Infinitesimal Probabilities*

Vieri Benci, Leon Horsten, and Sylvia Wenmackers

ABSTRACT

Non-Archimedean probability functions allow us to combine regularity with perfect additivity. We discuss the philosophical motivation for a particular choice of axioms for a non-Archimedean probability theory and answer some philosophical objections that have been raised against infinitesimal probabilities in general.

---

1 Introduction

We have proposed a specific non-Archimedean probability theory (henceforth called NAP), which allows the assignment of non-zero probabilities to infinitely unlikely events (Benci *et al.* [2013]). Examples of such events include the random or biased selection of an element from the set of the natural numbers or the integers, or from an interval of the rational or real numbers.[1] Like classical probability theory, NAP is applicable in a wide range of situations and can be employed to model different sources of uncertainty. As such, NAP is of relevance both to scholars who are interested in objective probability (or 'chance') and to those interested in subjective probability (and in particular in the rational kind thereof, 'credence'). Moreover, we think that NAP can be useful in the context of physics, where similar methods have found applications already (see Albeverio *et al.* [1986] and references in Cutland [1983]).

  NAP is motivated by four desiderata for a theory of probability: regularity, totality, perfect additivity, and weak Laplacianism. First, 'regularity' is the constraint that the probability of a possible event (that is, a non-empty subset of the sample space) should be strictly larger than that of the impossible event (that is, the empty set). It is a special case of the Euclidean principle, which requires that any set should be given a strictly larger probability than each of its strict subsets.[2] More generally, we want our probability function to be maximally sensitive to differences in this partial order (inclusion) between events. Second, 'totality' is the desideratum that all subsets of the sample space must be assigned a probability value. In other words, all sets should be measurable. Third, 'perfect additivity' is the requirement that the probability of an arbitrary union of mutually disjoint events is equal to the sum of the

---

[1]   Outside the context of this article, it may be better to refer to our theory as NAP–BHW, to distinguish it from related theories. Similar approaches have recently been developed by Hammond ([1994]), Pivato ([2014]), and Pedersen ([unpublished]). (For more on related work, see Pivato [2014], p. 55ff.)

[2]   Regularity as a norm for probability theory and the use of infinitesimals to attain this norm have been discussed both in the context of objective probability (see, for example, Hofweber [2014] and references therein) and in the context of subjective probability (see, for example, Pedersen [2014] and references therein). See also the references in (Hájek [unpublished]).

probabilities of the separate events, where 'sum' has to be defined in an appropriate way in the infinite case.[3] Fourth, 'weak Laplacianism' is the requirement that a probability theory should allow for a uniform probability distribution on sample spaces of any cardinality as well as many other probability ratios between the atomic events. That is, the theory should allow for a mathematical representation of any probabilistic situation that is conceptually possible (from a pretheoretic standpoint).

Non-Archimedean probability theories have been developed that are unobjectionable from a mathematical point of view. But in recent years, philosophical arguments have been developed by Williamson ([2007]) Easwaran ([2014]), and others that purport to show on conceptual grounds that an appeal to infinitesimal probability values is inherently problematic. The main purpose of the present article is to defend non-standard probability against these critiques. We shall argue that the mathematical details of the non-Archimedean probability theory matter in this discussion. In particular, we will show how NAP can provide a diagnosis of where the objections against appealing to infinitesimals in probability theory go wrong.

The structure of this article is as follows: First, we describe the limitations of the orthodox approach to probability theory (Section 2). Subsequently, we describe one particular theory of non-standard probability, called NAP (Section 3), which satisfies the four desiderata listed above. Then we discuss various objections against the use of infinitesimals in probability theory and we evaluate them in the context of NAP (Section 4). We show that the arguments against infinitesimal probability values do not establish what they seek to establish and we argue that the proposed account cannot be dismissed on the basis of the arguments that have been adduced in the literature. We then expand on the virtues of infinitesimal probabilities. We explain how NAP provides satisfactory models for the probabilistic scenarios that classical probability theory cannot adequately describe; we show how NAP yields total, regular, and perfectly additive probability functions even for uncountable domains, and we indicate the role that NAP can play in decision theory (Section 5). In the concluding section, we discuss some more general underlying concerns and evaluate the viability of NAP in light of its advantages and its drawbacks (Section 6). In the Appendix, more details are given about the construction of NAP models.

2 The Limits of Classical Probability Theory

The axioms of Kolmogorov constitute the basis of the received view of probability theory.

---

[3] The requirement for perfect additivity is closely related to that of regularity and totality. Skyrms ([1983]) called it 'ultra-additivity' and analysed it as a Zenonian intuition.

2.1 Classical probability functions

The set of atomic outcomes, Ω, is called the 'sample space'; the σ-algebra on Ω, $\mathfrak{A}$, is called the 'event space'. The following set of axioms is equivalent to those presented by Kolmogorov ([1956]):

K0. Domain and Range: The events are the elements of a σ-algebra $\mathfrak{A} \subseteq \mathcal{P}(\Omega)$ and the probability function is a function

$$P_{\mathrm{K}} : \mathfrak{A} \rightarrow \mathbb{R}.$$

K1. Non-negativity: $\forall A \in \mathfrak{A}$,

$$P_{\mathrm{K}}(A) \geq 0.$$

K2. Normalization:

$$P_{\mathrm{K}}(\Omega) = 1.$$

K3. Additivity: If A and B are events and $A \cap B = \emptyset$, then

$$P_{\mathrm{K}}(A \cup B) = P_{\mathrm{K}}(A) + P_{\mathrm{K}}(B).$$

K4. Continuity: Let

$$A = \bigcup_{n \in \mathbb{N}} A_n,$$

where $A_n \subseteq A_{n+1}$ are elements of $\mathfrak{A}$; then

$$P_{\mathrm{K}}(A) = \lim_{n \to \infty} P_{\mathrm{K}}(A_n). \tag{1}$$

Adding this last axiom to the previous ones is equivalent to requiring 'countable additivity' or 'σ-additivity'.

The triple $\langle \Omega, \mathfrak{A}, P_{\mathrm{K}} \rangle$ is called a 'classical' probability space. Classical probability theory is mathematically coherent and useful. The existence of models for the axioms proves its consistency and the wide range of applications by physicists, engineers, and economists shows its usefulness in modelling situations in the physical world. Nonetheless, there are probabilistic scenarios involving infinite sample spaces that cannot be described in a satisfactory manner in terms of probability functions that are governed by Kolmogorov's axioms.

2.2 Limitations

The axioms of Kolmogorov lead to a probability theory that does not respect any of the four principles mentioned in the introduction. By considering uncountable sample spaces, it is clear that the classical approach does not guarantee regularity, totality, or perfect additivity (see, for example, Skyrms [1983]). Moreover, the orthodox theory violates weak Laplacianism, since it does not allow us to represent uniform probability distributions on countable

sample spaces. Consider the fair lottery on $\mathbb{N} = \{1, 2, 3, \ldots\}$, which is sometimes called the 'de Finetti lottery' (also discussed in Wenmackers and Horsten [2013]). It is easy to see that there is no coherent way to describe it in terms of classical probability functions. Because of the Archimedean property of the real numbers (that are used in the value range of classical probability functions) and finite additivity, the probability of any particular ticket winning has to be set to zero.[4] This entails that either the normalization axiom or the continuity axiom has to be abandoned. The first option relates to proposals for unbounded probability (such as that of Rényi [1955]). With NAP, however, we opt for the second option.

In the context of a subjective interpretation of probability, de Finetti ([1974]) advocated merely finite additivity. Whereas de Finetti did not require that probabilities be assigned to all events of a σ-algebra $\mathfrak{A}$, which is part of K0, we also introduce the notion of 'semi-classical' probability functions that satisfy K0–K3, and are thus only 'finitely additive'. This is sufficient to satisfy weak Laplacianism at least to a minimal extent: the uniform zero distribution is now consistent with the axioms. However, probability theory lacks mathematical power if it cannot make use of limit behaviour for calculating probabilities.[5] In NAP, countable additivity is replaced by an axiom that is compatible with a stronger form of additivity (perfect additivity) and that does provide an alternative sense of limit operations (as explained further on).

There is a second aspect of a semi-classical description of the fair lottery on $\mathbb{N}$ that is unattractive: by assigning probability zero to each natural number, the semi-classical probability values collapse a distinction between any infinitely improbable but possible event ('remote contingency') on the one hand, and the impossible event (empty set) on the other hand. Of course, the same violation of regularity occurs for classical probability functions—whether fair or not—on uncountable domains, where they too have to assign probability zero to many contingent events.

The assignment of probability zero to remote contingencies creates problems for formally modelling conditionals,[6] utilities, and learning situations. The following observations are phrased in terms of fair infinite lotteries, but they apply to all situations that use (semi-)classical probability functions to model (countably) infinite event spaces.

---

[4] The Archimedean property of $\mathbb{R}$ says that given any strictly positive real number, r, and a larger real number, R, there exists a natural number, N, such that the product $N \times r$ exceeds R; in other words, $\mathbb{R}$ does not contain infinitesimals.

[5] In practice, statisticians do not run into problems with the failure of weak Laplacianism, since they have accustomed themselves to working on continuum-sized spaces, where σ-additivity is compatible with uniform distributions provided that all point events are given probability zero.

[6] In particular, when one uses Adams's thesis (see, for example, Arlo-Costa [2008]).

First, the (semi-)classical approach with the ratio formula alone does not give a satisfactory account of conditional probabilities. We are strongly inclined to think that in the fair lottery on $\mathbb{N}$ or on $\mathbb{R}$, the probability that ticket number 1 wins given that one ticket of the set {1, 2, 3} wins, is $\frac{1}{3}$ (see, for example, Bartha and Hitchcock [1999], p. 407). According to the way of defining conditional probability via a ratio, however, this conditional probability is undefined since it involves putting the probability of the conditioning event—zero in this case—in the denominator. In the context of uncountable sample spaces, it is well known that classical probability functions do not contain enough information to compute all conditional probabilities and the limiting operation must be specified separately.[7]

Second, the (semi-)classical approach leads to problems for decision theory. If the probability of a single ticket winning is zero, then given the standard mechanism for calculating expected utilities, participation in the fair lottery on $\mathbb{N}$ could not have non-zero expected utility for an agent, even if the prize for winning is very high. So, an agent should be indifferent between owning a single (or any finite set of) ticket(s) and owning none at all. For a fair infinite lottery on $\mathbb{R}$, the agent should be indifferent between owning the set of all rational number tickets and owning none at all. This seems incorrect.

Third, the (semi-)classical approach does not accommodate the possibility of learning from remote evidence. Suppose that an agent does participate in a fair lottery on an infinite sample space. Suppose that her credences are regulated by a (semi-)classical probability function, as they should be according to Bayesian accounts of subjective probability. Suppose further that our agent happens to have drawn the winning ticket. Then she will want to update her credences on the evidence that she has received, but she cannot update in the normal, Bayesian manner a probability that started out as zero to any other value. An important instance of this problem: if the probability of history going as it actually goes is zero (according to some system of laws), one cannot update on the present state (within this system).

2.3 Infinitesimals to the rescue?

The three problems discussed in the previous section can be avoided if the probability functions are regular.[8] The most straightforward suggestion to obtain regular probability functions is to make room for attributing

---

[7] When taking the limit of conditional probabilities to an empty conditioning event, this results in puzzles like the Borel–Kolmogorov paradox (Kolmogorov [1956], pp. 50–1). Concerning such cases, Jaynes ([2003], Section 15.7) has written: 'In general, the final result will and must depend on which limiting operation was specified'.

[8] We are not suggesting that this is the only viable option to address these problems; in Section 5.4, we briefly discuss the relation between our approach and approaches that are based on conditional probability functions or lexicographical probabilities.

infinitesimally small but non-zero probability values to events. In the fair lottery on $\mathbb{N}$, for instance, it seems reasonable to judge the probability that a given ticket wins to be infinitesimally small, but non-zero. This suggestion was pursued by several authors on several occasions (see Footnote 2).

Regularity can be introduced in the axioms by strengthening K1, but it also requires modifying principles K0 and K4. Regarding K0, the value range of generalized probability functions must be extended. One requirement is that we should be able to calculate with the generalized probability values much as we are able to calculate with classical probability values. In particular, there must be a natural way of adding and multiplying them (to allow computation of probabilities of unions and intersections of events). Fortunately, since the work of Robinson ([1961]), we know a precise sense in which systems of real numbers that include infinitesimals can be taken to form a field.

Nonetheless, the question remains of how non-standard probability values should be attached to events of a sample space. Ideally, we want to be able to do this in such a way that perfect additivity is also satisfied, by replacing K4 with a different infinite additivity axiom. Applied to the fair lottery on $\mathbb{N}$, there is a very simple (and, as we will see very shortly, naïve) proposal how this can be done. The 'measure' of the ordered set of the natural numbers $\mathbb{N}$, one might say, is ω: the smallest infinite ordinal number. Therefore, '$\frac{1}{\omega}$' might be thought of as an infinitesimally small number, which seems a good candidate for assigning to any point event as a probability value after normalization.[9] After all, this appears to yield a very natural countable additivity property: $\omega \cdot (\frac{1}{\omega}) = 1$.[10]

Unfortunately, this idea does not work. We want our generalized probability function to be maximally sensitive to distinctions in sizes of events. We want P to be such that

$$P(\{\emptyset\}) < P(\{1\}) < P(\{1, 2\}) < P(\{1, 2, 3\}) < \ldots$$

This means that we must exhaust all the finite numbers to measure the finite sets. But then the set $\{2, 3, 6, \ldots\}$ of even numbers, for instance, must surely already be assigned an infinite measure before normalization. And therefore the measure of the ordered set $\mathbb{N}$ (which must be strictly larger than that of the set of the even numbers) must be much larger than the first infinite ordinal ω.

