THE REVIEW OF SYMBOLIC LOGIC
Volume 11, Number 3, September 2018

TRIANGULATING NON-ARCHIMEDEAN PROBABILITY

HAZEL BRICKHILL
Graduate School of Engineering, Kobe University
and
LEON HORSTEN
Department of Philosophy, University of Bristol

Abstract. We relate Popper functions to regular and perfectly additive such non-Archimedean probability functions by means of a representation theorem: every such non-Archimedean probability function is infinitesimally close to some Popper function, and vice versa. We also show that regular and perfectly additive non-Archimedean probability functions can be given a lexicographic representation. Thus Popper functions, a specific kind of non-Archimedean probability functions, and lexicographic probability functions triangulate to the same place: they are in a good sense interchangeable.

Introduction. Popper functions were introduced in Popper (1959). They have been used to describe quantitative probabilistic concepts and situations that fall outside the scope of classical probability theory. In particular, Popper functions have proved useful in modelling learning from evidence, indicative and counterfactual conditionals, decision and utility theory McGee (1994), Leitgeb (2012).

Popper functions are typically defined axiomatically, by a list of properties that these functions are required to satisfy. Of course such presentations are accompanied by demonstrations that the list of axioms is consistent.

It is immediate from the axioms of Kolmogorov that a concept of probability is thereby defined: the Kolmogorov axioms represent intuitive properties of an informal concept of probability. This is not so for the axioms that define Popper functions. These axioms implicitly define a conditional function on a class of propositions. But it is not immediate from the Popper function axioms that they define a concept of conditional *probability*. It is difficult to see which models are allowed by the axioms governing Popper functions. So one may wonder whether there are some Popper functions that really cannot be taken to represent probability. It has been argued that the notion of conditional probability is actually prior to its definition in terms of absolute probability Hajek (2003), and so the idea of Popper functions is a reasonable one. But how can we tell that Popper hit upon the right set of axioms?

One principal aim of the present article is to make a case for the thesis that Popper functions do capture a pretheoretical concept of quantitative probability. We do this by relating Popper functions to a class of functions for which it is more immediate that they

Received: December 6, 2016.
2010 *Mathematics Subject Classification*: 03B42, 26E30, 28E05, 60A05.
Key words and phrases: nonstandard probability, non-Archimedean probability, Popper functions, lexicographic probability, infinitesimals.

© Association for Symbolic Logic, 2018

doi:10.1017/S1755020318000060

https://doi.org/10.1017/S1755020318000060 Published online by Cambridge University Press

capture an informal concept of quantitative probability: we will relate Popper functions to a specific class of non-Archimedean probability functions.

The idea of infinitely small numbers (infinitesimals) goes back at least to Leibniz and Newton. It was shown to be mathematically coherent in the middle of the last century by Abraham Robinson Robinson (1961).

The concept of fair lotteries on infinite sample spaces motivates the application of infinitesimals to the theory of probabilities. In particular, for such a lottery it seems prima facie reasonable to assign a nonzero but infinitesimally small probability value to every proposition that states that a given ticket is the winning one. A property that one might want to impose on non-Archimedean probability functions in infinite lottery situations is *regularity*, i.e., the condition that only the impossible event receives probability 0. Also, one might insist on *perfect additivity*, which says (roughly) that the probability of an event can be obtained as the sum of the probabilities of its atomic subevents (of which there may be infinitely many). In this article, we focus on one particular form that such a non-Archimedean probability theory can take, namely the theory that was developed in Benci, Horsten, & Wenmackers (2013) and Wenmackers & Horsten (2013).[1]

A key difference between Popper functions and non-Archimedean probability functions is that the former take values in the real [0, 1] interval whereas the latter take values in the [0, 1] interval of a non-Archimedean extension of $\mathbb{R}$. Yet we will show that Popper functions are closely related to non-Archimedean probability functions. We will prove a *representation theorem* that relates regular and perfectly additive non-Archimedean probability functions to Popper functions (§3). On the one hand, for every finitely additive Popper function, there is a regular and perfectly additive non-Archimedean probability function that is pointwise infinitesimally close to it. On the other hand, for every regular perfectly additive non-Archimedean function there is a Popper function that is pointwise infinitesimally close to it. The latter is fairly trivial; the former is not.

One direction of the theorem shows that Popper functions cannot do anything that regular and perfectly additive non-Archimedean probability functions cannot do. So if one believes that such non-Archimedean probability functions (as explicated in Benci *et al.*, 2013) model an intuitive notion of probability well, then Popper functions must too: while the latter may not give us as much information, they cannot go against our intuitions concerning concepts of probability to any greater extent than such non-Archimedean probability functions. The converse direction of the theorem shows that in all the applications of Popper functions to describe situations or elucidate concepts, the notion of regular and perfectly additive non-Archimedean probability may also be employed.

There are still reasons one may prefer non-Archimedean probability functions over Popper functions or vice versa: for instance, a non-Archimedean probability function will enable us to compare probabilities at a much finer level than a Popper function, distinguishing between events whose probabilities only differ by infinitesimal amounts. On the other hand, one may argue that this level of detail is uncalled for and not justified by our intuitions. But neither can go too far wrong if you believe the other is correct.

Non-Archimedean probability functions are themselves not completely intuitive. One source of mystery is the nonwellfoundedness of the degrees of infinitesimality that such

[1] There are of course other approaches to non-Archimedean probability: see for instance Nelson (1987). For a defence of the thesis that the theory that we are focussing on in this article constitutes a good framework for developing a theory of infinitesimal probabilities, see Benci, Horsten, & Wenmackers (2018).

https://doi.org/10.1017/S1755020318000060 Published online by Cambridge University Press

probability functions entail. This seems to open the prospect that the probability of one event can be smaller than that of another event even though there is no largest degree of infinitesimality at which they differ. This would of course make such functions very hard to picture.

In response to this, we prove a second representation theorem (§4). This second theorem relates the non-Archimedean probability functions under consideration to lexicographical probability functions as discussed in Blume, Brandenburger, & Dekel (1991) and in (Halpern, 2010, sec. 1–3). In particular, we show that despite the nonwellfounded structure of the range of non-Archimedean probability functions, probability values can be ordered lexicographically. This phenomenon gives us deeper insight into the structure of non-Archimedean probability functions. Conversely, it links the applications of lexicographic probability theory to non-Archimedean probability. In particular, it suggests ways in which non-Archimedean probability theory can be connected to utility and decision theory by supplying hierarchical orderings of 'null events'.[2]

In sum, the representation theorems that are proved in this article are intended to shore up the foundations of all three representations of probability. If one has concerns about any two of the three theories investigated in this article, but not about the third, then the results of this article offer some reassurance. We have a virtuous triangle of representations: starting anywhere on the triangle, one can get anywhere else.

Our representation theorems are related to representation theorems in this area that are explained and discussed in Halpern (2010). In the present article we focus on a *restricted* class of non-Archimedean probability functions, namely those that satisfy a natural and general infinite additivity principle. This is a particularly important class of non-Archimedean probability functions, since these functions systematically relate the probabilities of large events to their infinitely many sub-events. The known representation theorems hold for the class of *all* non-Archimedean probability functions. This class is much wider, so our representation theorems are stronger than those in the literature. To obtain our stronger results, we have to do significantly more and new mathematical work.

The construction of the non-Archimedean probability functions uses mathematical tools that some readers may not be familiar with—in particular ultrafilters and ultraproduct constructions. In §2 where we introduce these functions. We give the precise definitions but also describe the basic properties that result, and the short proofs here give a flavour of how these are used. The proofs of the main theorems in the article are technical, but only an understanding of the statements of the theorems is necessary to grasp their philosophical import.

