY   Skyrms ([1983b])
gave an intriguing analysis of the Zenonian intuition of regularity. His
text focused on length measurement, but the argument carries over to
probability measures; hence, we present it in some detail. 'Zeno's paradox
of measure' is a scholarly reconstruction of an argument against plurality
emerging from Zeno's four paradoxes of motion. The conclusion of this
argument is that something of non-zero, finite length cannot be composed
of infinitely many parts. The Zenonian argument starts by assuming
the opposite: if the whole is composed of infinitely many parts, then
either those parts all have no magnitude or they all have a non-zero
magnitude, but then the whole would either have no magnitude or an
infinite magnitude, respectively, both of which are in contradiction with the
whole having a non-zero, finite length. Skyrms argues that this argument
crucially relies on some implicit assumptions: that the parts all have equal
size (invariance), that they are not infinitesimal (Archimedean axiom), and
that we can make sense of an infinite sum of the individual magnitudes
(ultra-additivity). As such, 'Zeno's paradox of measure' has a very similar
structure to the proof that shows that there is no real-valued, countably
additive probability function that assigns equal probabilities to single
tickets in a lottery on the natural numbers (*cf.* section 8.3): it shows that
either assigning zero probability or non-zero probability to individual
tickets both fail to yield a normalizable measure, because either the sum
over all tickets is zero or it diverges. Analogous assumptions are in place
in both arguments: an invariant partition such that the parts have equal
magnitudes versus equiprobability; no infinitesimal magnitudes versus

real-valued probability; and a way to make sense of infinite sums of magnitudes versus countable additivity.

Skyrms names the additivity assumption in the Zenonian argument the principle of ultra-additivity, which he specifies as follows (p. 227):

> "the principle that the magnitude of the whole is the sum of the magnitudes of its parts continues to hold good when we have a partition of the whole into an infinite number of parts."

This way of phrasing it—as a property known for finite quantities that is assumed to hold for infinite quantities, too—resembles Leibniz's '*souverain principe*' (see Katz & Sherry, 2012, section 4.3), which in turn can be formalized by the Transfer principle of NSA (as was explained in section 4). In this light, it is curious to observe that the term for the Zenonian principle chosen by Skyrms, ultra-additivity, resonates well within the context of NSA, which is replete with ultrafilters. (This resonance may be curious, but it need not be coincidental—given Skyrms's familiarity with NSA.)

Skyrms also argues that the step in the Zenonian argument that implicitly assumes the principle of ultra-additivity was not contested by the school of Plato, the school of Aristotle, or the atomists. So, it appears that the principle of ultra-additivity was—possibly without reflection—widely accepted, which suggests that it represents a deeply anchored intuition about magnitudes: if finite magnitudes are to be infinitely divisible (which of course the Zenonian argument tries to refute), then it is hard to imagine for the magnitudes of the parts in the partition *not* to sum to the magnitude of the whole. Skyrms writes (p. 235): "It is ironic that it is just here that the standard modern theory of measure finds the fallacy."

In the context of measure theory, and thus of standard probability, the principle of ultra-additivity is formalized—and thereby restricted to countable collections—in terms of CA. However, as the failure of the existence of a countably additive fair probability measure on the natural numbers demonstrates, it does not do justice to the underlying intuition of universal summability.

LEWIS ON INFINITESIMAL CHANCES   In a postscript to 'Causations' (an article that appeared in 1973) and in a passage that appears between brackets, Lewis (1986b, pp. 175–176) discussed infinitesimal chances and presented real-valued probabilities as a rounding off of the true hyperreal chances (with original italics):[79]

> "They say that things with no chance at all of occurring, that
> is with probability zero, do nevertheless happen; for instance

---

[79] Hájek (2012a) cites this passage and calls Lewis's work on this topic "[t]he most important philosophical defence of regularity" of which he is aware (p. 414).

> when a fair spinner stops at one angle instead of another, yet any precise angle has probability zero. I think these people are making a rounding error: they fail to distinguish zero chance from infinitesimal chance. Zero chance is *no* chance, and nothing with zero chance ever happens. The spinner's chance of stopping exactly where it did was not zero; it was infinitesimal, and infinitesimal chance is still *some* chance."