More generally, we would like the generalized probability function to satisfy the 'Euclidean principle': For all events A and B, if A is a proper

---

[9] We have put it between quotation marks, since the reciprocal of an ordinal number is undefined.

[10] These considerations may provide motivation to explore the use of Conway's surreal numbers as probability values. We are grateful to Kenny Easwaran for this observation; see also his (Easwaran [2014], p. 38).

Table 1.   Quantitative probability theories

	Domain: Standard	Non-standard
Range: $\mathbb{R}$	Kolmogorov	Loeb
Non-Archimedean field	NAP	Nelson

sub-event of B, then $P(A) < P(B)$ (see Section 3.6; Benci *et al.* [2006]; Parker [2013]). This desideratum, which is equivalent to the aforementioned regularity demand, makes it a not altogether trivial task to describe even the fair lottery on $\mathbb{N}$ in terms of infinitesimals.

Consistent theories of probability functions that draw on ideas from non-standard analysis (NSA) have been proposed in the literature. A classification of them is given in Table 1. Two well-known examples are the theory of Loeb measures ([1975]) and Nelson's ([1987]) 'radically elementary probability theory'. However, Loeb's and Nelson's theories describe lotteries on non-standard domains, so they simply do not address the problem of describing the fair lottery on $\mathbb{N}$ (or its natural generalizations to $\mathbb{Q}$, $2^{\mathbb{N}}$,...). The task before us is to describe fair lotteries on such standard domains. Loeb's theory has a standard co-domain ($\mathbb{R}$), whereas in Nelson's theory, the co-domain is also non-standard.

It is not difficult to construct non-standard probability models that are the exact analogues of semi-classical models, namely, models with a standard sample space that assign non-standard real numbers to events and that in addition satisfy Kolmogorov's axioms except σ-additivity (and K0, of course). One can even force such models to be regular (McGee [1994]). However, it is not straightforward to construct a class of such models that in addition have plausible infinite additivity properties. This requires a new concept of a 'limit of probabilities', which was developed in (Benci *et al.* [2013]). The resulting theory, NAP, will be described in Section 3 below; NAP occupies the fourth quadrant in Table 1.[11]

Note that we have not started out from a non-standard measure theory that we then applied to the concept of probability. Instead, we started from four intuitive requirements about probability. The model for this happens to require a fine free ultrafilter (or equivalently, a maximal ideal), just like NSA does, but this does not reduce NAP to NSA—nor vice versa: NSA and NAP have different motivations and interpretations, which turn out to be related to the same underlying mathematical structure.

---

[11] Pivato ([2014]) develops an approach that should also be situated in this quadrant.

3 NAP Theory

In this section, we describe NAP in an axiomatic way so that it can be compared directly with Kolmogorov's axioms. We start with the first four axioms of NAP. Then we discuss the last axiom, which is the most delicate one.

3.1 First four axioms of NAP

The first four axioms of NAP are the following:

NAP0. Domain and Range: The events are all the subsets of Ω, which is a finite or infinite sample space. Probability is a total function

$$P : \mathcal{P}(\Omega) \rightarrow \mathfrak{R},$$

where $P(\Omega)$ is the powerset (set of all subsets) of Ω and $\mathfrak{R}$ is a superreal field (that is, an ordered field that contains the real numbers as a subfield).

NAP1. Regularity: $P(\emptyset) = 0$ and $\forall A \in \mathcal{P}(\Omega) \backslash \{\emptyset\}$,

$$P(A) > 0. \tag{2}$$

NAP2. Normalization:

$$P(\Omega) = 1. \tag{3}$$

NAP3. Additivity: If A and B are events and $A \cap B = \emptyset$, then

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

Observe that in the axioms NAP0–NAP3, the field $\mathfrak{R}$ is not specified.[12,13] It is important to notice that NAP uses the domain to build the range. For example, consider the case with $\Omega = \{a, b\}$, $P(\{a\}) = 1/\sqrt{2}$, and hence $P(\{b\}) = 1 - 1/\sqrt{2}$. In this case the natural field is $\mathbb{Q}(\sqrt{2})$. However, as long as there is no need to introduce infinitesimal probabilities, all these fields are contained in $\mathfrak{R}$ and hence it is simpler to take $[0, 1]_\mathbb{R}$ as the range.

Immediate consequences of the axioms are[14]:

Proposition 1
If NAP0, NAP1, NAP2, and NAP3 hold, then:

(1) $\forall A \in \mathcal{P}(\Omega), P(A) \in [0, 1]_\mathfrak{R}$
(2) $P(A) = 1 \Leftrightarrow A = \Omega$

---

[12] If we would specify some range upfront, even if it would be a non-Archimedean set, we would not be able to guarantee regularity. This can be seen from a cardinality argument introduced by Hájek ([unpublished]) and formalized by Pruss ([2013]): these impossibility results assume a fixed range and hence do not apply to NAP.

[13] This need not be surprising: in finite probability theory, the same happens. In many games, such as games involving fair dice, we have $\mathfrak{R} = \mathbb{Q}$, but it is possible to have less familiar fields.

[14] For a proof of this (elementary) proposition, see (Benci *et al.* [2013]).

(3) Moreover, assume that one of the following holds:

- ○ Ω is countable and the theory is fair, namely, $\forall \omega, \tau \in \Omega$, $P(\{\omega\}) = P(\{\tau\})$

- ○ Ω is uncountable

then $\mathfrak{R}$ is a non-Archimedean field.

This proposition demonstrates that non-Archimedean fields arise quite naturally from axiom NAP1.

3.2 Continuity and conditional probability

We have noted that a retreat to finite additivity is unsatisfactory because it does not allow the calculation of infinitely disjunctive events on the basis of limit behaviour. For the same reason the probability theory consisting solely of NAP0–NAP3 is too weak. We need to add an axiom that replaces axiom K4 of classical probability theory and that allows some kind of infinite sum. The trouble with this point, however, lies with the limit operation. In fact, if we want to take the limit of a sequence of points $a_n \in X$, it seems desirable for X to be complete since otherwise the Cauchy sequences might not be convergent and this fact prevents the development of any interesting calculus. In probability theory, X needs to be an ordered field as well. But the only complete ordered field is $\mathbb{R}$; no non-Archimedean field is complete. This is the main technical problem in dealing with non-Archimedean fields. We proposed to solve this problem by constructing a different notion of limit, which we here call the 'Ω-limit'. (In Benci *et al.* [2013] this limit was called the 'Λ-limit'.)

In order to present the Ω-limit in a natural way, we will introduce the following principle, which states that fixing the conditional probability, $P(A|\lambda_n)$, for a sufficiently large family of finite sets, λ_n, determines the value of $P(A|\Omega)$, which is nothing but the unconditional probability $P(A)$. The same idea is also present in Kolmogorov's classical setting, only the details of the limit operation are different:

Conditional Probability Principle (CPP): Let $\{\lambda_n\}$ be a family of events such that $\lambda_n \subseteq \lambda_{n+1}$ and $\Omega = \bigcup_{n \in \mathbb{N}} \lambda_n$; then, eventually

$$P_K(\lambda_n) > 0,$$

and, for any event, A, we have that

$$P_K(A) = \lim_{n \to \infty} P_K(A|\lambda_n).$$

It is easy to prove that CPP is equivalent to K4 (Benci *et al.* [2013]); the advantage of CPP is that it is easier to reformulate it in an NAP context. More precisely: we shall give an appropriate notion of limit that allows us to formulate a variant of CPP within NAP, and this will be the final axiom of NAP.

3.3 The final axiom of NAP Theory

It is possible to associate a hyperreal number with any real-valued function that is defined only on the finite subsets of the sample space, such that certain natural algebraic properties hold among these numbers. The details are given in the Appendix. We call this operation the 'Ω-limit' (denoted as $\lim_{\lambda\uparrow\Omega}$) and one may interpret the resulting hyperreal number as the function's 'value at infinity' (though it is not to be confused with the usual Archimedean limit, denoted as $\lim_{n\to\infty}$). With the help of this non-Archimedean limit operation, we can now formulate the CPP in NAP, which in some sense replaces axiom K4 of classical probability theory. Please note that $\mathcal{P}_{fin}(A)$ denotes the collection of finite subsets of a set A. This axiom is the keystone of NAP.

NAP4. CPP in NAP: For any $A \in \mathcal{P}(\Omega)$ and any $\lambda \in \mathcal{P}_{fin}(\Omega)$,

$$P(A|\lambda) \in \mathbb{R} \tag{4}$$

and

$$P(A) = \lim_{\lambda\uparrow\Omega} P(A|\lambda). \tag{5}$$

This limit can be rigorously defined and shown to exist (see Appendix), and the functions resulting from NAP0–NAP4 can be shown to satisfy the four desiderata for probability functions that were discussed in the introduction.

The intuitive meaning of Equation (5) is obvious: the probability of an event A is the Ω-limit of the conditional probability $P(A|\lambda)$ obtained by a finite sample set λ. We can give a suggestive interpretation to Equations (4) and (5) as follows: We may think of the real number $P(A|\lambda)$ as the result of experiments. The probability, $P(A)$, of event A is the 'abstract' extrapolation from the results of all possible finite experiments.

Formally, CPP and NAP4 are similar, and they are also similar in interpretation. But from a technical point of view, they are quite different. For example, since λ is a finite set, usually in classical theory $P_K(\lambda) = 0$ and hence $P_K(A|\lambda)$ cannot be defined. In NAP, in contrast, $P(A|\lambda)$ plays a central role.[15]

3.4 Infinite sums

The Weierstrass notion of the classical limit is assumed in the rigorous definition of the sum of an infinite sequence. Analogously, the Ω-limit allows us to define the sum of infinitely many real numbers. In this section, we will investigate this operation and, in the next section, it will be applied to NAP.

---

[15] By CPP, $P_K(\cdot|\lambda)$ is used to define $P_K(\cdot)$, but $P_K(\cdot|\lambda)$ cannot always be retrieved from the information encoded in $P_K(\ldots)$. In NAP, $P(\cdot|\lambda)$ is important both in the construction of $P(\cdot)$ and it can be retrieved from it by the usual ratio formula.

Let x_ω be a family of real numbers indexed by $\omega \in E \subseteq \Omega$. The Ω-sum of all x_ωs is defined as follows:

$$\sum_{\omega \in E} x_\omega = \lim_{\lambda \uparrow \Omega} \sum_{\omega \in E \cap \lambda} x_\omega. \tag{6}$$

Notice that, since λ is always finite, the function

$$\varphi(\lambda) := \sum_{\omega \in E \cap \lambda} x_\omega$$

of λ is well defined, always yielding a real number as function value.

The main differences between our new type of infinite sum and the classical series are:

- As shown in the Appendix, the Ω-sum depends on the choice of a free ultrafilter, $\mathcal{U}_\Lambda$. This is not the case with the usual series. So it would actually be more appropriate to write $\sum_{\omega \in E; \mathcal{U}_\Lambda} x_\omega$ rather than $\sum_{\omega \in E} x_\omega$.

- The Weierstrass-sum of a series exists only for certain countable sets of real numbers, while the Ω-sum exists for every family of real numbers indexed by $\omega \in E \subseteq \Omega$. In principle, Ω and hence E may have any cardinality.

- The Weierstrass-sum of a sequence—if it exists—is a real number, while the result of an Ω-sum is a hyperreal number in $\mathfrak{R}$.

3.5 Definition of NAP functions via infinite sums

One of the main consequences of axiom K4 is σ-additivity, which defines infinite sums and relates them to probabilities of unions of countably many events. In this section, we will see that its non-Archimedean counterpart, axiom NAP4, also allows us to relate the infinite sums defined in the previous section to (generalized) probability functions and to generalize well-known properties used in finite probability theory.

Weight Function:

$$w : \Omega \to \mathbb{R}.$$

A weight function describes the relative probability of elementary events. Notice that two different weight functions that are proportional to each other are equivalent for all practical purposes.

The following can be shown (Benci *et al.* [2013]):

Proposition 2

The function w takes its values in $\mathbb{R}$ and for any finite λ, the following holds:

$$P(A|\lambda) = \frac{\sum_{\omega \in A \cap \lambda} w(\omega)}{\sum_{\omega \in \lambda} w(\omega)}. \tag{7}$$

Taking the Ω-limit of both sides of Equation (7), we get

$$P(A) = \frac{\sum_{\omega \in A} w(\omega)}{\sum_{\omega \in \Omega} w(\omega)}. \tag{8}$$

More generally, we have for any $A, B \in \mathcal{P}(\Omega)$,

$$P(A|B) = \frac{\sum_{\omega \in A \cap B} w(\omega)}{\sum_{\omega \in B} w(\omega)}. \tag{9}$$

These properties generalize well-known properties that hold when Ω is a finite probability space. But these formulas say more: they say that in order to know the probability of any event, A, it is sufficient to know the relative probability, $w(\omega)$, of each elementary event, ω, and the rule that allows us to take an infinite sum, that is, the rule that allows us to take the Ω-limit (which is defined by Equation (A.2) via a free ultrafilter, $\mathcal{U}_\Lambda$; see Appendix). Since Equation (8) holds for arbitrary w, weak Laplacianism is fulfilled.

The main result, that NAP functions on infinite sample spaces exist, was shown in (Benci *et al.* [2013]), but it can be seen from combining the proof in the Appendix concerning Ω-limits, the definition in Equation (5), and Proposition 2.

So the NAP space is a triple $\langle \Omega, w, \mathcal{U}_{\mathcal{P}_{fin}(\Omega)} \rangle$ where:

- Ω is the sample space;
- $w : \Omega \rightarrow \mathbb{R}^+$ is a weight function;
- $\mathcal{U}_{\mathcal{P}_{fin}(\Omega)}$ is a free ultrafilter on $\mathcal{P}_{fin}(\Omega)$.