Our representation theorems relate Popper functions and lexicographic probability functions to one *specific* kind of non-Archimedean probability functions that are based on a relative frequency conception of probability. We would be remiss if we would not stress at this point that there are alternative non-Archimedean probability theories based on a relative frequency notion of probability.[3]

In particular, Khrennikov's p-adic probability theory is also non-Archimedean, based on a relative frequency idea, and was worked out before the non-Archimedean probability

[2] The connection between utility theory and non-Archimedean probability theory for infinite state spaces is investigated in Pedersen (2018).

[3] Thanks to an anonymous referee for drawing our attention to this and for encouraging us to discuss some of these alternative non-Archimedean approaches in this article.

https://doi.org/10.1017/S1755020318000060 Published online by Cambridge University Press

theory used in the present article was developed.[4] Moreover, within this framework of p-adic probability theory a natural theory of conditional probability can be developed Ilic-Stepic, Ognjanovic, & Ikodinovic (2014).

In p-adic probability theory, much emphasis has been placed on *applications*, whereas the non-Archimedean theory that we will be using has so far only been applied to fair infinite lotteries Wenmackers & Horsten (2013). Schumann has given an interesting application of non-Archimedean probability theory to higher-order fuzzy logic Schumann (2008). Moreover, he has shown how non-Archimedean probability theory in general, and p-adic probability theory in particular, can be used to give a mathematical description of classes of games for which Aumann's agreement theorem fails Schumann (2014).

We should also mention here that Popper functions are not the only approach for conditionalising on events that have probability zero. Various authors including de Finetti (1949), Renyi (1955) and Krauss (1968) have presented theories of *conditional probability*, and this literature seems to have developed largely independently from the more philosophical literature on Popper functions.[5] Krauss gives a representation theorem of *full conditional probability functions* in terms of strictly positive (i.e., regular) non-Archimedean probability functions (Krauss, 1968, Theorems 3.2 and 3.4), and so with the representation theorem of McGee presented in §3.1 (Theorem 3.2) doing the same for Popper functions, it is clear that the two approaches are essentially equivalent in our context.[6] Thus in this article we will focus solely on the Popper function presentation.

§1. Popper functions.

1.1. Axiomatic description. Popper functions were originally introduced so that conditional probabilities could be assigned where the event conditioned on would receive absolute probability zero. They are both a weakening and an extension of classical probability. Countable additivity is not imposed (it can be, but then Popper functions become less useful for modelling situations, like the infinite lotteries alluded to earlier, which classical probability cannot model), and additional conditional probabilities are defined. Popper functions are introduced *axiomatically*, the suggestion being that these are the right axioms to model conditional probability. We follow the presentation in Leitgeb (2012).

For some propositional language $\mathcal{L}$ let $C : \mathcal{L} \times \mathcal{L} \to [0, 1]$. C is Popper function if it satisfies the following:

1. $C(a, a) = 1$.
2. If $C(\neg a, a) \neq 1$ then $C(., a)$ is a probability function (i.e., satisfying the Kolmogorov axioms with finite additivity replacing countable additivity).
3. $C(a \wedge b, d) = C(a, d) \times C(b, a \wedge d)$.
4. If $C(a, b) = C(b, a) = 1$ then for all $d \in \mathcal{L}$, $C(d, a) = C(d, b)$.

The intended interpretation of $C(a, b)$ here is the probability of a occurring conditional on b occurring. It is not immediately obvious that conditional probabilities should satisfy

[4] p-adic probability theory goes back to Khrennikov (1996). For a good introductory overview of this theory, developed along analytical lines, see Khrennikov (2008). For a more logico-algebraic approach, see Khrennikov & Schumann (2006).

[5] Thanks to an anonymous referee for pointing this out to us.

[6] For a discussion of the relationship between these two approaches in a more general context, see Halpern (2010).

https://doi.org/10.1017/S1755020318000060 Published online by Cambridge University Press

these axioms but a little work shows they do, at least in intuitive situations. (If we have $C(\neg a, a) = 1$ then it can be shown that $C(., a)$ is the constant function 1, so that a is to be interpreted as impossible or contradictory event). Also note that absolute probabilities can be retrieved from a Popper function by taking the probability conditional on a tautology.

Of course none of this would be of any significance if the axioms governing Popper functions were inconsistent. But it can be shown that on all sample spaces, finite and infinite, Popper functions can be defined.

1.2. Uses. Popper functions do solve some of the problems outlined above for the standard theory of probability. In particular, the main selling point of Popper functions is of course reclaiming the conditional probabilities that are intuitively justified (McGee, 1994, sec. 3). Also, certain situations that cannot be modelled with standard probability functions (such as a fair lottery on $\mathbb{N}$), can be consistently modelled using a Popper function Van Fraassen (1976).

For these reasons, Popper functions have been quite useful in philosophy. In Bayesian theories of belief revision under new evidence, a rational agent will start with a credence function that obeys the standard laws of probability. Then this credence is revised under new evidence E, so that the new credence of a proposition is the old credence of that proposition conditional on E. But in standard probability theory, if the evidence E has probability zero, such conditionals are undefined. Should we be worried about how to rationally react to evidence that has probability zero? Well, given that this evidence may still be possible, we should. After all, a theory of rationality is not primarily a practical theory: what really concerns us is what it means to be rational, so we should take all possibilities into account Levi (1989). One obvious way of doing this is to use Popper functions.[7]

Further uses for Popper functions include theories of indicative and counterfactual conditionals. For an application of Popper functions to the theory of indicative conditionals, the reader may consult McGee (1994); for the use of Popper functions in the theory of counterfactual conditionals, Leitgeb (2012).

However, this usefulness really depends on our being willing to accept that Popper functions provide a good model of conditional probability. And it is not clear that we should be. The problem of probabilities of infinite sets not depending on the probabilities of their component subsets is even worse here than in the standard picture, as we don't even have countable additivity. It seems that the probabilities could somehow be arbitrary, floating free of the probabilities of point-events that make them up. And if so, this cannot be consistent with our intuitive notion of probability. But is that really possible? That is one of the questions this article seeks to answer, by showing how Popper functions are closely tied to a richer alternative theory: a theory of non-Archimedean probability.

§2. Non-Archimedean probability functions. From now on, we will focus on one particular non-Archimedean probability, namely the theory that we will call $PANAP$ ('Perfectly Additive Non-Archimedean Probability'), which was developed in Benci *et al.*, (2013).

In $PANAP$, the unconditional probability of an event is defined in terms of the conditional probability of an event. Loosely speaking, the probability $P(A)$ of event A is

[7] For a defence of the role of Popper functions in satisfactory theories of conditionalisation on evidence, see, e.g., Harper (1975). For a discussion of the limitations of the usefulness of Popper functions for such purposes, see Pruss (2015).

https://doi.org/10.1017/S1755020318000060 Published online by Cambridge University Press

conceived of as the *limit* of $P(A \mid \lambda)$, for the finite set λ "tending toward infinity": $P(A)$ is conceived of as the limit of the relative frequency of A's on finite snapshots of the sample space.

In this section, we sketch how $PANAP$ functions are constructed, and what their basic properties are.[8]

2.1. Limits. Let there be given a sample space Ω. And let Λ be a directed (under the inclusion ordering) subset of the collection $\mathcal{P}_{fin}(\Omega)$ (the finite subsets of Ω) such that $\bigcup \Lambda = \Omega$. For any $\lambda \in \Lambda$ and for any event A (i.e., any subset of Ω), $P(A \mid \lambda) \in \mathbb{R}$, where $P(A \mid \lambda)$ is taken to be defined as in *classical* probability theory using the *ratio formula* (since λ is finite):

$$P(A \mid \lambda) \equiv \frac{|A \cap \lambda|}{|\lambda|}.$$

Intuitively, Λ contains the finite subsets on which any given event A is "tested", and as the test sets move up the ordering, they "approach infinity". We want to conceive of the sought for probability function as a kind of *limit*. We are of course interested in the case where Ω is infinite.