Although they are not mentioned here, Lewis's wording is very reminiscent of Bernstein and Wattenberg (1969), who wrote "there is still some chance of hitting the point". Also observe that according to the definition that we gave in the introduction, zero is an infinitesimal. Hence, what Lewis is arguing for must be called 'non-zero infinitesimals' in our terminology.

NELSON'S RADICALLY ELEMENTARY PROBABILITY THEORY   Previously, Edward Nelson (1977) had provided the first axiomatic approach to NSA, which he called 'Internal Set Theory' (IST),[80] but he also provided an important alternative approach to infinitesimal probabilities. Nelson (1987) developed a 'Radically elementary probability theory', which relies on internal probability functions: these functions can be obtained by applying the Transfer principle (recall section 4) to sequences of standard Kolmogorovian probability functions on finite domains. Internal probability functions do not assign probability values to any infinite standard sets, but only to hyperfinite sets. The resulting additivity property is hyperfinite additivity. Nelson's probability functions are regular and they admit infinitesimal values. Unlike much previous work on non-standard probability functions, this approach does not aim at providing a real-valued probability measure (by the standard part function, *cf.* section 4). Precisely by leaving out this step, this framework has the benefit of making probability theory on infinite sample spaces equally simple and straightforward as the corresponding theory on finite sample spaces.

REFERENCES

Adams, E. W. (1962–63). On rational betting systems. *Archiv für mathematische Logik und Grundlagenforschung*, *6*, 7–29. Part 1 of 2.

---

80 According to Luxemburg (2007, p. xi):

> "[F]rom the beginning Robinson was very interested in the formulation of an axiom system catching his non-standard methodology. Unfortunately he did not live to see the solution of his problem by E. Nelson presented in the 1977 paper entitled 'Internal Set Theory'."

Adams, E. W. (1959). Two aspects of the theory of rational betting odds. *Technical Report, Berkeley (Univ. of Calif.) 1*, 9. Rotaprintvervielfälti-gung.

Adams, E. W. (1964). On rational betting systems. *Archiv für mathematische Logik und Grundlagenforschung, 6*, 112–128. Part 2 of 2.

Albeverio, S., Fenstad, J. E., Hoegh-Krøhn, R., & Lindstrøm, T. (1986). *Non-standard methods in stochastic analysis and mathematical physics*. Pure and Applied Mathematics. Orlando, FL: Academic Press.

Alexander, A. (2014). *Infinitesimal: How a dangerous mathematical theory shaped the modern world*. London, UK: Oneworld.

Anderson, R. M. (1976). A non-standard representation for Brownian motion and Itô integration. *Israel Journal of Mathematics, 25*, 15–46.

Bair, J., Błaszczyk, P., Ely, R., Henry, V., Kanovei, V., Katz, K. U., . . . Shnider, S. (2013). Is mathematical history written by the victors? *Notices of the American Mathematical Society, 60*, 886–904.

Barrett, M. (2010). The possibility of infinitesimal chances. In E. Eells & J. H. Fetzer (Eds.), *The place of probability in science* (pp. 65–79). Boston Studies in the Philosophy of Science. Springer.

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

Bartha, P. & Hitchcock, C. (1999). The shooting-room paradox and condi-tionalizing on measurably challenged sets. *Synthese, 118*, 403–437.

Bartha, P. & Johns, R. (2001). Probability and symmetry. *Philosophy of Science, 68*, S109–S122.

Benacerraf, P. (1965). What numbers could not be. *Philosophical Review, 74*, 47–73.

Benci, V. & Di Nasso, M. (2003). Numerosities of labelled sets: A new way of counting. *Advances in Mathematics, 173*, 50–67.

Benci, V., Di Nasso, M., & Forti, M. (2006). The eightfold path to nonstan-dard analysis. In N. J. Cutland, M. Di Nasso, & D. A. Ross (Eds.), *Nonstandard methods and applications in mathematics* (Vol. 25, pp. 3–44). Lecture Notes in Logic. Wellesley, MA: Association for Symbolic Logic, AK Peters.

Benci, V., Horsten, L., & Wenmackers, S. (2013). Non-Archimedean proba-bility. *Milan Journal of Mathematics, 81*, 121–151.