Regularity is imposed by requiring the ultrafilter to be 'fine' (see Kanamori [1994], p. 301).

3.6 Relation to numerosity theory

NAP is related to the theory of numerosity introduced in (Benci [1995]) and developed in various directions (Gilbert and Rouche [1996]; Benci and Di Nasso [2003]; Benci *et al.* [2006]). We briefly sketch the main tenets of this theory.

In order to count the elements of a set, A, it is necessary to have a set of numbers, $\mathfrak{N}$, and a rule, $\mathfrak{s}$, that specifies the number of elements, $\mathfrak{s}(A) \in \mathfrak{N}$, that belong to set A. More precisely, we can say that the operation of counting consists of a triple $(\mathfrak{U}, \mathfrak{N}, \mathfrak{s})$, where $\mathfrak{U}$ is the family of sets that can be counted, $\mathfrak{N}$ is the set of numbers, and $\mathfrak{s} : \mathfrak{U} \rightarrow \mathfrak{N}$ is a function. In the following, we shall call a triple, $(\mathfrak{U}, \mathfrak{N}, \mathfrak{s})$, that satisfies the basic properties related to our intuition of counting a 'counting system'. We highlight the following two principles governing counting systems, which are important for the basic view:

Euclidean Principle (EP): If $A \subsetneq B$, then $\mathfrak{s}(A) \lneqq \mathfrak{s}(B)$

Humean Principle (HP): If the elements of A can be put in a one-to-one correspondence with the elements of B, then $\mathfrak{s}(A) = \mathfrak{s}(B)$.

If we take $\mathfrak{U} = \mathfrak{Fin}$ (the class of finite sets), $\mathfrak{N} = \mathbb{N}$ (the set of natural numbers), and $\mathfrak{s} = |\cdot|$ (the usual function that gives the number of elements of a finite set), then we obtain the 'natural numbers counting theory' $(\mathfrak{Fin}, \mathbb{N}, |\cdot|)$. Of course, it satisfies all the intuitive properties of counting plus EP and HP, since those properties are extracted from intuitions that are largely based on dealing with finite sets and natural numbers. If infinite sets with strict subsets of the same cardinality are included in $\mathfrak{U}$, then it is well known that the properties EP and HP are inconsistent with each other. However, then there remain consistent counting theories that are based on either EP or HP.

Cantor was the first to realize this. He abandoned EP and constructed on the basis of HP the theory of cardinal numbers $(\mathfrak{S}, \mathrm{Card}, |\cdot|)$, where $\mathfrak{S}$ is the class of all sets and Card denotes the class of cardinal numbers.

Cantor also understood that if you count an infinite set, the result (that is, the type of number) obtained depends on the method that you employ for counting. In fact, he generalized the operation of counting in two different ways and he obtained not only the theory of cardinal numbers $(\mathfrak{S}, \mathrm{Card}, |\cdot|)$ but also the theory of ordinal numbers (WO, Ord, ord); here WO denotes the class of well-ordered sets, Ord the class of ordinal numbers, and ord the order type of a well-ordered set.[16] The two counting systems give different results when applied to infinite sets. Also, it is well known that the arithmetic in Card and Ord does not satisfy the usual algebraic rules that we are used to based on our experience with natural numbers. For instance, reciprocals are not defined for either (cf. Section 2.3).

Now, the following question arises naturally: 'Is there a different way to count the elements of infinite sets satisfying EP and such that the operations '+' and '·' satisfy the usual algebraic properties?'

The answer is yes, and we will see that the notion of Ω-limit can be used to construct such a counting system. A counting system $(\mathfrak{U}, \ \mathfrak{N}, \ \mathfrak{n})$ which satisfies EP will be called a 'numerosity theory': $\mathfrak{n}$ will be called the numerosity function and $\mathfrak{n}(E)$ is the numerosity of the set E.

The numerosity theory relevant for this article is given by $(\mathcal{P}(\Omega), \mathbb{N}_0^*, \mathfrak{n})$ where for every $A \in \mathcal{P}(\Omega)$, $\mathfrak{n}$ is given by

$$\mathfrak{n}(A) = \lim_{\lambda \uparrow \Omega} |A \cap \lambda|,$$

---

[16] For an account of the history of the measurement of sets of infinite size, see (Mancosu [2009]).

where λ ranges over finite subsets of Ω and $\mathbb{N}_0^*$ is a non-standard model of the natural numbers. In a numerosity theory, the numerosity of $\mathbb{N}$ is denoted by α.[17]

Given the following definition, numerosity theory can be related to NAP in the case of fair lotteries:

> Definition 3: If $\forall \omega_1, \omega_2 \in \Omega$, $w(\omega_1) = w(\omega_2)$, then the probability function (Ω, P) is called fair.

Without loss of generality, we can set $w(\omega) = 1$ for every $\omega \in \Omega$ in the fair case. Hence, if (Ω, P) is a fair lottery:

$$P(A) = \frac{\sum_{\omega \in A} w(\omega)}{\sum_{\omega \in \Omega} w(\omega)} = \frac{\lim_{\lambda \uparrow \Omega} \sum_{\omega \in A \cap \lambda} w(\omega)}{\lim_{\lambda \uparrow \Omega} \sum_{\omega \in \lambda} w(\omega)} = \frac{\lim_{\lambda \uparrow \Omega} |A \cap \lambda|}{\lim_{\lambda \uparrow \Omega} |\lambda|} = \frac{\mathfrak{n}(A)}{\mathfrak{n}(\Omega)}.$$

This formula is an expression of Laplace's famous 'First principle' ([1902], pp. 6, 9): the probability, $P(A)$, of an event, A, is the ratio between the number $\mathfrak{n}(A)$ of favourable cases and the number of all cases, $\mathfrak{n}(\Omega)$, provided that they are equiprobable.[18]

Consider again the infinite set $\mathbb{N} = \{1, 2, 3, \ldots\}$. Now consider the set $S = \{2, 3, \ldots\}$. We may describe this set relative to $\mathbb{N}$ in two ways, which may promote different intuitions about its relative size (cf. Section 4.2):

- If we describe S as $\{n | n \in \mathbb{N} \wedge n \neq 1\} = \mathbb{N} \backslash \{1\}$, then we emphasize that S is a strict subset of $\mathbb{N}$, which suggests that S has a smaller size than $\mathbb{N}$ (following the Euclidean principle of size).

- If we describe S as $\{n + 1 | n \in \mathbb{N}\} = \mathbb{N} + 1$, then we convey that S can be obtained via a re-labelling or translation of the elements of $\mathbb{N}$, which suggests that S has the same size as $\mathbb{N}$ (following the Humean principle of size).

Although the expression '$\mathbb{N} \backslash \{1\}$' refers to the same set S as the expression '$\mathbb{N} + 1$', the corresponding intuitions about the size of S relative to $\mathbb{N}$ cannot hold simultaneously: two sizes cannot both be different and the same. This is a simple illustration of the incompatibility of EP and HP.

The theory of cardinal numbers adopts HP as a criterion of identity. But probability theory cannot accept HP as a criterion of identity. Probability functions should not assign equal probability to all equinumerous sets of the sample space; otherwise it would lead to absurd conclusions, such as $[0, \frac{1}{2}]$ always being equiprobable to $[0, 1]$. But even though HP has to be

---

[17] α can be related to $\aleph_0$ or to ω, but it should not be confused with either. Their relation is quite involved and we refer interested readers to (Benci *et al.* [2006]), where this question has been analysed.

[18] For definiteness, the reader may consider a fair lottery on $\mathbb{N}$; see (Wenmackers and Horsten [2013], Section 6.2; Benci *et al.* [2013], Section 5.2). In particular, it is possible to assign probabilities that are equal to $1/n$ for each of the sets $n\mathbb{N} + i$ for any natural number, n, and $i \in \{0, \ldots, n\}$. This case has been discussed in terms of numerosities in (Mancosu [2009], Section 6.2).

abandoned, there is the possibility of adopting EP, which entails regularity in the context of probabilities. At least in infinite lottery situations, EP has a strong intuitive pull.[19] And we have seen that NAP constructs probability functions on infinite sample spaces that satisfy EP.

4 Objections and Replies

Various authors have formulated objections against the use of infinitesimal probability values. Some of these objections are presented as general arguments, not aimed against a particular theory, but rather against any hypothetical theory that involves non–Archimedean probability values. If any of these arguments are accepted as decisive, then any attempt to work out the details of such a theory is nipped in the bud. NAP shows that it is possible to develop a consistent non-Archimedean theory of probability meeting key conceptual desiderata. We now evaluate the general objections to infinitesimal probabilities in the light of NAP. It will turn out that the arguments against infinitesimal probabilities crucially depend on certain assumptions regarding the properties of probability functions that are taken to be uncontroversial by their authors, but which do require careful scrutiny.[20]

4.1 Cantor and the Archimedean property

We start off with a historically important argument against infinitesimal numbers in general that was formulated by Cantor ([1966], pp. 407–9). His argument is very cryptic: he merely states that not even a transfinite sum of infinitesimals can exceed a non-infinitesimal bound. What is behind this assertion can plausibly be spelled out as follows.

Consider the fair lottery on $\mathbb{N}$ again. Let us entertain the supposition that equal infinitesimal probability ϵ is assigned to each point event. Define

$$\text{Prob}(\{1, 2\}) = \text{Prob}(\{1\}) + \text{Prob}(\{2\}),$$

$$\text{Prob}(\{1, 2, 3\}) = \text{Prob}(\{1\}) + \text{Prob}(\{2\}) + \text{Prob}(\{3\}), \ldots$$

Then consider the following ω-sequence:

$$\text{Prob}(\{1\}), \text{Prob}(\{1, 2\}), \text{Prob}(\{1, 2, 3\}), \ldots$$

This sequence is bounded (by 1), so there must be a least upper bound, which we may call $\omega \cdot \epsilon$. But it is easy to see that between $\omega \cdot \epsilon - \epsilon$ and $\omega \cdot \epsilon$ at most

---

[19] For a vivid description of the intuitive pull of EP in infinite lottery situations, see (McCall and Armstrong [1989]).

[20] In this section, we use the notation '*Prob*' for probability functions in arguments that do not specify the formalism to which the function belongs. In cases where the formalism is clear, we stick to the notations P_K and P for Kolmogorovian and NAP functions, respectively.

one $n \cdot \epsilon$ (with $n \in \mathbb{N}$) can lie. After all, for any $n \in \mathbb{N}$ that lies in this open interval, we must then have:

$$(n - 1) \cdot \epsilon < (\omega \cdot \epsilon) - \epsilon < n \cdot \epsilon < \omega \cdot \epsilon < (n + 1) \cdot \epsilon.$$

This must mean that $\omega \cdot \epsilon$ does not exceed an infinitesimal value. This argument carries over to all limit ordinals, and the conclusion then is that $\lambda \cdot \epsilon$ does not exceed an infinitesimal value for any transfinite number λ. But this means that the infinitesimals are 'disconnected' from the standard numbers. It means that even an arbitrarily large transfinite ordinal cannot carry ϵ above a finite number value. One might sum this up by saying that infinitesimals are not even 'ordinal-Archimedean' (as opposed to 'natural number Archimedean').

4.1.1 Reply: The least upper bound principle

A response to Cantor's objection was given by Zermelo in his comments on Cantor's cryptic argument ([1966], p. 439).[21] He states that Cantor's argument establishes that the number $\omega \cdot \epsilon$ does not exist, rather than that it does not exceed the infinitesimals. In other words, multiplication of infinitesimals by transfinite ordinal numbers is meaningless.

The theory NAP is neutral about the existence or non-existence of transfinite ordinal numbers. So *a fortiori* it does not give a verdict about whether multiplication of transfinite ordinals with infinitesimals makes sense. Rather, it denies that the probability associated with the ω-sequence

$$P(\{1\}), P(\{1, 2\}), P(\{1, 2, 3\}), \ldots$$

is the least probability that is 'infinitely larger' than the probability of the point events. The above ω-sequence is indeed bounded, but it does not have a least upper bound, since the range of our non-Archimedean probability functions is not complete. Nonetheless, this sequence has a limit in a generalized sense (an Ω-limit). And the existence of this limit is sufficient for the existence of the probability of any event in the sample space of the fair lottery on $\mathbb{N}$.

In sum, there simply is no need for us to make sense of $\omega \cdot \epsilon$ in order to compute the probability of any event. Indeed, the theory of well-order types is not the right tool for computing limits of non-standard probabilities.

4.2 Ticket missing from an infinite lottery

Williamson's ([2007]) argument, which will be discussed shortly, involves ω-sequences of fair coin tosses that can be represented as a fair lottery on the Cantor space $2^{\mathbb{N}}$, which is a non-countably infinite sample space. We first

---

[21] For an extensive discussion of Cantor's objections against infinitesimals, see (Ehrlich [2006]).

present a new argument against infinitesimals that is inspired by Williamson's argument, but which only requires a countably infinite sample space. As far as we know, this variation is not endorsed by Williamson or any other author. We present it as an intermediate step to clarify both the structure of the coin toss argument and our reply to it.

Imagine an urn containing a countably infinite collection of tickets and a mechanism to implement a fair lottery on the tickets in the urn.

In situation (1), all tickets are in the urn and we denote the probability of winning of each arbitrary single ticket in such a lottery as $\text{Prob}(E_1)$, leaving open the possibility that this may be an infinitesimal.