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

$$\mathbb{R}^* \equiv \mathfrak{F}(\Lambda, \mathbb{R}) / \mathcal{U}_\Lambda, \tag{1}$$

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

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

The elements of $\mathbb{R}^*$ can then indeed be regarded as 'limits':

DEFINITION 2.1 (Ω-limit).

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

Thus $\lim_{\lambda \uparrow \Omega} \varphi(\lambda)$ is defined for any function $\varphi : \Lambda \to \mathbb{R}$. Note that the limit operator of the left hand side depends on the choice of ultrafilter $\mathcal{U}_\Lambda$. We suppress this dependence in our notation as the general properties discussed in this section hold for any such choice.

The set $\mathbb{R}^*$ will serve as the range of our Non-Archimedean Probability function. We want to identify $\mathbb{R}$ with a subset of $\mathbb{R}^*$. Therefore we identify the equivalence class of the function φ_c with constant value c with the real number c. We want to calculate (add, multiply) with the elements of $\mathbb{R}^*$. Therefore we define addition and multiplication on elements of $\mathbb{R}^*$ pointwise. It can then be verified, using standard arguments from nonstandard analysis, that $\mathbb{R}^*$ is a (non-Archimedean) field.

2.2. Infinitesimal probabilities. Using this notion of Ω-limit we can now define a non-Archimedean probability function as follows:

[8] For details of the proofs and constructions, see Benci *et al.* (2013). The only difference between the constructions in Benci *et al.* (2013) and the constructions in this article is that in the present article, the constructions are carried out using filters instead of ideals. But it is easily verified that the two formulations of the theory are equivalent.

https://doi.org/10.1017/S1755020318000060 Published online by Cambridge University Press

DEFINITION 2.2.

$$P(A) = \lim_{\lambda \uparrow \Omega} P(A \mid \lambda). \tag{3}$$

The intuitive meaning of the equation $P(A) = \lim_{\lambda \uparrow \Omega} P(A \mid \lambda)$ is that the probability of an event A is the Ω-*limit* of the conditional probability $P(A \mid \lambda)$ with λ a finite sample set. Note that because $\lim_{\lambda \uparrow \Omega} \varphi(\lambda)$ is always defined, our probability function P is defined for every subset of Ω.

It can then be shown that P satisfies the laws for being a finitely additive probability function (except that the value is taken not in $\mathbb{R}$ but in a non-Archimedean field):

- (NAP1) **Domain and range.** *The events are* ***all*** *the subsets of* Ω*, which is a finite or infinite sample space. Probability is a total function*

  $$P : \mathcal{P}(\Omega) \to \mathbb{R}^*,$$

  where $\mathbb{R}^*$ *is a superreal field, where a superreal field is an ordered field which contains the real numbers as subfield.*
- (NAP2) **Normalization.**

  $$P(\Omega) = 1. \tag{4}$$
- (NAP3) **Additivity.** *If* A *and* B *are events and* $A \cap B = \varnothing$*, then*

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

Since the non-Archimedean probability function that is thus defined depends on the initial choice of free ultrafilter $\mathcal{U}$, we should strictly speaking write $P_{\mathcal{U}}$ instead of P.

2.3. Regularity. We now look at how we can impose a general condition on the limit construction to ensure the probability functions that result are *regular*, i.e., that for such P:

DEFINITION 2.3 (Regularity). $\forall A \in \mathcal{P}(\Omega) \setminus \{\varnothing\}$:

$$P(A) > 0. \tag{5}$$

We have seen that the probability functions $P_{\mathcal{U}}$ that we have defined so far are determined by ultrafilters $\mathcal{U}$ on Λ—which is a directed (under the inclusion ordering) subset of the collection $\mathcal{P}_{fin}(\Omega)$. The regularity of $P_{\mathcal{U}}$ can be forced to hold by imposing a condition on the ultrafilter $\mathcal{U}$ on which it is based:

DEFINITION 2.4 (Fine ultrafilter). *An ultrafilter* $\mathcal{U}$ *on* Λ *is fine if and only if for every* $a \in \Omega$*, we have*

$$\{\lambda \in \mathcal{P}_{fin}(\Omega) : a \in \lambda\} \in \mathcal{U}.$$

Then we immediately have:

PROPOSITION 2.5. *If* $\mathcal{U}$ *is fine, then* $P_{\mathcal{U}}$ *is regular.*

Proof. Consider any $a \in \Omega$. Then $P_{\mathcal{U}}(\{a\}) \neq 0$ if and only if

$$\{\lambda \in \mathcal{P}_{fin}(\Omega) : P(\{a\} \mid \lambda) \neq 0\} \in \mathcal{U}.$$

But $\{\lambda \in \mathcal{P}_{fin}(\Omega) : P(\{a\} \mid \lambda) \neq 0\} = \{\lambda \in \mathcal{P}_{fin}(\Omega) : a \in \lambda\}$, which belongs to $\mathcal{U}$ by the fineness condition. □

2.4. Weights. The class of non-Archimedean probability functions that we have defined so far are those that are determined by fine (and free) ultrafilters according to the

https://doi.org/10.1017/S1755020318000060 Published online by Cambridge University Press

recipe described in the previous sections. But this class is still too narrow, because it turns out that all those functions $P_{\mathcal{U}}$ are *uniform*:

PROPOSITION 2.6. *For all fine ultrafilters $\mathcal{U}$ on Λ, and for all $a, b \in \Omega$:*

$$P_{\mathcal{U}}(\{a\}) = P_{\mathcal{U}}(\{b\}).$$

Proof. $P_{\mathcal{U}}(\{a\}) = P_{\mathcal{U}}(\{b\})$ if and only if

$$\{A \in \mathcal{P}_{fin}(\Omega) : P(\{a\} \mid A) = P(\{b\} \mid A)\} \in \mathcal{U}.$$

But $\{A \in \mathcal{P}_{fin}(\Omega) : P(\{a\} \mid A) = P(\{b\} \mid A)\} =$
$\{A \in \mathcal{P}_{fin}(\Omega) : P(\{a\} \mid A) \leq P(\{b\} \mid A)\} \cap \{A \in \mathcal{P}_{fin}(\Omega) : P(\{a\} \mid A) \geq P(\{b\} \mid A)\}$.
Now $\{A \in \mathcal{P}_{fin}(\Omega) : P(\{a\} \mid A) \leq P(\{b\} \mid A)\} =$
$\{A \in \mathcal{P}_{fin}(\Omega) : b \in A\} \cup \{A \in \mathcal{P}_{fin}(\Omega) : a \notin A \wedge b \notin A\}$. By fineness and the superset property for ultrafilters, we indeed have $\{A \in \mathcal{P}_{fin}(\Omega) : b \in A\} \cup \{A \in \mathcal{P}_{fin}(\Omega) : a \notin A \wedge b \notin A\} \in \mathcal{U}$, i.e., $\{A \in \mathcal{P}_{fin}(\Omega) : P(\{a\} \mid A) \leq P(\{b\} \mid A)\} \in \mathcal{U}$. Similarly, $\{A \in \mathcal{P}_{fin}(\Omega) : P(\{a\} \mid A) \geq P(\{b\} \mid A)\} \in \mathcal{U}$. So their intersection must also be in $\mathcal{U}$. □