Benci, V., Horsten, L., & Wenmackers, S. (2018). Infinitesimal probabilities. *British Journal for the Philosophy of Science, 69*, 509–552.

Berkeley, G. (1734). *The analyst, a discourse addressed to an infidel mathemati-cian*. London, England: Strand.

Bernstein, A. R. & Wattenberg, F. (1969). Nonstandard measure theory. In W. A. J. Luxemburg (Ed.), *Applications of model theory to algebra, analysis and probability* (pp. 171–185). New York, NY: Holt, Rinehard and Winston.

Błaszczyk, P., Katz, M. G., & Sherry, D. (2013). Ten misconceptions from the history of analysis and their debunking. *Foundations of Science*, *18*, 43–74.

Boolos, G. S., Burgess, J. P., & Jeffrey, R. C. (2007). *Computability and logic*. 5th ed. Cambridge, UK: Cambridge University Press.

Boyer, C. (1949). *The concepts of the calculus*. Hafner Publishing Company.

Brickhill, H. & Horsten, L. (2018). *Triangulating non-Archimedean probability*. Forthcoming in *Review of Symbolic Logic*; URL: https://arxiv.org/abs/1608.02850.

Carlyle, T. (1845). *Oliver Cromwell's letters and speeches: With elucidations*. New York, NY: Wiley and Putnam.

Carnap, R. (1950). *Logical foundations of probability*. Chicago, IL: University of Chicago Press.

Carnap, R. (1971a). A basic system of inductive logic, part I. In R. Carnap & R. C. Jeffrey (Eds.), *Studies in inductive logic and probability* (Vol. 1). Chicago, IL: University of Chicago Press.

Carnap, R. (1971b). Inductive logic and rational decisions. In R. Carnap & R. C. Jeffrey (Eds.), *Studies in inductive logic and probability* (Vol. 1). Chicago, IL: University of Chicago Press.

Carnap, R. (1980). The problem of a more general concept of regularity. In R. C. Jeffrey (Ed.), *Studies in inductive logic and probability* (Vol. 2, pp. 145–155). Written in 1960. Berkeley, CA: University of California Press.

Chen, E. & Rubio, D. (2018). *Surreal decisions*. Forthcoming in *Philosophy and Phenomenological Research*; doi:10.1111/phpr.12510.

Cutland, N. (1983). Nonstandard measure theory and its applications. *Bulletin of the London Mathematical Society*, *15*, 529–589.

de Finetti, B. (1931). Sul significato soggettivo della probabilità. *Fundamenta Mathematica*, *18*, 298–329. Translated in English as "On the subjective meaning of probability" in: P. Monari and D. Cocchi (eds.), "Probabilità e Induzione; Induction and Probability" (1993) Clueb, Bologna; pp. 291–321.

de Finetti, B. (1936). Les probabilités nulles. *Bulletin de Sciences Mathématiques*, *60*, 275–288.

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

de Finetti, B. (1974). *Theory of probability*. Translated by: A. Machí and A. Smith. London, UK: Wiley.

de Finetti, B. (2008) In A. Mura (Ed.), *Philosophical lectures on probability* (Vol. 340). Synthese Library. Introductory Essay by Maria Carla Galavotti; translated by: Hykel Hosni. London, UK: Springer.

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

DiBella, N. (2018). The qualitative paradox of non-conglomerability. *Synthese*, *195*, 1181–1210.

Easwaran, K. (2014). Regularity and hyperreal credences. *Philosophical Review*, *123*, 1–41.

Ehrlich, P. (2006). The rise of non-Archimedean mathematics and the roots of a misconception i: The emergence of non-Archimedean systems of magnitudes. *Archive for History of Exact Sciences*, *60*, 1–121.

Elga, A. (2004). Infinitesimal chances and the laws of nature. *Australasian Journal of Philosophy*, *82*, 67–76.

Feferman, S. (1979). Constructive theories of functions and classes. *Logic Colloquium*, *78*, 159–224.

Feferman, S. (1999). Does mathematics need new axioms? *The American Mathematical Monthly*, *106*, 99–111.

Foley, R. (2009). Beliefs, degrees of belief, and the Lockean thesis. In F. Huber & C. Schmidt-Petri (Eds.), *Degrees of belief* (Vol. 342, pp. 37–47). Synthese Library. Dordrecht, The Netherlands: Springer.