In situation (2), one ticket is removed from the urn prior to the drawing of the winning ticket. There is one competing ticket less, so the probability of winning of each remaining ticket is $\text{Prob}(E_2) = \frac{1}{1-\text{Prob}(E_1)}\text{Prob}(E_1)$ (renormalization). Taken in isolation, however, situation (2) looks exactly as before the removal of a ticket, which is situation (1). Because of this isomorphism between situation (1) and situation (2), we find that the probability of winning of each individual ticket is equal to $\text{Prob}(E_2) = \text{Prob}(E_1)$.

We have thus arrived at the following equations:

$$\text{Prob}(E_2) = \frac{1}{1 - \text{Prob}(E_1)}\text{Prob}(E_1),$$

$$\text{Prob}(E_2) = \text{Prob}(E_1).$$

Even in a non-Archimedean field, these equalities can only hold simultaneously if $\text{Prob}(E_1) = \text{Prob}(E_2) = 0$; it cannot be the case that $\text{Prob}(E_1)$ or $\text{Prob}(E_2)$ is a non-zero infinitesimal.

4.2.1 Reply: Changing the sample space mid-game

For standard probability functions, the range is fixed to be the unit interval of $\mathbb{R}$. Nevertheless, changing the sample space mid-game is, in general, not allowed, because the actual probability assignments still depend on the sample space. For NAP functions, the dependence on the sample space is more pronounced than for real-valued functions, because (the collection of finite subsets of) the sample space is used explicitly to construct the hyperreal field on which the function takes its values (see the Appendix for details).[22] Moreover, even if the sample space is kept fixed, the way in which the event of interest is embedded in the event space may influence the probabilities that are assigned to events. This is clear in the uniform case (in which NAP coincides with a non-Archimedean measure of relative sparseness, or normalized numerosity), but the issue generalizes to the non-uniform case.

---

[22] We are grateful to Marcus Pivato to encourage us to make this aspect more explicit.

These observations can be related to what has earlier been discussed as a re-labelling paradox (Bartha and Hitchcock [1999], Section 5). However, it is not the labelling itself that is essential,[23] but rather the choice of sample space and the embedding of events therein, which requires a form of holism in the assignment of probabilities that is captured by our demand for perfect additivity (see also Hofweber [2014]).

We can construct an NAP function, P, that describes a fair lottery on $\mathbb{N}$ in a regular way. This function will assign an infinitesimal probability to each singleton, namely, $P(\{n\}) = 1/\alpha$ for all $n \in \mathbb{N}$ (with α is the numerosity of $\mathbb{N}$). The function that we have thereby constructed crucially co-depends on the choice of the sample space (in this case $\mathbb{N}$). In particular, the non-Archimedean field on which the probability function takes it values depends on this choice.

Given some countable collection of tickets, situation (1) is a fair lottery on all of the tickets in this collection. The original argument only requires this collection to be countably infinite, without further specification. Hence, we need to fix a choice for the sample space before we can apply NAP to this scenario. In model A, we choose $\mathbb{N}$ to play the role of the sample space Ω_A of the probability function P. We use $\mathrm{Prob}_A(E_1)$ as shorthand for the probability of winning of an arbitrary single ticket in situation (1) on model A. On model A, event E_1 is represented by singleton $S_{E_1} = \{n\}$ (for some n) of Ω_A. Clearly, $\mathrm{Prob}_A(E_1)$ is equal to $P(\{n\}) = 1/\alpha$ for any $n \in \mathbb{N}$.

Situation (2) is a fair lottery on all but one of the initial collection of tickets (call this i). We use $\mathrm{Prob}_A(E_2)$ as shorthand for the probability of winning of an arbitrary single ticket in situation (2) on model A. Since we are using the same sample space as in the previous step, we can use the same probability function, P, and find $\mathrm{Prob}_A(E_2)$ via conditionalization:

$$\mathrm{Prob}_A(E_2) = P(\{n\}|\mathbb{N}\backslash\{i\})$$

$$= \frac{1}{\alpha - 1} \text{ (assuming } n \in \mathbb{N}\backslash\{i\}; 0 \text{ otherwise).}$$

This entails that

$$\mathrm{Prob}_A(E_2) = \frac{1}{1 - \mathrm{Prob}_A(E_1)} \mathrm{Prob}_A(E_1).$$

Since in situation (2) one ticket is not playing any role, and we are still faced with a countably infinite collection of tickets, we may consider representing the remaining tickets by $\mathbb{N}$ instead of by $\mathbb{N}\backslash\{i\}$. This is fine too, but we should realize that we can only do this by changing the sample space: we are now switching from model A to a new model B. In model B, we use the same probability function on the same sample space as in model A, but now

---

[23] We are grateful to Thomas Hofweber for pressing us on this point.

there is a different correspondence between sets in the event space and situations in the (hypothetical) world. In model A, we express the probability of a single ticket from the whole collection of tickets as $P(\{n\}) = 1/\alpha$, whereas in model B, $P(\{n\}) = 1/\alpha$ is used to express the probability of a single ticket from all but one of the initial collection of tickets.

The observation that E_2 can be described with the same labels as E_1 does not show that $\text{Prob}(E_1) = \text{Prob}(E_2)$ (as was claimed in the initial presentation of the objection), but only that $\text{Prob}_B(E_2) = \text{Prob}_A(E_1)$.

In sum, we have:

$$\text{Prob}_A(E_2) = \frac{1}{1 - \text{Prob}_A(E_1)}\text{Prob}_A(E_1),$$

$$\text{Prob}_B(E_2) = \text{Prob}_A(E_1).$$

This is insufficient to rule out the possibility that these probabilities might be infinitesimals. For this to follow, it would need to be the case that $\text{Prob}_A(E_2) = \text{Prob}_B(E_2)$. The fact that model A and model B can both be used to model the same situation, namely, situation (2), does not force this. Using the initial notation, however, $\text{Prob}_A(E_2) = \text{Prob}_B(E_2)$ would be glossed as $\text{Prob}(E_2) = \text{Prob}(E_2)$, making it impossible to tell them apart. To model the situation both before and after the removal of a ticket from the urn, we need a model like model A. From the viewpoint of NAP, a 'fair and countable lottery' is a highly underdetermined specification of a probability function (see also Wenmackers and Horsten [2013]).

4.3 Williamson's infinite sequence of coin tosses

Williamson ([2007]) has proposed an argument that purports to show that infinitesimals cannot be used to describe the probability of a fully specific outcome (for example, 'all heads') of a countably infinite sequences of tosses with a fair coin (endorsed in Hájek [unpublished]).

Williamson considers two infinite sequences of fair and independent coin tosses that all land heads: $H(1\ldots)$ and $H(2\ldots)$. $H(1\ldots)$ is an ω-sequence of coin tosses that all land heads. $H(2\ldots)$ is the subsequence of $H(1\ldots)$, which consists of the second toss of $H(1\ldots)$ and all the coin tosses that follow it. Williamson argues that

$$\text{Prob}(H(1\ldots)) = \frac{1}{2}\text{Prob}(H(2\ldots)),$$

$$\text{Prob}(H(1\ldots)) = \text{Prob}(H(2\ldots)).$$

The assertion $\text{Prob}(H(1\ldots)) = \frac{1}{2}\text{Prob}(H(2\ldots))$ follows from the fairness and independence of the coin tosses, together with the finite additivity property.

The assertion $\text{Prob}(H(1\ldots)) = \text{Prob}(H(2\ldots))$ is motivated by a symmetry consideration: as physical processes, $\text{Prob}(H(1\ldots))$ and $\text{Prob}(H(2\ldots))$ are isomorphic.

But even in non-Archimedean fields these two equalities can hold simultaneously only if $\text{Prob}(H(1\ldots)) = \text{Prob}(H(2\ldots)) = 0$; it cannot be the case that $\text{Prob}(H(1\ldots))$ or $\text{Prob}(H(2\ldots))$ is a non-zero infinitesimal. The conclusion that $\text{Prob}(H(1\ldots)) = \text{Prob}(H(2\ldots)) = 0$ is of course exactly what classical probability theory tells us it has to be.

4.3.1 Reply: The embedding of events in a sample space

Williamson's argument turns on the claim that the sequences $H(1\ldots)$ and $H(2\ldots)$ are identical in all relevant aspects (they are 'isomorphic'). This, however, is challenged by Weintraub ([2008]). She claims that the fact that $H(2\ldots)$ is a proper subset of $H(1\ldots)$ is significant.

Weintraub's reply is correct as far as it goes, but it seems incomplete. A further point is that, as discussed in Section 4.2, the assignment of probabilities does not make sense in the absence of a well-defined sample space that is applied in a consistent way. In the case of Williamson's argument, a crucial aspect of fixing the sample space is an answer to the question, 'When does the count of events start?'.

Let $2^{\mathbb{N}}$ be the sample space of model A, which reflects that the count of events starts at the first toss of $H(1\ldots)$. Let $C = 2^{\mathbb{N}\setminus\{1\}}$ be the conditioning event, which reflects that the count of events starts at the first toss of $H(2\ldots)$. In model B, the sample space is also $2^{\mathbb{N}}$. Although it is the same set as in model A, this set is now used in a different way, namely, to reflect that the count of events starts at the first toss of $H(2\ldots)$.

We again introduce some shorthand notations. In situation T1, a coin is tossed on all of some countably infinite collection of occasions. We use $\text{Prob}_A(H(1\ldots))$ as shorthand for the probability that such a coin comes up heads each time on model A. In situation T2, a coin is tossed on all but one (the first) of the countably infinite collection of occasions. We use $\text{Prob}_A (H(2\ldots))$ as shorthand for the probability that such a coin comes up heads each of the remaining occasions on model A and $\text{Prob}_B(H(2\ldots))$ for the corresponding probability on model B.

Williamson exploits the intuition that $\text{Prob}_A(H(1\ldots)) = \text{Prob}_B(H(2\ldots))$. But he glosses this as $\text{Prob}(H(1\ldots)) = \text{Prob}(H(2\ldots))$, thus turning the probabilities involved into evaluations within the same model. On the other hand, Williamson convincingly argues that $\text{Prob}(H(1\ldots)) = \frac{1}{2}\text{Prob}(H(2\ldots))$. Although we would rather represent this as $\text{Prob}_A(H(1\ldots)) = \frac{1}{2}\text{Prob}_A(H(2\ldots))$, leaving out the choice of model here is not as harmful as before, since we are now comparing probabilities within the same model, in

the sense that the sample space $2^{\mathbb{N}}$ is used in the same way. The two glosses indeed contradict each other unless $\text{Prob}(H(1\ldots)) = \text{Prob}(H(2\ldots)) = 0$. But the contradiction can only be obtained when the difference between the sample spaces is glossed over. In particular, there is no uniform application of an NAP model, M, such that $\text{Prob}_M(H(1\ldots)) = \text{Prob}_M(H(2\ldots))$ can be obtained.

At this point, it may be asked what the 'correct' sample space for evaluating the probability of $H(1\ldots)$ and of $H(2\ldots)$ is. In this article, we do not commit ourselves to there being a single correct sample space for evaluating probabilities associated with idealized scenarios such as lotteries on infinite spaces (see Section 6.1.1). But if what Williamson is envisaging is a possible physical universe consisting of an ω-sequence of coin tosses starting with toss number 1 (rather than with toss number 2), and if we regard the probabilities of the two sequences as objective probabilities, then evidently model A is the correct setting for evaluating both $\text{Prob}(H(1\ldots))$ and $\text{Prob}(H(2\ldots))$.[24]

4.4 Point sets on a circle

Williamson's argument crucially turns on translation symmetry.[25] Other symmetry considerations can be invoked to arrive at the same conclusion in other examples with a similar structure. Parker ([2013]) has given one such argument that turns on rotation symmetry; see also (Bernstein and Wattenberg [1969]; Barrett [2010]).

Consider the unit circle. Select the point on the circle with coordinates $(1, 0)$, and let it be called p_1. Now move an arc length 1 clockwise along the circle from this point; call this point on the circle p_2. Again move arc length 1 clockwise along the circle to obtain p_2. Continuing in this way, and taking account of the irrationality of the length of the unit circle, we obtain an ω-sequence $p_1, p_2, p_3, \ldots$ of points on the unit circle. Now abstract from the ordering to obtain set $\{p_1, p_2, p_3, \ldots\}$ of points on the circle; call this set S_1. Rotating the ω-sequence $p_1, p_2, p_3, \ldots$ by arc length 1 yields the set $\{p_2, p_3, p_4 \ldots\}$; call this set S_2. Now consider the probabilities, $\text{Prob}(S_1)$ and $\text{Prob}(S_2)$, of a point on the circle being in S_1 or S_2, respectively. Invariance of probability under rotation symmetry suggests that $\text{Prob}(S_2) = \text{Prob}(S_1)$. But if, in addition, the probability of a point being identical with p_i is equal for each $i \in \mathbb{N}$ (uniformity), then since $\text{Prob}(S_1) = \text{Prob}(S_2) + \text{Prob}(\{p_1\})$, p_i must

---

[24] Williamson ([2007]) also considers a second sequence of tosses with a separate coin, which is tossed at the same points in time as the first coin, except that the second coin's first toss occurs at the first coin's second toss. Even without analysing it in full, it ought to be clear that considering two coins allows for even more freedom in choosing the sample space and embedding the events in it. So, we agree with Hofweber ([2014]) and Easwaran ([2014]) that the probability associated with the second sequence resulting in all heads does not need to equal that of $H(2\ldots)$.

[25] In particular, on a temporal translation symmetry. It could be turned into a spatial translation symmetry by considering a countably infinite row of coins that are tossed simultaneously.

be zero for each $i \in \mathbb{N}$. In particular, point events cannot be assigned non-zero infinitesimal probability values. Again we have an *a priori*, conceptual argument for the conclusion that classical probability theory models fair lotteries on uncountable sets correctly.