But we do not want to build uniformity into the *definition* of our non-Archimedean probability functions. So we allow also probability functions $P_{\mathcal{U},w}$ that are "tempered" by a real-valued *weight function* w. To be more precise, our official definition of non-Archimedean probability functions is as follows:

DEFINITION 2.7 ($PANAP$ functions). *If $\mathcal{U}$ is a free and fine ultrafilter on a directed set $\Lambda \subseteq \mathcal{P}_{fin}(\Omega)$, and we have a weight function $w : \Omega \mapsto \mathbb{R}^+$, and $P_{\mathcal{U}}$ is as defined in the preceding sections, then*

$$P(A) = \lim_{\lambda \uparrow \Omega} P(A \mid \lambda)$$

is a ***non-Archimedean probability function*** *(NAP function), where*

$$P(A \mid \lambda) \equiv \frac{\Sigma_{a \in A \cap \lambda}(w(a))}{\Sigma_{a \in \lambda}(w(a))}.$$

This then is the class of non-Archimedean probability functions that we will be concerned with in this article. Clearly P is then always a finitely additive probability function, is regular (if w takes strictly positive values everywhere), and may be but need not be uniform.

As before, we should strictly speaking write $P_{\mathcal{U},w}$ instead of P. Often in what follows it will be clear from the context what w and $\mathcal{U}$ are; in such cases, we will omit the subscripts from $P_{\mathcal{U},w}$.

Our non-Archimedean probability functions P are then determined by a triple $\langle \Omega, \mathcal{U}, w \rangle$, where

- Ω is the sample space;
- $\mathcal{U}$ is a free and fine ultrafilter on some directed subset Λ of $\mathcal{P}_{fin}(\Omega)$ such that $\bigcup \Lambda = \Omega$;
- $w : \Omega \to \mathbb{R}^+$ is a weight function.

A triple $\langle \Omega, \mathcal{U}, w \rangle$ is called a *NAP space*. A standard (Kolmogorov) probability function is of course also determined by a triple, but there one chooses, instead of an ultrafilter, an event algebra (a σ-algebra). In the $PANAP$ setting one need not choose an event algebra, since $PANAP$ functions are always defined on the full power set of the sample

https://doi.org/10.1017/S1755020318000060 Published online by Cambridge University Press

space. Against that, must be noted that the behaviour of an $PANAP$ function can be very sensitive to the choice of ultrafilter.[9]

2.5. Infinite sums and perfect additivity. In general, σ-additivity cannot be had for PANAP functions (Wenmackers & Horsten, 2013, sec. 5.2). Indeed, the absence of σ-additivity for non-Archimedean probability is a key element in what makes its important applications (such as Schumann (2014)) possible. And yet we claim that PANAP functions are *perfectly additive*, so what do we mean? The concept of perfect additivity is not mathematically defined, but rather expresses the intuition that the probability of the union of disjoint events is some sort of sum of the probability of those component events—no matter how many of these component events there are. The question is how to interpret this sum when we have infinitely many disjoint events. For PANAP functions it can be shown that (Benci *et al.*, 2013, sec. 3.4):

PROPOSITION 2.8 (Infinite sum rule). *If* $A = \bigcup_{i \in I} A_i$, *with* $A_i \cap A_j = \emptyset$ *for all* $i, j \in I$, *then*:

$$P(A) = \lim_{\lambda \uparrow \Omega} (\Sigma_{i \in I} P(A_i \mid \lambda)).$$

Here, instead of taking a limit of finite sums defined by restricting on I, as in the standard infinite sum, we restrict with λ; as λ is always finite, this sum is well defined. This natural infinite sum rule holds not only for countable families of events, but for families of events of any cardinality. This is why we call $PANAP$ functions perfect additive.[10]

The infinite sum rule gives a way to calculate the probability of an event from the probabilities of smaller events, but we can also express the probability of an event in terms of the weight function using a new notion of infinite sum. The Weierstrass limit allows us to give a rigorous definition of the sum of an infinite sequence. Analogously, the Ω-limit allows the definition of the sum of infinitely many real numbers. We now investigate this operation.

Let x_ω be a family of real numbers indexed by $\omega \in E \subseteq \Omega$; the Ω-sum of all $x_\omega{}'s$ is defined as follows:

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

Notice that, since λ is always finite, the function

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

is well defined, always yielding a real number as function value. So as with the usual definition of infinite sum, we are taking a limit of finite sums.

Our new type of infinite sum of differs in several important respects from the usual Weierstrass-sum. First of all, 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 $\Sigma_{\omega \in E; \mathcal{U}_\Lambda} x_\omega$ rather than $\Sigma_{\omega \in E} x_\omega$. Second, the Weierstrass-sum of a series exists only for

[9] For a striking example of this, see Kremer (2014). For further philosophical discussion of the sensitivity to choice of ultrafilter, see Benci *et al.* (2018, sec. 6.1).

[10] For a discussion of the virtues of perfect additivity, see Schervish, Seidenfeld, & Kadane (2017). A consequence of their work is that no probability function which models a fair lottery on a the natural numbers can be additive in the sense that $P(A) = \Sigma P(A_i)$ for any reasonable limit sum Σ. Thus a different limit, such as ours restricting with λ, is necessary. Thanks to an anonymous referee for pointing this out.

https://doi.org/10.1017/S1755020318000060 Published online by Cambridge University Press

certain *denumerable* 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. Lastly, the Weierstrass-sum of a series—if it exists—is a real number, while the result of a Ω-sum is a hyperreal number in $\mathbb{R}^*$.

This notion of infinite sum also allows us to express the (non-Archimedean) probability of an event in a new way, namely as

$$P(A) = \frac{\Sigma_{a \in A}(w(a))}{\Sigma_{a \in \Omega}(w(a))}.$$

This is the infinite analogue of the weight adjusted frequency formula for calculating the probability of an event if a finite event space, and gives us another justification for claiming perfect additivity.

2.6. Uses. Non-Archimedean probability theory has been used to generate probability functions on infinite (and even uncountable) sample spaces that reflect the distinction between possibility and impossibility (Lewis, 1980). More in particular, they have recently been used to model fair lotteries on infinite sample spaces (Wenmackers & Horsten, 2013; Norton, 2018).

Recently, non-Archimedean probability functions have also been used in natural language semantics to generate semantic values for paradoxical sentences such as the liar sentence (Campbell-Moore, Horsten, & Leitgeb, forthcoming).

§3. Relating Popper functions to non-Archimedean probability functions. The main aim of the present section is to prove the following:

THEOREM 3.1. 1. *For every finitely additive Popper function on any sample space, there is a regular NAP function that is pointwise infinitesimally close to it.*

2. *For every regular PANAP function on any sample space, there is a finitely additive Popper function that is infinitesimally close to it.*

This theorem extends an earlier theorem that was proved by McGee in McGee (1994).

The second part of the theorem that we seek to prove is straightforward. Indeed, it is routine to verify that given any regular non-Archimedean probability function P, if $C(B \mid A)$ is defined as

$$st\left(\frac{P(B \wedge A)}{P(A)}\right)$$

(where $st(a)$ denotes the unique real number that is closest to a), then C satisfies the axioms governing Popper functions.

However, the first part of the theorem that we seek to establish is nontrivial. Given a Popper function on some sample space, we will construct a suitable $PANAP$ function, but the $PANAP$ function will not have the same underlying sample space.