Gaifman, H. (1986). Towards a unified concept of probability. In R. B. Marcus, G. J. W. Dorn, & P. Weingartner (Eds.), *Logic, methodology and philosophy of science vii* (Vol. 114, pp. 319–350). Studies in Logic and the Foundations of Mathematics. Amsterdam, The Netherlands: Elsevier.

Goldblatt, R. (1998). *Lectures on the hyperreals; an introduction to nonstandard analysis*. Graduate Texts in Mathematics. New York, NY: Springer.

Hacking, I. (1975). *The emergence of probability: A philosophical study of early ideas about probability, induction and statistical inference*. Cambridge, UK: Cambridge University Press.

Hájek, A. (2003). Waging war on Pascal's wager. *The Philosophical Review*, *112*, 27–56.

Hájek, A. (2012a). Is strict coherence coherent? *Dialectica*, *66*, 411–424.

Hájek, A. (2012b). *Staying regular?* Unpublished manuscript. Retrieved from `http://philrsss.anu.edu.au/sites/default/files/Staying%20Regular.December%2028.2012.pdf`

Halpern, J. Y. (2010). Lexicographic probability, conditional probability, and nonstandard probability. *Games and Economic Behavior*, *68*, 155–179.

Heath, T. L. (Ed.). (1897). *The works of Archimedes; edited in modern notation with introductory chapters*. Cambridge, UK: Cambridge University Press.

Henson, C. W. (1972). On the nonstandard representation of measures. *Transactions of the American Mathematical Society*, *172*, 437–446.

Herzberg, F. (2007). Internal laws of probability, generalized likelihoods and Lewis's infinitesimal chances—A response to Adam Elga. *British Journal for the Philosophy of Science*, *58*, 25–43.

Herzberg, F. (2010). The consistency of probabilistic regresses. a reply to Jeanne Peijnenburg and David Atkinson. *Studia Logica*, *94*, 331–345.

Hewitt, E. (1948). Rings of real-valued continuous functions I. *Transactions of the American Mathematical Society*, *64*, 54–99.

Hilbert, D. (1900). Mathematische Probleme. *Göttinger Nachrichten*, 253–297. Translated as "Mathematical Problems", *Bulletin of the American Mathematical Society*, **8**, no. 10 (1902), pp. 437–479.

Hofweber, T. (2014). Infinitesimal chances. *Philosophers' Imprint*, *14*, 1–14.

Hofweber, T. & Schindler, R. (2016). Hyperreal-valued probability measures approximating a real-valued measure. *Notre Dame Journal of Formal Logic*, *57*, 369–374.

Hoover, D. (1980). A note on regularity. In R. C. Jeffrey (Ed.), *Studies in inductive logic and probability* (Vol. 2, pp. 295–297). Berkeley, CA: University of California Press.

Howson, C. (2017). Regularity and infinitely tossed coins. *European Journal for Philosophy of Science*, *7*, 97–102.

Hrbáček, K. (2007). Stratified analysis? In I. van den Berg & N. Neves (Eds.), *The strength of nonstandard analysis* (pp. 47–63). Vienna, Austria: Springer.

Hurd, A. E. & Loeb, P. A. (1985). *An introduction to nonstandard real analysis*. Pure and Applied Mathematics. Orlando, FL: Academic Press.

Kanovei, V., Katz, M. G., & Mormann, T. (2013). Tools, objects, and chimeras: Connes on the role of hyperreals in mathematics. *Foundations of Science*, *18*, 259–296.

Katz, M. G. (2014). Leibniz's infinitesimals: Their fictionality, their modern implementations, and their foes from Berkeley to Russell to beyond. *Erkenntnis*, *78*, 571–625.

Katz, M. G. & Sherry, D. (2013). Leibniz's infinitesimals: Their fictionality, their modern implementations, and their foes from Berkeley to Russell to beyond. *Erkenntnis*, *78*, 571–625.

Katz, M. G. & Sherry, D. M. (2012). Leibniz's laws of continuity and homogeneity. *Notices of the American Mathematical Society*, *59*, 1550–1558.