It will be clear to the reader by now that our diagnosis of the argument from rotational symmetry against infinitesimal probabilities is structurally identical to our diagnosis of Williamson's argument. Hence, we do not describe it in detail here.

4.5 Easwaran and Pruss

Easwaran ([2014]) has proposed a conceptually new argument to the effect that infinitesimals cannot be used to describe the fair lottery on the Cantor space $2^{\mathbb{N}}$. He aims to show that for every infinitesimal ϵ we have that $\text{Prob}(H(1\ldots)) > \epsilon$, where $H(1\ldots)$ is again Williamson's infinite sequence of coin tosses. If this is so, then $\text{Prob}(H(1\ldots))$ can indeed not be an infinitesimal. The subsequence $H(2\ldots)$ does not play a role in Easwaran's argument. Instead, he considers $H(1\ldots N)$: a sequence of heads of non-standard length, where N is an infinite hypernatural number of heads.

His argument goes as follows: Consider a standard infinite sequence of coin tosses of all heads, $H(1\ldots)$. Now take any infinitesimal, ϵ. Then $\frac{1}{\epsilon} > n$ for any $n \in \mathbb{N}$. Now take a non-standard integer power 2^N such that

$$2^N < \frac{1}{\epsilon} \leq 2(2^N).$$

Such a non-standard integer power must exist,[26] and this number, N, must then be an infinite hypernatural number. Now consider the probability $\text{Prob}(H(1\ldots N))$ of the non-standard (hyperfinite) sequence of N heads. Then we have

$$\text{Prob}(H(1\ldots)) \geq \text{Prob}(H(1\ldots N)) = \frac{1}{2^N} > \epsilon,$$

which yields the desired result. Easwaran's conclusion is not that the probabilities involved should all be zero, but rather that it is illegitimate to compare probabilities on standard sample spaces to probabilities on non-standard sample spaces.

---

[26] By the 'transfer principle' for NSA. This principle says that if a first-order property holds of the standard real numbers, then it also holds of the non-standard reals. In the case under consideration, the property in question is

$$y > 1 \rightarrow \exists n \in \mathbb{N} : 2^n < y < 2^{n+1}.$$

Easwaran's argument is closely related to the following recent argument against infinitesimals put forward by Pruss ([2014], Section 3): Suppose that Prob_1 is a probability function on $\mathbb{N}$ that assigns an infinitesimal $\delta > 0$ to every singleton and represents a countably infinite fair lottery. Choose a hypernatural infinite number M so that $\frac{1}{\delta}$ is much larger than M. Now consider an infinite lottery on the infinite set $\{1, \ldots, M\}$. While in the case of a fair lottery on $\mathbb{N}$ it is perhaps not clear which infinitesimal should represent the probability of a singleton, there is an obvious answer as to what that infinitesimal probability here should be: $\text{Prob}_2(\{n\}) = \frac{1}{M}$ for all $n \in \{1, \ldots, M\}$. But we know that $\mathbb{N} \subsetneq \{1, \ldots, M\}$. So we are in a situation where the probability of drawing the winning ticket in a fair lottery with more tickets is higher than that of drawing the winning ticket in a lottery with fewer tickets. This is patently unreasonable.

Before presenting the former argument, Pruss offers an argument that he calls an 'intuition pump'. He compares an assignment of (not necessarily equal) infinitesimal probability values to each singleton of $\mathbb{N}$ to the countably additive function that assigns probability $1/2^n$ to the singleton n for each $n \in \mathbb{N}$.[27] Pruss observes that, in the former case, the probability assigned to each singleton is infinitely smaller than the probability assigned to it in the latter case. By an analogy to the former case, he argues that this cannot be correct.

4.5.1 Reply: Internal versus external probabilities

The crucial move in Easwaran's argument was the introduction of a probability assigned to a non-standard sequence N of heads; he assumes this probability has to be $\frac{1}{2^N}$. This is not true in NAP, which will assign a strictly smaller probability value to this event. Likewise, the crucial move in Pruss's argument was the introduction of a probability assigned to a particular outcome of a lottery on infinite hyperfinite set $\{1, \ldots, M\}$; he assumes this probability has to be $\frac{1}{M}$. Again, NAP assigns a strictly smaller probability value to this event.

To argue for these claims, we first make two statements that are indisputably correct, then try to combine them into a contradiction, and show why this fails. For a fair lottery on $\mathbb{N}$, NAP assigns $\frac{1}{\alpha}$ as the probability of a particular singleton outcome. For a fair lottery on the infinite hyperfinite set $\{1, \ldots, N\}$, an internal probability theory assigns $\frac{1}{N}$ as the probability of a

---

[27] Observe that the countably additive function $1/2^n$ is not an NAP function. In particular, it is not normalized: the NAP sum yields $1 - 1/2^\alpha$ rather than 1. For the relevant NAP function, it holds $\forall n \in \mathbb{N}$ that

$$P(\{n\}) = \frac{1}{2^n} \times \frac{2^\alpha}{2^\alpha - 1},$$

which differs pointwise by an infinitesimal from the function Pruss considers.

particular outcome.[28] Now it may seem like we can arrive at a contradiction: $\mathbb{N} \subsetneq \{1, \ldots, N\}$ for any infinite N. Yet, by choosing $N \lneqq \alpha$, 'the probability' of a singleton event is smaller on a strictly smaller sample space ($\frac{1}{\alpha} \lneqq \frac{1}{N}$).

The contradiction is only apparent, however, since we are mixing assignments across probability theories—that is why we put 'the probability' between quotation marks: different theories may assign probability values differently. If we want to compare probability values in a meaningful way, we have to do this in a context that allows us to assign probabilities to all events of interest.

In order to explain this, we have to introduce a technical distinction familiar from non-standard model theory: it is the distinction between internal and external objects in the non-standard universe. 'Internal' refers to information available from within the non-standard model using the transfer principle, which only applies to first-order properties. 'External' refers to information available in the non-standard universe in which the non-standard model is embedded—concerning properties that cannot be obtained by transfer alone. Internal approaches to probability theory do not assign probabilities to lotteries on $\mathbb{N}$ (or on any infinite standard domain), since these sets are external objects in the non-standard model. Hence, such theories do not allow us to compare a lottery on $\mathbb{N}$ to a hyperfinite lottery at all. In such a context, these sets are incommensurable, and there is no contradiction.

In contrast, NAP is an external approach to probability theory. It has been claimed that external probability functions cannot assign probabilities to hyperfinite lotteries (Easwaran [2014], p. 27). However, NAP functions can do this; it requires a second iteration of the range-building procedure and the results are different from those of an internal approach. Since this is an important point, we elaborate on it. Although we originally intended NAP theory to model probability functions on standard domains (cf. Table 1), our formalism is perfectly general. Its machinery can be used to construct a uniform probability distribution on an infinite hyperfinite set $\{1, \ldots, N\}$. Moreover, since NAP yields total probability functions, the NAP function on $\{1, \ldots, N\}$ will also assign a probability to the sub-event $\mathbb{N}$: a fundamental reason for taking NAP to be preferable over Nelson's theory. The NAP value associated with a singleton event on $\{1, \ldots, N\}$ will not be equal to $\frac{1}{N}$. Instead, it will be a strictly smaller infinitesimal. This ought to be clear immediately, since NAP functions respect the Euclidean principle. For instance, for the probability of ticket 1 winning in a fair lottery on $\{1, \ldots, N\}$ we have:

$$P(1|\{1, \ldots, N\}) < P(1|\mathbb{N}) = \frac{1}{\alpha}.$$

So, in NAP the events are comparable, yet there is no contradiction either.

---

[28] This can be obtained directly by considering fair lotteries on finite sample spaces and by applying the transfer principle from NSA to it; see, in particular, Nelson's ([1987]) probability theory.

Another way to see this is via an analogy with set size: the numerosity assigned to $\mathbb{N}$ is α and the internal cardinality of $\{1, \ldots, N\}$ is N, which may be chosen smaller than α. Yet, this does not suffice to conclude that 'the size' of $\mathbb{N}$ is strictly smaller than that of a strict superset. To compare sizes, we have to pick a particular theory for assigning sizes. $\mathbb{N}$ has no internal cardinality, so this measure is not useful for comparing its size with a hyperfinite set. However, $\mathbb{N}$ and $\{1, \ldots, N\}$ do have an external cardinality and a numerosity and for $\{1, \ldots, N\}$ both of these measures are (much) larger than N.[29]

In short, the way in which NAP describes a (fair) lottery on a hyperfinite set is fundamentally different from the way in which Nelson's theory describes it. In an internal theory, we perform one ultrafilter construction and within the resulting ultrapower model find a non-standard number, N, and the probability values associated with a fair lottery on $\{1, \ldots, N\}$. In the NAP description, however, we need two ultrafilter constructions to find an NAP function that describes a fair lottery on a hyperfinite set. In particular, if we already have an NAP function that describes a fair lottery on $\mathbb{N}$, in general the probability values required to describe a fair lottery on $\mathbb{N}^*$ or infinite hyperfinite subsets thereof will not yet be in the range of this function. A second NAP construction is required to obtain the required range.

Regarding Pruss's intuition pump, in this example all finite initial segments of sums of probabilities of singletons are such that the former is infinitely smaller than the latter. The intuition pump only works if we assume that this implies that the same holds for the infinite sum. However, in the case of NAP and the corresponding non-Archimedean limit of the sum, the implication does not hold. The key observation is that in order to determine which sum is larger, one has to compare the summands (as Pruss does) as well as the sum operation. The standard infinite sum, appropriate for adding CA probability values, is fundamentally different to the non-Archimedean sum (Section 3.4) that is appropriate for summing NAP values.

5 Dividends

Now that the objections against non-Archimedean probabilities have been addressed, we turn to the advantages of using non-Archimedean probabilities. We describe problems that classical probability theory cannot model in a satisfactory manner, but which can be modelled elegantly by NAP.

---

[29] N is the internal cardinality of $\{1, \ldots, N\}$, which means that there is no internal bijection of $\{1, \ldots, N\}$ onto a smaller initial segment of the hypernatural numbers. However, the external cardinality of $\{1, \ldots, N\}$ is much bigger than N: there exist no external bijection of $\mathbb{N}$ onto $\{1, \ldots, N\}$, so the set $\{1, \ldots, N\}$ has uncountably many elements. For the distinction between internal and external cardinality, see, for instance (Albeverio *et al.* [1986], p. 67).

5.1 Measure and utility

According to classical probability theory, uncountable sample spaces contain non-measurable subsets (Truss [1997], Chapter 11, Section 4). The probability functions that are produced by NAP are 'total': they assign probabilities to all subsets of the sample space, which can be finite, countably infinite, or uncountable. This is a virtue of the account, as non-measurable sets are widely regarded as 'pathologies' by probability theorists.

If NAP is adopted, then utilities can be calculated in the usual way even for events that are judged to be non-measurable by classical probability theory. Contingent events that have measure zero on the classical theory can have a non-zero probability if NAP is adopted. Thus NAP seems to provide a suitable background theory for a utility theory for infinite spaces.

Mixing hyperreal probability assignments with standard utility theory may lead to sub-optimal results. A non-Archimedean utility theory for infinite outcome spaces has been worked out by Pivato ([2014]) and by Pedersen ([unpublished]).[30] In the resulting theory, utilities do not satisfy the Archimedean principle. As Pivato himself notes, this is not a defect of the theory since many utility theorists are wary of this principle as a constraint on utilities (see, for example, Krantz *et al.* [1971], Sections 1.5.2, 6.5.1, and 9.1).

5.2 Regularity and uniformity

Lewis ([1980]) has argued that only impossible events must be assigned probability zero, that is, probability functions should be 'regular'. The main reason for this requirement is that probability functions ought to be maximally fine-grained; they are expected to distinguish between impossibility and infinitely improbable contingency. Lewis's stance harmonizes with our preferred non-standard probability theory. We have seen that NAP gives us ways of building regular probability functions even for infinite sample spaces. Indeed, NAP functions satisfy the Euclidean principle. The fineness of the grain is always sufficient for the problem at hand, since the range of the NAP function is constructed using the relevant domain.

It is sometimes held that for reasons of symmetry, certain uniformity assumptions should also be imposed. This lies behind versions of Laplace's ([1902]) principle of insufficient reason, later called the 'principle of indifference'. So one might require that in the absence of evidence to the contrary, all atomic propositions should be given the same probability value. It is well known that the fact that sample spaces can be carved up in different ways spells trouble for the principle of indifference (see van Fraassen [1989],

---

[30] As mentioned before, Pivato's construction of non-standard utility functions for infinite sets of outcomes is very similar to the construction of NAP models.

Chapter 12). Nonetheless, uniform probability distributions on infinite spaces, and even regular uniform distributions, certainly seem conceptually possible. So our probability theory should at least allow for them; this is captured in our demand of weak Laplacianism that was discussed in the introduction, and which is satisfied by NAP.

In sum, even if we agree with those philosophers of probability who argue that the principle of indifference involves an illegitimate inference from ignorance to knowledge, the regular uniform probability distributions on infinite sample spaces should fall within the scope of the mathematical treatment of probability.

5.3 Credence and chance

Perhaps the best known principle relating subjective probability and chance is Lewis's ([1980]) principal principle, which can be roughly stated as follows:

$$\text{Prob}(A|Ch(A) = x) = x,$$

where 'Prob' is subjective probability, Ch is a chance measure, and x is a real number.

In classical probability theory, it would seem that if A represents the value of a continuous observable (say, a position measurement for an electron in a superposition state), $Ch(A)$ will be zero in a non-determinist context for every value A. Hence, according to Lewis's principal principle $\text{Prob}(A|Ch(A) = 0) = 0$. This will render any probability conditional on the posterior $\text{Prob}(A)$ undefined, essentially leading to the problems flagged in Section 2.2. NAP can guarantee regularity, so non-zero infinitesimal values can unproblematically be assigned to $Ch(A)$ in such cases. Indeed, Lewis himself advocated the use of infinitesimal probabilities both for subjective credences and for objective chances for precisely such reasons.