Indeed, the following argument shows that we cannot in general expect the $PANAP$ function to have the same sample space.[11] Consider the following total, regular non-Archimedean (but not $PANAP$!) probability function on $\mathbb{N}$:

$$P_1(A) \equiv \epsilon \cdot \delta_0(A) + (1 - \epsilon) \cdot \sum_{n \in A \cap \mathbb{N}_0} \frac{1}{2^n},$$

[11] Due to Arthur Pedersen, personal communication.

https://doi.org/10.1017/S1755020318000060 Published online by Cambridge University Press

where ϵ is some positive infinitesimal and δ_0 is the $0-1$-valued function such that for every $A \subseteq \mathbb{N}$, $\delta_0(A) = 1$ if $0 \in A$, and $\delta_0(A) = 0$ otherwise. Let C be the Popper function $st(P_1)$. Since $P_1(\{0\} \mid \{0,1\})$ is infinitesimally small, $C(\{0\},\{0,1\})$ must be 0. Now for every $PANAP$ function P, $P(\{0\} \mid \{0,1\})$ must be a real number (different from 0), since $\{0,1\}$ is a finite set. So there can be no $PANAP$ function P such that $|C(\{0\},\{0,1\}) - P(\{0\} \mid \{0,1\})|$ is infinitesimally small.

3.1. McGee's theorem. McGee effectively proves the following theorem (McGee, 1994, pp. 181–184):

THEOREM 3.2. *For every Popper function for a propositional language $\mathcal{L}$, there is a regular non-Archimedean probability function on $\mathcal{L}$ that is pointwise infinitesimally close to it.*

The structure of McGee's proof is roughly as follows.

Let $\mathcal{L}$, and a Popper function $C : \mathcal{L} \times \mathcal{L} \mapsto \mathbb{R}$, be given.

Stage 1

Let $\mathcal{L}_1, \ldots, \mathcal{L}_i, \ldots$ be an enumeration of the *finitely generated* sub-languages of $\mathcal{L}$. By "finitely generated language", McGee means a language that is generated by the familiar boolean operations from a finite set of basic events. For each $\mathcal{L}_i$, McGee generates a non-Archimedean probability function Pr_i defined on $\mathcal{L}_i$ which is infinitely close to C restricted to $\mathcal{L}_i$. In other words, he generates a probability function for any finitely generated event-algebra on the sample space Ω.

Van Fraassen has shown how on the basis of the given Popper function C, for every $\mathcal{L}_i$ a finite number n of *ranks* or degrees of infinitesimality are distinguished Van Fraassen (1976).[12] Every rank k is marked by a 'maximally probable' sentence $a_k \in \mathcal{L}_i$ of that rank, so that:

- $C(b, a_k) > 0$ for every sentence b of rank k;
- $C(b, a_k) = 0$ for every sentence b of rank $> k$;
- $C(a_k, b) = 0$ for every sentence b of rank $< k$.

McGee then shows how in terms of these ranks an infinitesimal probability function Pr_i can be expressed that is infinitely close to C.

What McGee's construction shows is effectively that the probability of a sentences $b \in \mathcal{L}_i$ can be represented as

$$Pr_i(b) = r_0 \cdot \epsilon^0 + \cdots + r_n \cdot \epsilon^n,$$

where ϵ is an infinitesimal number and $r_0, \ldots, r_n \in [0,1]$ (Halpern, 2010, §4.1). (r_0 is then $st[Pr_i(b)]$.)

In other words, $Pr_i(b)$ can be seen as a finite sequence $\langle r_0, \ldots, r_n \rangle$ of classical probability values. These finite sequences are lexicographically ordered. So the non-Archimedean probabilities for $\mathcal{L}_i$ can be given a *lexicographical* representation.

Stage 2

In the second part of the proof, an ultrafilter on the index set of the $\mathcal{L}_i$'s yields ultraproduct model Pr for the whole of $\mathcal{L}$. By the fact that for each i, j,

[12] See also the similar constructions in Krauss (1968) based on work in Renyi (1955).

https://doi.org/10.1017/S1755020318000060 Published online by Cambridge University Press

$$Pr_i(b \mid a) \approx C(b, a) \approx Pr_j(b \mid a),$$

we have that $Pr(b \mid a) \approx C(b \mid a)$, and we are done.

Note, incidentally, that the function Pr is not given as a triple $\langle \Omega, \mathcal{U}, w \rangle$, so the non-Archimedean function that McGee produces is not an $PANAP$ function. Also, the function Pr is only finitely additive, and not countably additive.

McGee then reflects on the familiar notion of σ-additivity (McGee, 1994, p. 184):

> For some purposes, it is useful to look at sentential calculi in which we can form infinite conjunctions and disjunctions and probability measures that are countably additive, rather than merely finitely additive. Thus, we may add to the definition of a Popper function this following requirement:
>
> If b is the disjunction of $a_0, a_1, a_2, \ldots$ and if $a_i \wedge a_j$ is inconsistent whenever $i \neq j$, then $C(b, c) = \Sigma_{i=0}^{\infty} C(a_i, c)$.
>
> For the corresponding condition in terms of nonstandard probability assignments, the first thing that comes to mind is nonsensical:
>
> If b is the disjunction of $a_0, a_1, a_2, \ldots$ and if $a_i \wedge a_j$ is inconsistent whenever $i \neq j$, then $P(b) = \Sigma_{i=0}^{\infty} P(a_i)$.
>
> The nonstandard model of analysis will not be topologically complete, and thus this infinite sum will not normally exist. Instead, the condition we require is:
>
> If b is the disjunction of $a_0, a_1, a_2, \ldots$ and if $a_i \wedge a_j$ is inconsistent whenever $i \neq j$, then $P(b) \approx \Sigma_{i=0}^{\infty} P(a_i)$.
>
> It follows immediately from [Theorem 3.2] that with these two additional requirements, the two approaches will again coincide.

Of course McGee is right that in a non-Archimedean context, σ-additivity cannot be the right infinite additivity rule.[13] But this not a valid reason for retreating to insisting only that the infinite sum (in the Weierstrass sense) of the probabilities and the probability of the infinite union should agree up to an infinitesimal. Indeed, we expect from our theory of probability that the probability of an infinite sum is *exactly* computable on the basis of the probabilities of its components: the probability of an infinite union of a family of events should be expressible as an infinite sum of the probabilities of the events in the family. We have seen in the previous section how, using a generalised limit concept, non-Archimedean probability functions are able to do this.

We have noted that $PANAP$ does have a natural infinite additivity rule. Indeed, $PANAP$ satisfies *perfect additivity*. So our representation Theorem 3.1 is a substantial strengthening of McGee's Theorem 3.2. We now turn to the proof of the main theorem.

3.2. Finitely generated languages. In this section we start, as McGee does (McGee, 1994, pp. 181–183), by proving the result for finitely generated languages, and then in

[13] The reasons for this are explained in Wenmackers & Horsten (2013).

https://doi.org/10.1017/S1755020318000060 Published online by Cambridge University Press

the next section we will use this to construct an $PANAP$ space for an arbitrary language (giving part 1 of Theorem 3.1, and so completing our main theorem). The construction of the $PANAP$ function in this section is fairly hands on, and mathematically not very difficult.

First we look at the language which describes the events. Let $\mathcal{L}_f$ be a finitely generated propositional language. So $\mathcal{L}_f$ has a finite number of atomic propositions, and is closed under the Boolean operations of conjunction, disjunction and negation. As $\mathcal{L}_f$ is finitely generated we can choose $b_1, b_2, \ldots, b_m$ from $\mathcal{L}_f$ such that:

1. $b_i \wedge b_j \leftrightarrow \bot$ for all $i, j \leq m$ with $i \neq j$
2. $\top \leftrightarrow b_1 \vee b_2 \vee \cdots \vee b_m$
3. For any sentence a of $\mathcal{L}_f$ which is not a contradiction, there exists $\{i_1, \ldots i_k\} \subseteq \{1, 2, \ldots, m\}$ such that $a \leftrightarrow b_{i_1} \vee \cdots \vee b_{i_k}$.