Keisler, H. J. (2010). The ultraproduct construction. In V. Bergelson, A. Blass, M. Di Nasso, & R. Jin (Eds.), *Ultrafilters across mathematics* (Vol. 530, pp. 163–179). Contemporary Mathematics. American Mathematical Society.

Kelly, K. T. (1996). *The logic of reliable inquiry*. Oxford, UK: Oxford University Press.

Kemeny, J. G. (1981). Fair bets and inductive probabilities. *The Journal of Symbolic Logic*, *20*, 263–273.

Kerkvliet, T. & Meester, R. (2016). Uniquely determined uniform probability on the natural numbers. *Journal of Theoretical Probability*, *29*, 797–825.

Kolmogorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitrechnung*. Ergebnisse der Mathematik. Translated by N. Morrison, *Foundations of probability.* Chelsea Publishing Company, 1956 (2nd ed.) Berlin, Germany: Springer.

Kolmogorov, A. N. (1948). Algèbres de Boole métriques complètes. *VI Zjazd Matematyków Polskich*, 21–30. Translated by R. C. Jeffrey as "Complete metric Boolean algebras" *Philosophical Studies* **77** pp. 57–66, 1995.

Komjáth, P. & Totik, V. (2008). Ultrafilters. *American Mathematical Monthly*, *115*, 33–44.

Kossak, R. & Schmerl, J. (2006). *The structure of models of Peano Arithmetic*. Oxford Logic Guides. Oxford, UK: Clarendon Press.

Kremer, P. (2014). Indeterminacy of fair infinite lotteries. *Synthese*, *191*, 1757–1760.

Kyburg, H. E., Jr. (1961). *Probability and the logic of rational belief*. Middletown, CT: Wesleyan University Press.

Laplace, P.-S. (1814). *Essai philosophique sur les probabilités*. 3th edition printed by V. Courcier, Paris, France, 1816. Translated by Truscott, F. W., Emory, F. L. *Philosophical Essay on Probabilities.* Wiley (1902) New York, NY. Paris, France.

Lehman, R. S. (1955). On confirmation and rational betting. *The Journal of Symbolic Logic*, *20*, 251–262.

Lewis, D. K. (1980). A subjectivist's guide to objective chance. In R. C. Jeffrey (Ed.), *Studies in inductive logic and probability* (Vol. 2, pp. 263–293). Berkeley, CA: University of California Press.

Lewis, D. K. (1986a). Philosophical papers. (Chap. A Subjectivist's Guide to Objective Chance, Vol. 2). Oxford, UK: Oxford University Press.

Lewis, D. K. (1986b). *Philosophical papers*. Oxford, UK: Oxford University Press.

Lindley, D. V. (Ed.). (1991). *Making decisions*. 2nd edition. UK: Wiley.

Lindley, D. V. (Ed.). (2006). *Understanding uncertainty*. UK: Wiley.

Loeb, P. A. (1975). Conversion from nonstandard to standard measure spaces and applications in probability theory. *Transactions of the American Mathematical Society*, *211*, 113–122.

Łoś, J. (1955). Quelques remarques, théorèmes, et problèmes sur les classes définissables d'algèbres. In *Mathematical interpretation of formal systems (symposium, amsterdam 1954)* (Vol. 98, pp. 1–13). Studies in Logic and the Foundations of Mathematics. Amsterdam, The Netherlands: North-Holland Publishing Co.

Luxemburg, W. A. (2007). Foreword. In I. van den Berg & N. Neves (Eds.), *The strength of nonstandard analysis* (pp. v–x). Vienna, Austria: Springer.

Mancosu, P. (2009). Measuring the size of infinite collections of natural numbers: Was Cantor's theory of infinite number inevitable? *The Review of Symbolic Logic*, *2*, 612–646.

Martin-Löf, P. (1990). Mathematics of infinity. In P. Martin-Löf & G. Mints (Eds.), *Colog-88 computer logic* (Vol. 417, pp. 146–197). Lecture Notes in Computer Science. Berlin, Germany: Springer.

McCall, S. & Armstrong, D. M. (1989). God's lottery. *Analysis*, *49*, 223–224.

McGee, V. (1994). Learning the impossible. In E. Eells & B. Skyrms (Eds.), *Probability and conditionals: Belief revision and rational decision* (pp. 179–199). Cambridge, UK: Cambridge University Press.