In addition, NAP contains the resources for resolving the so-called zero-fit problem for classical probability, which goes as follows: Suppose that the actual world is a 'Williamson-world'. It consists entirely of an ω-sequence, A, of coin tosses (not necessarily fair). And suppose that the world is chancy. In particular, it is governed by a law that states that the limiting relative frequency of heads is $\frac{1}{10}$.[31] Then we can define a notion of 'goodness of fit'

---

[31] This example is taken from (Elga [2004]). Observe that an NAP function describing a coin with a fixed bias is not the same as an NAP function that expresses a law concerning its limiting frequency. The prior NAP function is regular on $2^{\mathbb{N}}$ and assigns some non-zero probability to sequences with a limiting frequency unequal to the bias (such as a sequence of all heads produced by a fair coin). By conditionalizing on the relevant hypothesis, a posterior NAP function can assign probability zero to all sequences in $2^{\mathbb{N}}$ that do not have the required limiting frequency.

of a hypothesis with respect to the actual world. We say that hypothesis T_1 has better fit than T_2 if and only if $Ch(A|T_1) > Ch(A|T_2)$.

But this presents a problem for classical probability. Let T_1 be the 'right' theory: it says that the limiting frequency of actual coin tosses is $\frac{1}{10}$. And let T_2 be a theory that fits less well: it says that the limiting frequency lies in the interval $[\frac{1}{20}, \frac{1}{2}]$. Kolmogorov's principles require that $Ch(X|T_1)$ must be zero for almost all ω-sequences X that satisfy T_1 (since there are continuum many such). Indeed, if there are no further laws governing the actual world beside T_1, then $Ch(X|T_1)$ will have to be zero for all ω-sequences X that satisfy T_1. But that means that classical probability predicts that the fit of T_1 is no better than that of T_2, which is incorrect.

Again, NAP can be used to generate chance functions that assign non-zero (but infinitesimal) chances only to ω-sequences X that satisfy T_1. Then, by the Euclidean principle (satisfied by NAP), $Ch(A|T_1) > Ch(A|T_2)$, which is the right outcome.

Lewis's principal principle assumes that for each moment in time there exists a unique objective physical chance function that governs physical events in the actual world. In the present article we do not wish to commit ourselves to this assumption. Indeed, whereas this assumption is compatible with everything that is claimed in this article, it is not a view that is congenial to the spirit of NAP. We will return to this issue below (see Section 6.1.1).

5.4 Conditional probability

In Section 2, we discussed three problems due to the assignment of probability zero to contingent events in the (semi-)classical approach and mentioned that they can also be addressed without the use of non-Archimedean probabilities. One of these approaches consists in regarding conditional probabilities as the fundamental notion; Popper functions are one way of realizing this idea. Another approach is to consider sequences of probability functions: lexicographical probabilities. In this section, we explore the relations between NAP and these alternatives (although we will not be able to do justice to their history here).

NAP is based on axioms phrased in terms of unconditional (or absolute) probability functions, just as Kolmogorov's axioms. Nevertheless, conditional (or relative) probabilities play a central role in the construction of unconditional NAP functions, just as they do in the classical theory (recall the CPP in Section 3.2). So, the mathematics of NAP harmonizes with the philosophical observation that the notion of conditional probability is at least as fundamental as that of unconditional probability (see, for example, Hájek [2003]). Hence, it should not come as a surprise that deep connections exist between

NAP functions and axiomatizations phrased in terms of conditional probability functions.

Even if conditional probabilities with impossible antecedents are meaningless, conditional probabilities involving infinitely unlikely contingencies should be well defined. As mentioned before, in the fair lottery on $\mathbb{N}$ or $\mathbb{R}$, for instance, it seems that we ought to be able to say that Prob($\{1\}|\{1, 2, 3\}$) $= \frac{1}{3}$. On the (semi-)classical description of infinite lotteries, all such conditional probabilities are undefined. Precisely for this reason, Popper functions have been introduced.

A Popper function is a (non-classical) conditional probability function $C(A, B)$ that is defined for all $A, B \in \mathcal{P}(\Omega)$ (where Ω is a finite or infinite sample space). Popper functions take their values in the real interval $[0, 1]$, just like classical probability functions. If $E \neq \emptyset$, then $C(\cdot, E)$ is required to be a classical probability function.[32] The conditional probability $C(\cdot, \emptyset)$ is defined as one. This is an arbitrary choice; it reflects the fact that we do not care what value is assigned to events conditional on an impossible event.[33] A notion of unconditional probability can be defined in terms of a given Popper function as follows: Prob(A) $= C(A, \Omega)$, where Ω is the sample space. The crucial point is that Popper functions impose restrictions on $C(\cdot, E)$ even if E is an event that has unconditional probability zero. This is where Popper functions differ from classical probability functions. In the example concerning the lottery of the natural numbers above, a description in terms of Popper functions will indeed predict that $C(\{1\}, \{1, 2, 3\}) = \frac{1}{3}$. Thus Popper functions generate an interesting account of conditional probabilities.

NAP can be seen as a generalization of classical probability. It is not hard to see that the following representation theorem holds (Benci *et al.* [2013]):

Theorem 4

(1) For every classical probability function, P_K, there exists an NAP function P that is pointwise infinitely close to it, that is, for every E such that $P_K(E)$ is defined, $|P_K(E) - P(E)| < r$ for every $r > 0 \in \mathbb{R}$.

(2) For every NAP function, P, there is a classical probability function, P_K, that is infinitely close to it for every event on which the latter is defined.

There exists a representation theorem relating regular non-standard probability functions that only satisfy finite additivity (and not necessarily NAP's

---

[32] So, in that case we require $C(\cdot, E)$ to satisfy σ-additivity. Totality and σ-additivity are not imposed by all authors and in general cannot be satisfied jointly.

[33] For a complete list of the axioms governing Popper functions and a discussion of it, see (McGee [1994]). It may be more natural to leave $C(\cdot, \emptyset)$ undefined; see, for instance, (Easwaran [2014], p. 16).

infinite additivity property) on the one hand, and Popper functions on the other hand (see Krauss [1968]; McGee [1994]). However, Brickhill and Horsten ([unpublished]) has recently shown that this result can be strengthened to a representation theorem that relates regular (and perfectly additive) NAP functions with Popper functions[34]:

Theorem 5

    (1) For every Popper function, $C(\cdot, \cdot)$, there exists an NAP function, P, that is pointwise infinitely close to it.

    (2) For every NAP function, P, there is a Popper function, $C(\cdot, \cdot)$, that is pointwise infinitely close to it.

So, NAP functions can be regarded as generalizations of Popper functions. However, since even infinitesimal differences may change the order of NAP values, corresponding Popper functions and NAP functions may lead to different decisions (see the following section).

Popper functions are related by means of representation theorems to classes of classical probability functions.[35] So, indirectly, the representation theorems of Krauss, McGee, and Brickhill connect non-standard probabilities to classes of classical probability functions.

Since the work of Adams, classical probability functions play a central role in the theory of indicative and subjunctive conditionals. But it is well known that conditionals with contingent probability zero antecedents cannot be treated in a satisfactory manner.[36] Popper functions have been used to construct better theories of conditional sentences (Leitgeb ([2012]). Given the connection between Popper functions and NAP functions, it is clear that there is an important role to be played here for non-standard probability theories, too. Popper functions have in turn been related to possible worlds semantics for counterfactuals (Leitgeb [2012]), Part A, Section 3). So, again, there is a deep relation between possible worlds semantics for conditionals and NAP models.

Conditional probabilities play a prominent role in learning from evidence. They figure crucially in standard update rules. In classical probability theory, we cannot use Bayes's rule, for instance, to learn from infinitely improbable contingencies that actually obtain. They cannot be used to update our subjective probability distribution in such situations. But if our subjective

---

[34] The proof of the second part of this theorem is easy; the proof of the first part is much more complicated.

[35] See (van Fraasen [1976]), which associates finite dimensions with Popper functions. For NAP functions, van Fraasen's notion of dimension can be extended into the transfinite. It would take us too far afield to pursue this in the present article. See also (Császár [1955]); Rényi [1956]).

[36] Such as, 'If in the lottery on the natural numbers ticket 3 is (was) drawn, then one of the tickets 1, 2, 3 is (was) drawn'.

probability distribution is regular, then Bayes's rule can be used to revise our probability distribution.[37]

Halpern ([2010]) has carried out a systematic investigation of the relationships between non-standard probability spaces, conditional probability spaces (including Popper spaces), and lexicographical probability spaces. However, his results do not suffice to show how NAP spaces relate to the two other approaches, since the non-standard probability spaces considered by Halpern lack the perfect additivity property that is encoded by axiom NAP4 (which was not formulated at the time). Instead, Halpern ([2010], p. 166) considered a pointwise limit on the hyperreals to obtain a proxy for countable additivity that applies to non-standard probabilities. We suspect that considering NAP spaces may lead to an interesting restriction on the class of lexicographical probability spaces with which they are equivalent.

6 General Considerations

In this final section, we consider misgivings that are somewhat more diffuse, and that relate to more general philosophical questions about the nature of probability and about the way in which probability relates to the real world. In particular, we address the non-uniqueness of NAP functions and revisit the failure of NAP to validate certain invariance principles. We will see that these issues are connected.

In the introduction, we motivated NAP on the basis of four desiderata (regularity, totality, perfect additivity, and weak Laplacianism). The existence of models for NAP theory shows that these four initial desiderata can be combined in a consistent way. We may consider further desiderata, but not all subsets of these desiderata can be combined harmoniously within a single, consistent probability theory. In particular, the desideratum of totality is in tension with that of uniqueness and the desideratum of regularity is in tension with that of invariance.[38]

6.1 Non-uniqueness

NAP functions are non-unique in the following sense: Given a conceptually possible probabilistic scenario, there may be uncountably many NAP functions that describe the scenario equally well. The functions differ from each other because their construction relies on a different free ultrafilter (see

---

[37] Pruss ([2012]) argues that updating on infinitesimal probabilities is coherent but can give counterintuitive results in certain situations. We reserve discussion of Pruss's objection for another occasion.

[38] It has also been observed by Skyrms ([1980], Appendix 4) that there is a trade-off between various requirements—he considered additivity, translation invariance, totality, and regularity—for standard and non-standard measures.

Appendix). Although not any free ultrafilter will do,[39] the relevant collection still contains uncountably many filters.[40]

Given a particular subset of the sample space, this plurality of ultrafilters may lead to variations in the associated NAP values. The difference between two NAP assignments to an event that does not receive a classical probability value may be non-infinitesimal (and, in some cases, as high as one minus an infinitesimal). Concerning a non-standard probability function for a fair lottery on the natural numbers from (Wenmackers and Horsten [2013]), Kremer ([2014]) has shown that there is a set that can be given any rational number between zero and one as (the standard part of) its probability value; this observation generalizes to NAP.

Offhand, this non-uniqueness seems undesirable. However, Kremer suggests that 'maybe this indeterminacy is a feature, not a bug', because intuitively it is not clear at all what probability should be assigned to the set that he constructs. In order to assess the issue of non-uniqueness, we first discuss all the parameters that need to be fixed to define an NAP space and how these choices relate to the notion of uniquely determined physical chance.

6.1.1 Parameters and objective probability

An NAP space is a triple $\langle \Omega, w, \mathcal{U}_{\mathcal{P}_{fin}(\Omega)} \rangle$, with Ω a sample space, w a real-valued weight function defined on the elements of Ω, and $\mathcal{U}_{\mathcal{P}_{fin}(\Omega)}$ an ultrafilter on the class of finite subsets of Ω (see Section 3.5). (We focus on situations where Ω is infinite.)

This means that in order to model a given conceptually possible probabilistic situation, three choices need to be made. The first two choices are familiar from classical probability theory. If one believes in the existence of objective probabilities, then these choices can be constrained: one can require Ω to be a subset of the 'universal sample space' of physically possible point events, and one can take the weight function to be physically determined.

In the classical setting, a third choice has to be made (in uncountably infinite sample spaces): one has to pick a σ-algebra of events. The defender of NAP also has a third choice to make: the choice of an ultrafilter. In both approaches, this third choice may involve arbitrariness. In the classical setting, when totality fails, we have to take some sets to be non-measurable. It is not easy to convince oneself that the subsets of Ω that are non-measurable have no probability (physical propensity) of occurring, but we leave that as a problem for the classical probabilist who is also a fan of objective probabilities. Of course, the classical probabilist is not forced to take this position; if she is a

---

[39] We need a fine ultrafilter (see Kanamori [1994], p. 301) and may impose further conditions on the filter (as discussed, for instance, in (Benci *et al.* [2013], Section 5.2).
[40] For an infinite set of cardinality κ, there are 2^{2^κ} ultrafilters to choose from.

subjectivist, then she will take the weight function to be an expression of a person's subjective expectations.

In NAP, the ultrafilter is defined on a 'directed set', Λ, which is included in, but does not have to be all of, $\mathcal{P}_{fin}(\Omega)$. In (Benci *et al.* [2013], Section 5), we have shown how choosing a smaller Λ (which also reduces the number of ultrafilters) influences the properties of the resulting NAP function. Nonetheless, it is hard to see how one could ever have conclusive grounds for preferring one particular ultrafilter over all the others (see also Wenmackers and Horsten [2013], Section 6.2). A fan of objective probability can probably still maintain that the choice of ultrafilter is, ultimately, physically determined. Nonetheless, given the empirical inaccessibility of this ultrafilter,[41] this position does not help us to select a unique NAP function in our models.