So the b_i's are the most fine-grained description of the "state of the world" that our language can give. Let us call these the *normal atoms* of the language (noting they are not necessarily atomic propositions). We call a sentence *normal* if it is the disjunction of b_i's; it is clear from the above argument that each sentence is equivalent to a normal sentence.

Now suppose we have a Popper function $C : \mathcal{L}_f \times \mathcal{L}_f \to \mathbb{R}$. For any sentences a, b of $\mathcal{L}_f$, $C(a, b)$ is to be interpreted as the probability of a conditional on b. We will assume that the Popper function is *regular* in the sense that the only x's such that $C(., x)$ is the constant function 1 are contradictions (and any atomic proposition is not a contradiction). However, this is not an important constraint as all such sentences that are not contradictions can still be interpreted as the empty event, which will receive probability zero.

We now wish to construct a $PANAP$ space with probability function P which will agree with the Popper function C up to an infinitesimal difference. To define an $PANAP$ space we have seen that (§2) it is sufficient to assign a sample space Ω, a weight function $w : \Omega \to \mathbb{R}^+$ and a directed set Λ of finite subsets of Ω. In order to compare a Popper function defined on a propositional language (where events are just propositions) to an $PANAP$ function defined on a sample space (where events are subsets of that sample space) we must also specify the interpretation of the propositions within the sample space. Then for any sentences a, b of $\mathcal{L}_f$ if $\bar{a}, \bar{b} \subseteq \Omega$ are their interpretations we want P to be such that $st(P(\bar{a}|\bar{b})) = C(a, b)$. Contradictions will be interpreted as the empty set, and tautologies as the full sample space, but for the details we first define the sample space Ω.

We follow McGee's construction at first (McGee, 1994, p. 182), although in slightly different terms. Define a_0 to be the normal equivalent of a tautology. Set a_{k+1} to be the disjunction of all b_i's such that b_i logically entails a_k and $C(b_i, a_k) = 0$. Set $rk(C)$ to be the largest n such that a_n is defined (i.e., where there are normal atoms b_i with $C(b_i, a_{n-1}) = 0$). To see that $rk(C)$ is well defined, note that as there are only finitely many normal atoms, and Popper functions are finitely additive, each a_k must be the disjunction of strictly fewer normal atoms than the previous one. We call $rk(C)$ the *rank* of the Popper function Van Fraassen (1976) and also set, for each b_i, $rk(b_i)$ to be the least k such that $C(b_i, a_k) > 0$.

We now depart from McGee to carry out the construction of the $PANAP$ space. Importantly, as mentioned above, we need to give a sample space which is not simply the normal atoms of the language. Instead we will interpret each normal atom by an infinite subset of the sample space. We also specify the weight function and a directed set, and this triple will then yield the $PANAP$ function. After this, we need to check that the probabilities assigned agree with the Popper function up to infinitesimals. The construction

https://doi.org/10.1017/S1755020318000060 Published online by Cambridge University Press

of the $PANAP$ function is fairly straightforward, and it is intuitive to see why it will work, though the proof is not quick.

Sample space: We set $\Omega = \aleph_0$.

Interpretation: For the interpretation of $\mathcal{L}_f$ it is enough to assign each b_i to a subset of Ω, as then the interpretation of any proposition a of $\mathcal{L}_f$ will just be the union of the interpretations of each b_i that compose the normal equivalent of a. To do this we assign each b_i to a set $\bar{b}_i$ such that $|\bar{b}_i| = \aleph_0$, each $\bar{b}_i$ is disjoint from the other $\bar{b}_i$'s, and $\bigcup\{\bar{b}_i : i \in \{1, 2, \ldots m\}\} = \Omega$ (so all the points in the sample space are assigned to some b_i).

Weight function: We define $w : \Omega \to \mathbb{R}^+$, by

$$w(x) = C(b_i, a_k) \text{ where } x \in \bar{b}_i \text{ and } k = rk(b_i).$$

Note that this is well defined as for $x \in \Omega$ there is a unique b_i such that $x \in \bar{b}_i$, and by definition of $rk(b_i)$ we have $w(x) > 0$.

Directed set: To show agreement with the Popper function, we will use a directed set Λ such that all $\lambda \in \Lambda$ satisfy the following two properties:

$$\forall i, j \in \{1, \ldots m\}\ rk(b_i) = rk(b_j) \to |\bar{b}_i \cap \lambda| = |\bar{b}_j \cap \lambda| \tag{7}$$

$$\forall i, j \in \{1, \ldots m\}\ rk(b_i) < rk(b_j) \to |\bar{b}_i \cap \lambda| > |\bar{b}_j \cap \lambda|^2. \tag{8}$$

Property (7) will ensure that conditional probabilities using propositions that are all of the same rank agree with the Popper function (in fact they will be the same, not just infinitesimally close), and property (8) will ensure that propositions with a lower rank will dominate, the squaring here makes sure that those of a higher rank are not only smaller in probability, but infinitesimal so.

LEMMA 3.3. *We can find $\Lambda \subset \mathcal{P}_{fin}(\Omega)$ such that properties (7) and (8) hold for every $\lambda \in \Lambda$.*

Proof. This is easy as we can just take

$$\Lambda = \{\lambda \in \mathcal{P}_{fin}(\Omega) : \lambda \text{ satisfies (7) and (8)}\}.$$

Then Λ is directed: any $x \in \Omega$ can be incorporated into such a set so $\bigcup \Lambda = \Omega$, and for any $\lambda_1, \lambda_2 \in \Lambda$ it is straightforward to expand $\lambda_1 \cup \lambda_2$ to satisfy (7) and (8). □

Now we have defined an $PANAP$ space (up to the choice of ultrafilter $\mathcal{U}_\Lambda$, but any choice will have the properties we need) and so we have specified an $PANAP$ function $P : \mathcal{P}(\Omega) \to \mathbb{R}^*$. So now we need to show that P, the $PANAP$ function, agrees with C, the Popper function, up to the standard part. This takes rather more work than one might expect, becoming fairly technical but the intuition is in 7 and 8 above.

We start with a simple case where we are just dealing with normal atoms of the same rank.[14]

[14] Henceforth we just use, e.g., b for $\bar{b}$, which will be unambiguous.

https://doi.org/10.1017/S1755020318000060 Published online by Cambridge University Press

LEMMA 3.4. *For any normal atom b_i and $b = b_i \vee b_{j_1} \vee \cdots \vee b_{j_k}$ with $rk(b_i) = rk(b_{j_1}) = \cdots = rk(b_{j_k})$ we have*

$$P(b_i|b) = C(b_i, b). \tag{9}$$

So in this case the value given by the $PANAP$ function is actually the same as that given by the Popper function, not just infinitesimally close.