McGee, V. (2002). Nonstandard models of true arithmetic. Lecture notes for course 'Logic II' at MIT; http://web.mit.edu/24.242/www/NonstandardModels.pdf.

Nelson, E. (1977). Internal set theory: A new approach to nonstandard analysis. *Bulletin of the American Mathematical Society*, *83*, 1165–1198.

Nelson, E. (1987). *Radically elementary probability theory*. Princeton, NJ: Princeton University Press.

Oberschelp, W. (1962–63). Über die Begründung wahrscheinlichkeitstheoretischer Axiome durch Wetten. *Archiv für mathematische Logik und Grundlagenforschung*, *6*, 35–51.

Oppy, G. (1990). On Rescher on Pascal's wager. *International Journal for Philosophy of Religion*, *30*, 159–168.

Painlevé, P. (1967). *Analyse des travaux scientifiques*. Reprinted in: "Œuvres de Paul Painlevé", Éditions du CNRS, Paris (1972), Vol. 1, pp. 72–73. Paris, France: Albert Blanchard.

Palmgren, E. (1998). Developments in constructive nonstandard analysis. *The Bulletin of Symbolic Logic*, *4*, 233–272.

Parikh, R. & Parnes, M. (1974). Conditional probabilities and uniform sets. In A. Hurd & P. Loeb (Eds.), *Victoria symposium on nonstandard analysis* (Vol. 369, pp. 177–188). Lecture Notes in Mathematics. Berlin, Germany: Springer.

Parker, M. (2013). Set size and the part–whole principle. *Review of Symbolic Logic*, *6*, 589–612.

Parker, M. (2018). *Symmetry arguments against regular probability: A reply to recent objections*. Unpublished manuscript; URL: http://philsci-archive.pitt.edu/14362/.

Pascal, B. (1670 / 1995). *Pensées*. Translated by A.J. Krailsheimer. Penguin Classics.

Peano, G. (1889). *Arithmetices principia, nova methodo exposita*. Translated as "The principles of arithmetic, presented by a new method" by

J. Van Heijenoort in J. Van Heijenoort, editor, *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*, Harvard University Press, Cambridge, MA (1977) 83–97; http://books.google.be/books?id=v4tBTBlU05sC&pg=PA83. Turin, Italy: Bocca.

Pedersen, A. P. (2014). Comparative expectations. *Studia Logica*, *102*, 811–848.

Perkins, E. (1981). A global intrinsic characterization of Brownian local time. *The Annals of Probability*, *9*, 800–817.

Pivato, M. (2014). Additive representation of separable preferences over infinite products. *Theory and Decision*, *77*, 31–83.

Pruss, A. (2012). Infinite lotteries, perfectly thin darts, and infinitesimals. *Thought*, *1*, 81–89.

Pruss, A. (2013). Probability, regularity, and cardinality. *Philosophy of Science*, *80*, 231–240.

Pruss, A. (2014). Infinitesimals are too small for countably infinite fair lotteries. *Synthese*, *191*, 1051–1057.

Ramsey, F. P. (1931). Truth and probability. In R. B. Braithwaite (Ed.), *The foundations of mathematics and other logical essays* (Vol. 5, pp. 156–198). International library of psychology, philosophy, and scientific method. (Original paper from 1926). London, UK: Routledge & P. Kegan.

Robinson, A. (1961). Non-standard analysis. *Proceedings of the Royal Academy of Sciences, Amsterdam, ser. A*, *64*, 432–440.

Robinson, A. (1966). *Non-standard analysis*. Amsterdam, The Netherlands: North-Holland.

Schechter, E. (1997). *Handbook of analysis and its foundations*. San Diego, CA: Academic Press (Elsevier).

Schmieden, C. & Laugwitz, D. (1958). Eine Erweiterung der Infinitesimalrechnung. *Mathematisches Zeitschrift*, *69*, 1–39.

Schurz, G. & Leitgeb, H. (2008). Finitistic and frequentistic approximation of probability measures with or without σ-additivity. *Studia Logica*, *89*, 257–283.

Shafer, G. (2008). The game-theoretic framework for probability. In B. Bouchon-Meunier, C. Marsala, M. Rifqi, & R. R. Yager (Eds.), *Uncertainty and intelligent information systems* (pp. 3–15). Hackensack, NJ: World Scientific.