So on both the classical approach and in NAP, probability is partially arbitrary, in the sense that it involves a choice that is not empirically accessible. Once a choice has been made (for a particular σ-algebra of events or an ultrafilter, respectively), the probability function is unique (relative to this choice).[42] Even if it is assumed that there is a single true σ-algebra of events or ultrafilter, it is empirically inaccessible, so making a particular choice to represent this physical chance function remains partially arbitrary.

Thus, at most, we can hold that the probability of physical events is objective in a weaker sense. What one could say is the following: There is such a thing as physical chance. And it is a legitimate task of our mathematical models to track this property. But our models can only track physical chance in a mediated way. In order to describe a physical system and its behaviour, our probabilistic models have to select a sample space and label the point events (that is, establish a connection between reality and point events in the model). For finite sample spaces, the labelling does not matter; but for infinite sample spaces, different labellings can result in different probability assignments. All this induces a degree of relativity in probability values of events. But it in no way contradicts the objectivity of physical chance.

We advocate an even weaker position *vis-à-vis* the objectivity issue. As stated in the introduction, we see the task of probability as being one of mathematically modelling 'conceptually possible' probabilistic situations (weak Laplacianism). The resulting mathematical models should preserve as much of our pre-theoretical intuitions concerning probability as possible. Viewed in this light, the choice of an ultrafilter does not appear to be a

---

[41] We can only perform finitely many experiments. Hence, even if we would assume that there is such a thing as a particular ultrafilter in reality, there is no way to establish empirically which one it is.

[42] In the real-valued case, it is well known that we can trade this partial arbitrariness of the event space (due to the failure of totality) for the partial arbitrariness of measure values (due to failure of uniqueness) by introducing generalized limits (or Banach limits); see, for instance, (Wenmackers and Horsten [2013], Section 3.2).

problem, for there is no reason to assume that there is a unique best way to model certain infinite probabilistic situations (such as infinite lotteries).

6.1.2 Order and subjective probability

The regularity axiom NAP1 and the associated Euclidean principle require probability values to respect the ordering induced by the subset relation on the event space. The sum and product rule require numerical probabilities to have more structure than just this partial order: we require the probability values to be part of a totally ordered field of numbers. Hence the subset relation underdetermines the order of the probability values. We need an additional degree of freedom to allow for various kinds of probability assignments to the same event space.

In part, this is accomplished by selecting a weight function (Section 3.5), which encodes probability relations between atomic events, and thus the order between many disjoint events (in particular, for finite and co-finite events). But the subset relation together with a specification of a weight function still underdetermines the order of the probability values.[43] For some events, the difference in probability can be more than infinitesimal, depending on the properties of the free ultrafilter.[44] Yet, even infinitesimal differences may change the order (for example, in a fair lottery on the natural numbers, the probability of the set of even numbers may be equal to or infinitesimally smaller than that of the set of odd numbers).

The observation that the subset relation is a partial order whereas any sort of numbers (real or hyperreal) require a total order has been made by others, and recently by Hofweber ([2014]) and Easwaran ([2014]). Easwaran observes that real-valued probability functions leave out part of the structure of the partial order, whereas hyperreal-valued probability functions add structure in an arbitrary way. He prefers the former approach, since even if one assigns probabilities to an algebra, one can still consult the order of the subset relation on the algebra. Nothing prevents you from using both sources of information when you have to make decisions. For instance, if you get the opportunity to choose between betting on the occurrence of {1} and of {1, 2} in a lottery on $\mathbb{N}$ or $\mathbb{R}$, both real-valued probabilities are zero, but that does not prevent you from preferring the largest event.

---

[43] In the context of real-valued probabilities, this has prompted some researchers to strengthen the notion of 'uniformity' on a countably infinite sample space beyond the assignment of equal weights to atomic events. Kerkvliet and Meester ([2016]) assign uniquely determined, real-valued probabilities to a large collection of subsets of the sample space. Although their results may lead to interesting suggestions to explore in the context of NAP, we do not pursue those here.

[44] In the case of a fair lottery in $\mathbb{N}$, these are the sets that do not have an asymptotic density (see, for example, Wenmackers and Horsten [2013], Section 3.1). In general, a necessary (not sufficient) condition is for the event to be infinite and not co-finite.

One might think that ignoring relevant existing structure (a sin of omission) is not as grave as adding structure (a sin of misinformation).[45] However, it has to be borne in mind that one can always consider the entire family of NAP functions modelling a given situation, rather than an—arbitrary—representative of it (see also Wenmackers and Horsten [2013]). Such a family is the set of all NAP functions that meet a common specification, such as 'a fair lottery on ℝ', which fixed the sample space and the weight function, and possibly additional constraints on the directed set. As a whole, the family shows us how much the probabilities of a given event, and the order of probabilities of multiple events, can vary (dependent on the choice of ultrafilter).[46]

To put it differently,[47] there may be multiple, equally good ways to model the same situation, corresponding to different choices of the ultrafilter. What matters is what is true (or false) on all ways of making these arbitrary choices—what is supertrue (or superfalse)—as well as the spread of possible assignments. We need not project these arbitrary choices onto what is being modelled.

In the context of decision theory, a family of NAP functions leads to more subtle decisions. Let us consider a fair lottery on the natural numbers (starting at one) and suppose you are given a choice between betting on the occurrence of the set of even numbers and of the set of odd numbers. If you make decisions based on real-valued probabilities, you are indifferent between these two options. Since the choice concerns disjoint events, taking into account the subset relation on the event space (as Easwaran [2014] suggests) does not change this, either. By considering the family of NAP functions, however, you may reasonably favour the set of odd numbers: some NAP functions assign a higher probability to this set than to the even numbers, whereas the others assign equal probabilities to both events.[48] For some other subsets of the sample space, you do not have any information based on real-valued probability assignments. By considering the family of NAP functions, it turns out that some events that are non-measurable on the classical account have probabilities that vary between zero plus an infinitesimal and unity minus an infinitesimal; but others vary within a smaller interval, for instance, between one-third and two-thirds (plus minus an infinitesimal) (see Kremer [2014]; Kerkvliet and Meester [2016]). Given the choice between such a subset of ℕ, S, and the set ℕ mod 4, you do not need to know the exact hyperreal-

---

[45] But recall from the previous subsection that the classical approach may also be subject to this criticism: it adds structure by declaring which sets are measurable and which are not.

[46] This is familiar from the context of real-valued probabilities, where a family of probability functions is used to represent imprecise probabilities; see, for example, (Walley [2000]; Halpern [2003]).

[47] We are grateful to Alan Hájek for this suggestion.

[48] See also Halpern's ([2010], p. 167) example in which infinitesimal probabilities lead to decisions that cannot be captured by Popper functions.

valued probability of the former to know that it has a higher probability of winning than the latter (see also Wenmackers [unpublished]).

At the same time, a family of NAP functions is more definite since it specifies the relevant limit process. The Borel–Kolmogorov paradox (mentioned in Footnote 7) demonstrates that classical probability functions alone do not contain enough information to define all conditional probabilities; information on the limiting process has to be supplemented. In the context of NAP, specifying the limit process is reflected in a reduction of the family of free ultrafilters and the corresponding family of NAP functions.

6.1.3 Domain and co-domain

NAP functions in the same family do not even have the same domain and co-domain. Believers in objective probability hold that there is one ultimate universal sample space. But NAP does not want to be restricted to sub-domains of this sample space (if it exists at all). We also want to model situations that are outside the physical realm. It may (or may not) be the case, for instance, that the universal physical sample space is of the size of the continuum. Then we would still want NAP to be able to model lotteries on the function spaces on the real numbers, for instance.

Easwaran has objected to hyperreal credences using a complexity argument ([2014], Section 5.4). His conclusion, that physical agents cannot have hyperreal credences, relies on four premises, two of which 'might be controversial' ([2014], p. 29): 'Credences supervene on the physical' and 'All physical quantities can be entirely parametrized using the standard real numbers'. Bascelli *et al.* ([2014], p. 850) reject the second controversial premise, by referring to physical models that do employ hyperreal numbers. We want probability theory to be applicable to thought experiments, as well as to models in physics. Hence, we argue that a probability theory should not depend on current physical theories, nor on considerations about credences of actual agents (informed by such theories).

In classical probability theory, the interval $[0, 1] \subset \mathbb{R}$ serves as the value range of all probability functions, even those that have $\mathbb{R}$ as its sample space. The situation is different for non-standard probability theory. The co-domain, $\mathcal{R}$, of an NAP function, P, depends, *inter alia*, on its domain, $\mathcal{P}(\Omega)$.

Hájek ([unpublished]) remarks that a probability theory that allows for regularity will have to look very different from Kolmogorov's theory. By examining NAP as an example of such a theory, we can make this claim more precise. A cardinality argument easily shows that regularity cannot be ensured for arbitrary sample spaces if the range of the function is fixed (cf. Footnote 12). One crucial difference between NAP and Kolmogorov's theory is precisely that NAP constructs the range based on the sample space. However, this difference need not be

a problematic one. On the assumption that there is a strongly inaccessible cardinal, it can easily be shown that there are regular NAP probability functions that are defined on all sets according to certain models of Zermelo–Fraenkel set theory (with the axiom of choice).

6.2 Invariance

Now we return to the observation that the desideratum of regularity is in tension with that of invariance. The arguments by Williamson, Parker, and Barrett do show that NAP is incompatible with certain constraints on probabilities that have some intuitive pull. One-to-one correspondence is a celebrated criterion of identity for cardinal number of sets: this is the Humean principle (Section 3.6). But one-to-one correspondence is not the correct criterion of identity for an ordinal number of a well-ordering: for finite sets, one-to-one correspondence works fine; for infinite sets, it does not. So the criterion of identity in terms of one-to-one correspondence is a symmetry principle that holds for one concept (cardinal number) and not for another (ordinal number). Does it hold for the probability of a set in a uniform distribution context? As we have argued in Section 3.6, for finite sets it does, but for infinite sets it does not; otherwise, we would be forced in any countably infinite sample space to give all infinite sets a measure of one. This is something we do not want, for the concept of a sparse infinite sets lies within the scope of our pretheoretical concept of probability. We have seen that we also have to give up certain other invariance principles, such as translation-invariance (Williamson's coin tosses) and rotation-invariance (Barrett and Parker). This again becomes clear only when one considers infinite sample spaces: if we want a maximally fine-grained concept of probability, then we are forced to accept the Euclidean principle. And this principle imposes limits on the amount of invariance that a fine-grained probability function can support (Benci *et al.* [2013], Section 5.4).

It is likely that Williamson intends his argument to be given a physical interpretation. We know, one might say, that the laws of physics are time-translation invariant. But Williamson's argument purports to show that theories assigning infinitesimal probabilities to particular infinite sequences of fair coin tosses are not time-translation invariant. So, there is something wrong with modelling infinite sequences of coin tosses in this way. Williamson's infinite series of fair coin tosses probably already transports us out of the physical world. The scenario as described by Williamson is presumably inconsistent with our best current scientific theories. But ignoring that, it is still not easy to see why the NAP treatment of Williamson's scenario has to violate time-translation invariance. We have to keep track of the sample

space in which we are working. In the time-translated scenario, there just is a new sample space (containing one point event less than the original scenario).

For some purposes, we may want certain invariance properties even when dealing with a fine-grained concept of probability. For instance, when considering uniform distributions over the rational numbers, we may want the probability of an interval to be proportional to the length of the interval. To a large extent, such intuitions can be accommodated (Benci *et al.* [2013], Sections 5.3 and 5.4).

For other purposes, invariance behaviour (and simplicity) is more important than fine-grainedness. In all such situations, classical probability suffices. Still, as we have argued, there are kinds of probabilistic situations that cannot be modelled by classical probability but that can be modelled well by NAP theory.

We end by concluding that there is a legitimate place for non-Archimedean theories of probability. In the philosophical literature, infinitesimal probabilities have received much criticism, but most of it does not hold up to scrutiny. We have looked into NAP as a particular theory that contains infinitesimal probabilities. Although this theory has some counterintuitive consequences, it also has advantages over classical probability theory: it exhibits regularity, totality, perfect additivity, and weak Laplacianism. On balance, we find NAP to be a serious contender for a theory of probability, which we expect to be fruitful in shedding new light on old puzzles that combine probability and infinity.

Appendix: Λ, the Ω-limit

Let Λ be the family of the finite subsets of Ω ($\Lambda = \mathcal{P}_{fin}(\Omega)$)[49] and consider the class of real-valued functions, $\mathfrak{F}(\Lambda, \mathbb{R})$, defined on Λ. Notice that if we fix $A \in \mathcal{P}(\Omega)$, we have that, for $\lambda \in \Lambda$, the conditional probability $P(A|\lambda) \in \mathfrak{F}(\Lambda, \mathbb{R})$. Thus, in order to formulate CPP for NAP, we are led to the following axiomatic definition of 'Ω-limit'[50]:

Axiom 1 (Existence Axiom)
Every function $\varphi \in \mathfrak{F}(\Lambda, \mathbb{R})$ has a unique 'Ω-limit' in a superreal field $\mathfrak{R} \supseteq \mathbb{R}$, which will be denoted by

$$\lim_{\lambda \uparrow \Omega} \varphi(\lambda).$$

Axiom 2 (Real Numbers Axiom)
Let $\varphi \in \mathfrak{F}(\Lambda, \mathbb{R})$ be eventually equal to $c \in \mathbb{R}$, namely, assume that

---

[49] In general, Λ is a directed set $\subseteq \mathcal{P}_{fin}(\Omega)$. As explained in (Benci *et al.* [2013]), some properties of the resulting probability functions depend on the choice of Λ. However, in order to understand the construction of the limit, it is easiest to think of Λ as $\mathcal{P}_{fin}(\Omega)$.