Proof. We have for all $\lambda \in \Lambda$

$$\begin{aligned}
\frac{P(b_i \cap \lambda)}{P(b \cap \lambda)} &= \frac{\sum_{x \in b_i \cap \lambda} w(x)}{\sum_{x \in b \cap \lambda} w(x)} && \text{by definition of } P \\
&= \frac{|b_i \cap \lambda| C(b_i, a_{rk(b_i)})}{\sum_{b_j \in b} |b_j \cap \lambda| C(b_j, a_{rk(b_i)})} && \text{by definition of } w \\
&= \frac{C(b_i, a_{rk(b_i)})}{\sum_{b_j \in b} C(b_j, a_{rk(b_i)})} && \text{by (7)} \\
&= \frac{C(b_i, a_{rk(b_i)})}{C(b, a_{rk(b_i)})} = C(b_i, b) && \text{by (2) and (3) of Popper functions.}
\end{aligned}$$

This property is then preserved in the Ω-limit. □

We are now ready to show we have agreement up to an infinitesimal in the general case:

THEOREM 3.5. *For any $a, b \in \mathcal{L}$ we have*

$$C(a, b) = st\left(\frac{P(\bar{a} \cap \bar{b})}{P(\bar{b})}\right). \tag{10}$$

Proof. Fix $a, b \in \mathcal{L}$. We have

$$P(a|b) = \frac{P(a \cap b)}{P(b)} = \frac{\sum_{b_i \in a \cap b} P(b_i)}{P(b)} = \sum_{b_i \in a \cap b} \frac{P(b_i)}{P(b)}$$

as P is finitely additive.

Let d be the minimum rank of all the normal atoms in b. We will see that only the normal atoms of this rank are relevant for the standard part. Set b_d to be the union of all the normal atoms b_j of rank d and $b_{>d}$ to be the union of all those with rank greater than d, so $b = b_d \cup b_{>d}$. Now we have for b_i in b:

$$P(b_i|b) = \frac{P(b_i)}{P(b_d)} \frac{P(b_d)}{P(b)}.$$

Now for any $\lambda \in \Lambda$, taking k to be the number of b_i's in $b_{>d}$, by property (8) of Λ we have

$$k^2 |b_d \cap \lambda| > |b_{>d} \cap \lambda|^2 \Rightarrow k^2 |b \cap \lambda| > |b_{>d} \cap \lambda|^2.$$

Now by property (7) and the definition of the weight function we have

$$\frac{P(b_{>d} \cap \lambda)}{P(b \cap \lambda)} = \frac{|b_{>d} \cap \lambda| \sum_{b_i \in b_{>d}} C(b_i, a_{rk(b_i)})}{|b_d \cap \lambda| \sum_{b_i \in b} C(b_i, a_{rk(b_i)})}$$

but

$$M = \frac{\sum_{b_i \in b_{>d}} C(b_i, a_{rk(b_i)})}{\sum_{b_i \in b} C(b_i, a_{rk(b_i)})}$$

https://doi.org/10.1017/S1755020318000060 Published online by Cambridge University Press

is constant for any λ so

$$\frac{P(b_{>d} \cap \lambda)}{P(b \cap \lambda)} = M \frac{|b_{>d} \cap \lambda|}{|b_d \cap \lambda|} < Mk^2 \frac{|b_{>d} \cap \lambda|}{|b_{>d} \cap \lambda|^2} = Mk^2 \frac{1}{|b_{>d} \cap \lambda|}.$$

Thus we have for any $n \in \mathbb{N}$ we can choose $\mu \in \Lambda$ such that

$$\forall \lambda \in \Lambda \text{ with } \lambda \supseteq \mu \quad \frac{P(b_{>d} \cap \lambda)}{P(b \cap \lambda)} < 1/n.$$

By preservation to the Λ limit (fineness) this implies $st(\frac{P(b_{>d})}{P(b)}) = 0$ and so $st(\frac{P(b_d)}{P(b)}) = 1$ by standard laws of probability.

Then

$$st(P(b_i|b)) = st \frac{P(b_i)}{P(b_d)} \frac{P(b_d)}{P(b)} = st \frac{P(b_i)}{P(b_d)} st \frac{P(b_d)}{P(b)} = st(P(b_i|b_d)).$$

Going back to the general case,

$$P(a|b) = \sum_{b_i \in a \cap b} \frac{P(b_i)}{P(b)} = \sum_{b_i \in a \cap b_d} \frac{P(b_i)}{P(b)} + \sum_{b_i \in a \cap b_{>d}} \frac{P(b_i)}{P(b)}$$

and the latter term is infinitesimal, as it is less than $\frac{P(b_{>d})}{P(b)}$. Now we can apply Lemma 3.4: taking the standard part (and noting the sums here are all finite) we get

$$st(P(a|b)) = st \sum_{b_i \in a \cap b_d} \frac{P(b_i)}{P(b)} = st \sum_{b_i \in a \cap b_d} \frac{P(b_i)}{P(b_d)} = \sum_{b_i \in a \cap b_d} C(b_i, b_d).$$

Now $C(b_i, b_d) = C(b_i, b)$ as by axiom 3. of Popper functions $C(b_i, b) = C(b_i, b_d)$ $C(b_d, b)$ and $C(b_d, b) = 1$. So we must have

$$\sum_{b_i \in a \cap b_d} C(b_i, b_d) = \sum_{b_i \in a \cap b_d} C(b_i, b) = C(a, b).$$

Putting this all together we get

$$st(P(a|b)) = C(a, b)$$

as required. □

Thus part 1 of Theorem 3.1 is proved.

We now describe a slight extension of this construction which we will use to construct an $PANAP$ space for an infinite language. Let $\mathcal{L}_i$ be a finitely generated language with $\{b_1 \ldots b_m\}$ the normal atoms and $(\Omega_i, w_i, \Lambda_i)$ be the $PANAP$ space generated as above. Let $\bar{a}^i \subset \Omega_i$ be the interpretation of the proposition a of $\mathcal{L}_i$. Suppose we have a different language $\mathcal{L}_p$ (which may not be finitely generated, and may include some of the propositions of $\mathcal{L}$) interpreted on a domain Ω_p (which may be uncountable), and let $\bar{a}^p \subset \Omega_p$ be the interpretation of the proposition a of $\mathcal{L}_p$. Suppose we also have a weight function $w_p : \Omega_p \to \mathbb{R}^+$.

We take $\mathcal{L}$ to be the language generated by the atomic propositions of $\mathcal{L}_i$ and $\mathcal{L}_p$ augmented by the atomic proposition p not in either $\mathcal{L}_i$ or $\mathcal{L}_p$ (p stands for 'previous'), and define a new $PANAP$ space as follows. We set $\Omega = \Omega_i \cup \Omega_p$ (assuming Ω_i and Ω_p are disjoint, otherwise we could make them so be introducing an index), and the give the elements the same weighting as before so $w(x) = w_i(x)$ if $x \in \Omega_i$ and $w(x) = w_p(x)$ if $x \in \Omega_p$. We interpret p as $\Omega_p \subset \Omega$ and any other atomic a in $\mathcal{L}$ as $\bar{a}^i \cup \bar{a}^p$ (where

https://doi.org/10.1017/S1755020318000060 Published online by Cambridge University Press

$\bar{a}^x = \emptyset$ if $a \notin \mathcal{L}_x$). Note that for each of the normal atoms b_js of $\mathcal{L}_i$ there is a sentence $b'_j = b_i \wedge \neg p$ of $\mathcal{L}$ which has the interpretation $\bar{b}_j$ as before.

We want $st(P(p|a)) = 0$ if a is a noncontradictory proposition of $\mathcal{L}_f$, so we treat p as having a rank greater than any of the normal atoms of $\mathcal{L}_i$: so we can define a directed set Λ on Ω in the same way as in Lemma 3.3, such that Λ satisfies (7) and (8) as above, and in addition $|\bar{b}_j \cap \lambda| > |\bar{p} \cap \lambda|^2$ for any normal atom b_j of $\mathcal{L}_i$.