Shimony, A. (1955). Coherence and the axioms of confirmation. *The Journal of Symbolic Logic*, *20*, 1–28.

Skolem, T. A. (1923). Einige bemerkungen zur axiomatischen begründung der mengenlehre. *Proc. 5th Scandinaviska Matematikerkongressen, Helsingfors, July 4–7, 1922*, 217–232. Translated as "Some remarks on axiomatized set theory" by S. Bauer-Mengelberg in J. Van Heijenoort, editor, *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–*

1931, Harvard University Press, Cambridge, MA (1977) 290–301; http://books.google.be/books?id=v4tBTBlU05sC&pg=PA290.

Skolem, T. A. (1934). Über die Nicht-charakterisierbarkeit der Zahlenreihe mittels endlich oder abzählbar unendlich vieler Aussagen mit ausschliesslich Zahlenvariablen. *Fundamenta Mathematicae*, *23*, 150–161.

Skyrms, B. (1980). *Causal necessity*. New Haven, CT: Yale University Press.

Skyrms, B. (1983a). Three ways to give a probability assignment a memory. In J. Earman (Ed.), *Testing scientific theories* (Vol. 10, pp. 157–161). Minnesota Studies in the Philosophy of Science. Minneapolis: University of Minnesota Press.

Skyrms, B. (1983b). Zeno's paradox of measure. In R. S. Cohen & L. Laudan (Eds.), *Physics, philosophy and psychoanalysis: Essays in hounour of Adolf Grunbaum* (pp. 223–254). Dordrecht, The Netherlands: Reidel.

Skyrms, B. (1995). Strict coherence, sigma coherence and the metaphysics of quantity. *Philosophical Studies*, *77*, 39–55.

Stillwell, J. (1977). Concise survey of mathematical logic. *Australian Mathematical Society Journal (Series A)*, *24*, 139–161.

Stolz, O. (1883). Zur Geometrie der Alten, insbesondere über ein Axiom des Archimedes. *Mathematische Annalen*, *22*, 504–519. Based on an earlier publication in "Berichten des naturwissenschaftlich-medicinischen Vereines in Innsbruck", 1882, volume 12, p. 74.

Stolz, O. (1885). *Vorlesungen über allgemeine Arithmetik*. Leipzig, Germany: Teubner.

Tao, T. (2007–2012). http://terrytao.wordpress.com/tag/nonstandard-analysis/.

Tao, T. (2007). http://terrytao.wordpress.com/2007/06/25/ultrafilters-nonstandard-analysis-and-epsilon-management/.

Tao, T. (2012). http://terrytao.wordpress.com/2012/04/02/a-cheap-version-of-nonstandard-analysis/.

Weintraub, R. (2008). How probable is an infinite sequence of heads? A reply to Williamson. *Analysis*, *68*, 247–250.

Wenmackers, S. (2011). *Philosophy of probability: Foundations, epistemology, and computation* (Doctoral dissertation, University of Groningen, Groningen, The Netherlands). http://philpapers.org/archive/WENPOP.

Wenmackers, S. (2012). Ultralarge and infinite lotteries. In B. Van Kerkhove, T. Libert, G. Vanpaemel, & P. Marage (Eds.), *Logic, philosophy and history of science in belgium ii; proceedings of the young researchers days 2010* (pp. 59–66). Belgium, Brussels: Koninklijke Vlaamse Academie van België voor Wetenschappen en Kunsten.

Wenmackers, S. (2013). Ultralarge lotteries: Analyzing the lottery paradox using non-standard analysis. *Journal of Applied Logic*, *11*, 452–467.

Wenmackers, S. (2018). *Do infinitesimal probabilities neutralize the infinite utility in Pascal's wager?* Forthcoming in P. Bartha and L. Pasternack (eds.) *Classic Arguments in the History of Philosophy: Pascal's Wager*, Cambridge, UK: Cambridge University Press.

Wenmackers, S. & Horsten, L. (2013). Fair infinite lotteries. *Synthese*, *190*, 37–61.

Williamson, T. (2007). How probable is an infinite sequence of heads? *Analysis*, *67*, 173–180.