[50] An anonymous referee suggested that it is more intuitive to require the following: if $\varphi \geq c$ eventually, then $\lim_{\lambda \uparrow \Omega} \varphi(\lambda) \geq c$. Please observe that this follows from the axioms.

$\exists \lambda_0 \in \Lambda,\ \forall \lambda \supset \lambda_0,\ \varphi(\lambda) = c$. Then:

$$\lim_{\lambda \uparrow \Omega} \varphi(\lambda) = c.$$

Axiom 3 (Sum and Product Axiom)

If $\varphi, \psi \in \mathfrak{F}(\Lambda, \mathbb{R})$, then

$$\lim_{\lambda \uparrow \Omega} (\varphi(\lambda) + \psi(\lambda)) = \lim_{\lambda \uparrow \Omega} \varphi(\lambda) + \lim_{\lambda \uparrow \Omega} \psi(\lambda),$$

$$\lim_{\lambda \uparrow \Omega} (\varphi(\lambda) \cdot \psi(\lambda)) = \lim_{\lambda \uparrow \Omega} \varphi(\lambda) \cdot \lim_{\lambda \uparrow \Omega} \psi(\lambda).$$

First of all, we want to show that these axioms are consistent, so we will build a model. If Ω is finite, the above axioms are trivially satisfied taking $\mathfrak{R} = \mathbb{R}$ and defining

$$\lim_{\lambda \uparrow \Omega} \varphi(\lambda) = \varphi(\Omega).$$

If Ω is infinite, we take a fine and free ultrafilter, $\mathcal{U}_\Lambda$, over Λ, and we set[51,52]

$$\mathfrak{R} = \mathfrak{F}(\Lambda, \mathbb{R})/\mathcal{U}_\Lambda. \tag{A.1}$$

The Ω-limit is defined by

$$\lim_{\lambda \uparrow \Omega} \varphi(\lambda) = [\varphi]_{\mathcal{U}_\Lambda}. \tag{A.2}$$

This model then has the required properties, with the fineness of the ultrafilter guaranteeing regularity.

It is also possible to show that all the models of Ω-limit have this form. More precisely, assume that we have a structure $(\mathfrak{R}, \lim_{\lambda \uparrow \Omega})$, where $\mathfrak{R}$ is a superreal field and

$$\lim_{\lambda \uparrow \Omega} \mathfrak{F}(\Lambda, \mathbb{R}) \rightarrow \mathfrak{R}$$

is an operator that satisfies axioms (1), (2), and (3). Then there is an ultrafilter such that Equations (A.1) and (A.2) hold. We refer to (Benci *et al.* [2013]) for further details on the Ω-limit.

By Equation (A.1), it follows that $\mathfrak{R}$ is a non-standard model of the real numbers. For this reason, we refer to $\mathfrak{R}$ as a field of hyperreal numbers. The relation of NAP with NSA is quite deep, particularly from the technical point of view and we refer to (Benci *et al.* [2013]) for a discussion of this point as well.

---

[51] $\mathfrak{F}(\Lambda, \mathbb{R})/\mathcal{U}_\Lambda$ denotes the set of equivalence classes $[\varphi]_{\mathcal{U}_\Lambda}$ with respect to the relation $\approx_{\mathcal{U}_\Lambda}$, defined by

$$\varphi \approx_{\mathcal{U}_\Lambda} \psi \Leftrightarrow \exists Q \in \mathcal{U}_\Lambda,\ \forall \lambda \in Q,\ \varphi(\lambda) = \psi(\lambda).$$

[52] Since we want to identify $\mathbb{R}$ with a subset of $\mathfrak{R}$, the equivalence class of a function φ_c identically equal to c must be identified with the real number c.

Funding

Acknowledgements

Vieri Benci
Dipartimento di Matematica
Università degli Studi di Pisa
Pisa, Italy
and
Centro Linceo Interdisciplinare Beniamino Segre
Rome, Italy
benci@dma.unipi.it

Leon Horsten
School of Arts
University of Bristol
Bristol, UK
leon.horsten@bristol.ac.uk

Sylvia Wenmackers
Institute of Philosophy
KU Leuven
Leuven, Belgium
sylvia.wenmackers@kuleuven.be

References

Albeverio, S., Fenstad, J. E., Hoegh-Krøhn, R. and Lindstrøm, T. [1986]: *Non-standard Methods in Stochastic Analysis and Mathematical Physics*, Orlando, FL: Academic Press.

Arlo-Costa, H. [2008]: 'The Logic of Conditionals', in N. Zalta (*ed.*), *Stanford Encyclopedia of Philosophy*, < http://plato.stanford.edu/archives/fall2008/entries/logic-conditionals/>.

Barrett, M. [2010]: 'The Possibility of Infinitesimal Chances', in E. Eells and J. H. Fetzer (*eds*), *The Place of Probability in Science*, Boston: Springer, pp. 65–79.

Bartha, P. and Hitchcock, C. [1999]: 'The Shooting-room Paradox and Conditionalizing on Measurably Challenged Sets', *Synthese,* **118**, pp. 403–37.

Bascelli, T., Bottazzi, E., Herzberg, F., Kanovei, V., Katz, K. U., Katz., M. G. and Shnider, S. [2014]: 'Fermat, Leibniz, Euler, and the Gang: The True History of the Concepts of Limit and Shadow', *Notices of the American Mathematical Society,* **61**, pp. 848–64.

Benci, V. [1995]: 'I numeri e gli insiemi etichettati', in *Conferenze del seminario di matematica dell'Universita' di Bari,* Bari, Italy: Laterza, p. 29.

Benci, V. and Di Nasso, M. [2003]: 'Alpha-theory: An Elementary Axiomatic for Nonstandard Analysis', *Expositiones Mathematicae,* **21**, pp. 355–86.

Benci, V., Di Nasso, M. and Forti, M. [2006]: 'An Aristotelian Notion of Size', *Annals of Pure and Applied Logic,* **143**, pp. 43–53.

Benci, V., Horsten, L. and Wenmackers, S. [2013]: 'Non-Archimedean Probability', *Milan Journal of Mathematics,* **81**, pp. 121–51.

Bernstein, A. R. and Wattenberg, F. [1969]: 'Nonstandard Measure Theory', in W. A. J. Luxemburg (*ed.*), *Applications of Model Theory to Algebra, Analysis, and Probability,* New York: Holt, Rinehard and Winston, pp. 171–85.

Brickhill, H. and Horsten, L. [unpublished]: 'Popper Functions, Lexicographical Probability, and Non-Archimedean Probability'.

Cantor, G. [1966]: *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts*, Hildesheim: Olms.

Császár, A. [1955]: 'Sur la structure des espaces de probabilité conditionnelle', *Acta Mathematica Academiae Scientiarum Hungarica,* **6**, pp. 337–61.

Cutland, N. [1983]: 'Nonstandard Measure Theory and Its Applications', *Bulletin of the London Mathematical Society,* **15**, pp. 529–89.

de Finetti, B. [1974]: *Theory of Probability,* London: Wiley.

Easwaran, K. [2014]: 'Regularity and Hyperreal Credences', *Philosophical Review,* **123**, pp. 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**, pp. 1–121.

Elga, A. [2004]: 'Infinitesimal Chances and the Laws of Nature', *Australasian Journal of Philosophy,* **82**, pp. 67–76.

Gilbert, T. and Rouche, N. [1996]: 'Y a-t-il vraiment autant de nombres pairs que de naturels?', in A. Pétry (*ed.*), *Méthodes et Analyse Non Standard,* Louvain-la-Neuve, Belgium: Bruylant-Academia, pp. 99–139.

Hájek, A. [2003]: 'What Conditional Probability Could Not Be', *Synthese,* **137**, pp. 273–323.

Hájek, A. [unpublished]: 'Staying Regular?'.

Halpern, J. Y. [2003]: *Reasoning about Uncertainty,* Cambridge, MA: MIT Press.

Halpern, J. Y. [2010]: 'Lexicographic Probability, Conditional Probability, and Nonstandard Probability', *Games and Economic Behavior,* **68**, pp. 155–79.

Hammond, P. J. [1994]: 'Elementary Non-Archimedean Representations of Probability for Decision Theory and Games', in P. Humphreys (*ed.*), *Patrick Suppes: Scientific*

Philosopher, Volume 1: Probability and Probabilistic Causality, Dordrecht: Springer, pp. 25–61.

Hofweber, T. [2014]: 'Infinitesimal Chances', *Philosophers' Imprint,* **14**, pp. 1–14.

Jaynes, E. T. [2003]: *Probability Theory: The Logic of Science,* Cambridge: Cambridge University Press.

Kanamori, A. [1994]: *The Higher Infinite: Large Cardinals from Their Beginnings*, Berlin: Springer.

Kerkvliet, T. and Meester, R. [2016]: 'Uniquely Determined Uniform Probability on the Natural Numbers', *Journal of Theoretical Probability*, doi:10.1007/s10959-015-0611-2.

Kolmogorov, A. N. [1956]: *Foundations of the Theory of Probability*, New York: Chelsea Publishing Company.

Krantz, D., Luce, R., Suppes, P. and Tversky, A. [1971]: *Foundations of Measurement, Volume 1: Additive and Polynomial Representations,* New York: Academic Press.

Krauss, P. H. [1968]: 'Representation of Conditional Probability Measures on Boolean Algebras', *Acta Mathematica Academiae Scientiae Hungaricae,* **19**, pp. 229–41.

Kremer, P. [2014]: 'Indeterminacy of Fair Infinite Lotteries', *Synthese,* **191**, pp. 1757–60.

Laplace, P. S. [1902]: *Philosophical Essay on Probabilities.* New York: Wiley.

Leitgeb, H. [2012]: 'A Probabilistic Semantics for Counterfactuals', *Review of Symbolic Logic,* **5**, pp. 26–121.

Lewis, D. K. [1980]: 'A Subjectivist's Guide to Objective Chance', in R. C. Jeffrey (*ed.*), *Studies in Inductive Logic and Probability,* Volume 2, Berkeley, CA: University of California Press, pp. 263–93.

Loeb, P. A. [1975]: 'Conversion from Nonstandard to Standard Measure Spaces and Applications in Probability Theory', *Transactions of the American Mathematical Society,* **211**, pp. 113–22.

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**, pp. 612–46.

McCall, S. and Armstrong, D. M. [1989]: 'God's Lottery', *Analysis,* **49**, pp. 223–4.

McGee, V. [1994]: 'Learning the Impossible', in E. Eells and B. Skyrms (*eds*), *Probability and Conditionals: Belief Revision and Rational Decision,* Cambridge: Cambridge University Press, pp. 179–99.

Nelson, E. [1987]: *Radically Elementary Probability Theory,* Princeton, NJ: Princeton University Press.

Parker, M. [2013]: 'Set Size and the Part–Whole Principle', *The Review of Symbolic Logic,* **6**, pp. 589–612.

Pedersen, A. P. [2014]: 'Comparative Expectations', *Studia Logica,* **102**, pp. 811–48.

Pedersen, A. P. [unpublished]: 'Strictly Coherent Preferences, No Holds Barred'.

Pivato, M. [2014]: 'Additive Representation of Separable Preferences over Infinite Products', *Theory and Decision,* **77**, pp. 31–83.

Pruss, A. [2012]: 'Infinite Lotteries, Perfectly Thin Darts, and Infinitesimals', *Thought,* **1**, pp. 81–9.

Pruss, A. [2013]: 'Probability, Regularity, and Cardinality', *Philosophy of Science,* **80**, pp. 231–40.

Pruss, A. [2014]: 'Infinitesimals Are Too Small for Countably Infinite Fair Lotteries', *Synthese,* **191**, pp. 1051–7.

Rényi, A. [1955]: 'On a New Axiomatic Theory of Probability', *Acta Mathematica Hungarica,* **6**, pp. 285–335.

Rényi, A. [1956]: 'On Conditional Probability Spaces Generated by a Dimensionally Ordered Set of Measures', *Theory of Probability and Its Applications,* **1**, pp. 55–64.

Robinson, A. [1961]: 'Non-standard Analysis', *Proceedings of the Royal Academy of Sciences, Amsterdam, Series A,* **64**, pp. 432–40.

Skyrms, B. [1980]: *Causal Necessity,* New Haven, CT: Yale University Press.

Skyrms, B. [1983]: 'Zeno's Paradox of Measure', in R. S. Cohen and L. Laudan (*eds*), *Physics, Philosophy, and Psychoanalysis: Essays in Hounour of Adolf Grunbaum*, Dordrecht: Reidel, pp. 223–54.

Truss, J. [1997]: *Foundations of Mathematical Analysis,* Oxford: Oxford University Press.

van Fraassen, B. C. [1976]: 'Representation of Conditional Probabilities', *Journal of Philosophical Logic,* **5**, pp. 417–30.

van Fraassen, B. C. [1989]: *Laws and Symmetry,* Oxford: Oxford University Press.

Walley, P. [2000]: 'Towards a Unified Theory of Imprecise Probability', *International Journal of Approximate Reasoning,* **24**, pp. 125–48.

Weintraub, R. [2008]: 'How Probable is An Infinite Sequence of Heads? A Reply to Williamson', *Analysis,* **68**, pp. 247–50.

Wenmackers, S. [unpublished]: 'Infinitesimal Probability', in R. Pettigrew and J. Weisberg (*eds*), *Open Handbook of Formal Epistemology*.

Wenmackers, S. and Horsten, L. [2013]: 'Fair Infinite Lotteries', *Synthese,* **190**, pp. 37–61.

Williamson, T. [2007]: 'How Probable Is an Infinite Sequence of Heads?', *Analysis,* **67**, pp. 173–80.