Let P be the $PANAP$ function defined by (Ω, w, Λ). It then follows that for any a and b in $\mathcal{L}_i$ we have $st(P(a|b)) = C(a, b)$. To see this we can just treat p as an extra normal atom and the argument before will go through, as any a or b in $\mathcal{L}_i$ that is not a contradiction will be implied by one of the b_i's, so $st(P(a)) = st(P(a \wedge \neg p))$ and $st(P(b)) = st(P(b \wedge \neg p))$. Thus,

$$st(P(a|b)) = st(\frac{P(a \wedge p) + P(a \wedge \neg p)}{P(b \wedge p) + P(b \wedge \neg p)}) = st(\frac{P(a \wedge \neg p)}{P(b \wedge \neg p)}) = C(a, b).$$

3.3. Infinitely generated languages. In this section we demonstrate how to extend the above result to an arbitrary language $\mathcal{L}$. McGee uses an ultraproduct construction to do this, but his straightforward approach[15] will not work here. The problem is that taking an ultraproduct will yield a weight function taking values in a nonstandard extension of the real numbers, which is not how $PANAP$ works.[16] The construction given here is rather different, although ultraproduct techniques are employed. Readers who are not familiar with ultrafilters and ultraproducts may skip this proof, as it is somewhat technical, and the basic idea is really the same is taking an ultraproduct: we combine, with some interesting mathematics, the models for finitely generated sublanguages into a model for the whole language.

Let $\mathcal{L}$ be an infinitely generated propositional language and C a Popper function on $\mathcal{L}$. Let $\kappa = |\mathcal{L}|$ and $\langle i_\alpha : \alpha < \kappa \rangle$ be a wellordering of all finite systems of atomic sentences from $\mathcal{L}$. For each $\alpha < \kappa$ let $\mathcal{L}_{i_\alpha}$ be the language generated by the atomic propositions in i_α, so each finitely generated sublanguage of $\mathcal{L}$ is just some $\mathcal{L}_{i_\alpha}$.
It is easy to see that for every for every finitely generated sublanguage $\mathcal{L}_{i_\alpha}$ we can construct is an $PANAP$ space $(\Omega_{i_\alpha}, w_{i_\alpha}, I_{\Lambda_{i_\alpha}})$ which satisfies *standard part agreement* (10)—we simply apply the method of the previous section for the language $\mathcal{L}_{i_\alpha}$ and the Popper function C restricted to this sublanguage. We now build a new $PANAP$ space from these.

The sample space: set

$$\Omega = \{\langle x, \alpha \rangle : x \in \Omega_{i_\alpha}\} = \bigcup_{\alpha < \kappa} (\Omega_{i_\alpha} \times \{\alpha\}).$$

Essentially Ω is just the union of all the Ω_i's—the ordered pair construction is just to make sure the elements for different sublanguages are kept distinct.

The interpretation: For any proposition a in $\mathcal{L}$ the interpretation of a is $\{\langle x, \alpha \rangle : a \in \mathcal{L}_{i_\alpha} \wedge x \in \bar{a}_{i_\alpha}\}$ where $\bar{a}_i$ is the interpretation of a in our model for $\mathcal{L}_i$. Note that this is consistent with Ω being the interpretation of a tautology.

[15] The ultraproduct construction in itself is not exactly simple, but its application is well developed, and a standard technique in model theory.

[16] There is no prima facie reason why $PANAP$ should not be extended to allow for this, after all the same additivity principle could be introduced in such a space, so that the probability values would take values in a further extension of the real numbers. However, how we could get the infinite additivity properties of $PANAP$ in the ultraproduct is not obvious.

https://doi.org/10.1017/S1755020318000060 Published online by Cambridge University Press

The weight function: For $\langle x, \alpha \rangle \in \Omega$ set $w(\langle x, \alpha \rangle) = w_{i_\alpha}(x)$. Clearly this is well defined on all of Ω and always strictly positive.

A directed set?

Here we must switch our approach from simply using a directed set, where any ultrafilter including that directed set will give the desired properties, to actually using the ultrafilter directly.

For each $\alpha < \kappa$ we will define an $PANAP$ space $(\Omega_\alpha, w_\alpha, \Lambda_\alpha)$, distinct from $(\Omega_{i_\alpha}, w_{i_\alpha}, \Lambda_{i_\alpha})$ above, except for in the first case where we set $(\Omega_0, w_0, \Lambda_0) = (\Omega_{i_0}, w_{i_0}, I_{\Lambda_{i_0}})$. For $\alpha > 0$ we construct the $PANAP$ space as in the extension at the end of the previous section. Take $\mathcal{L}_i, \Omega_i$, etc. there as $\mathcal{L}_{i_\alpha}, \Omega_{i_\alpha} \times \{\alpha\}$, etc.

We take $\mathcal{L}_{p_\alpha}$ etc. to encompass the previous finite sublanguages, being the language generated by all the atomic propositions of the $\mathcal{L}_{i_\beta}$ for $\beta < \alpha$. Set $\Omega_{p_\alpha} = \bigcup_{\beta<\alpha} \langle \Omega_{i_\beta} \times \{\beta\} \rangle$, the interpretation of any a in $\mathcal{L}_{p_\alpha}$ just as on Ω but restricted to Ω_{p_α}, and $w_{p_\alpha} = w \restriction \Omega_{p_\alpha}$.

Then following the earlier construction we have $\Omega_\alpha = \Omega_{i_\alpha} \cup \Omega_{p_\alpha} = \bigcup_{\beta \leq \alpha} \langle \Omega_{i_\beta} \times \{\beta\} \rangle \subset \Omega$ and $w_\alpha = w \restriction \Omega_\alpha$. Set Λ_α to be the directed set as constructed there and take $\mathcal{U}_\alpha$ to be a corresponding ultrafilter. Note that the probability function P_α generated by the $PANAP$ space $(\Omega_\alpha, w_\alpha, \mathcal{U}_\alpha)$ will satisfy *standard part agreement* (10) for any propositions from $\mathcal{L}_{i_\alpha}$, although not necessarily for all propositions in $\mathcal{L}_{i_\alpha} \cup \mathcal{L}_{p_\alpha}$.

We construct the ultrafilter for our full $PANAP$ space from these $\mathcal{U}_\alpha$ together with an ultrafilter on κ. Let $\tilde{\mathcal{U}}$ be an ultrafilter on κ such that for each atomic proposition a from $\mathcal{L}$, $\{\alpha \in \kappa : a \in \mathcal{L}_{i_\alpha}\} \in \tilde{\mathcal{U}}$[17] and note that all *end-seqments* $\{\alpha \in \kappa : \alpha > \beta\}$ are in $\tilde{\mathcal{U}}$. This is because, as $\mathcal{L}$ has κ many atomic propositions, for any $\beta < \kappa$ there is some proposition $a \in \mathcal{L}$ which is not in any $\mathcal{L}_{i_\alpha}$ for $\alpha < \beta$.

Now we can define the ultrafilter $\mathcal{U}$ on $\mathcal{P}_{fin}(\Omega)$ which we will use for our $PANAP$ space. First to ease notation, for $X \subset \mathcal{P}_{fin}(\Omega)$ set $X_\alpha := X \cap \mathcal{P}_{fin}(\Omega_\alpha)$. Define $\mathcal{U} \subset \mathcal{P}(\mathcal{P}_{fin}(\Omega))$ by

$$X \in \mathcal{U} \leftrightarrow \{\alpha : X_\alpha \in \mathcal{U}_\alpha\} \in \tilde{\mathcal{U}}.$$

LEMMA 3.6. *$\mathcal{U}$ is a nonprincipal ultrafilter, and $\mathcal{U}$ is fine.*

Proof.

1. **Supersets**
   Let $Y \subseteq \Omega$, $X \subseteq Y$. Then we have $Y_i \supseteq X_i$ for all α so
   $$X \in \mathcal{U} \Rightarrow \{\alpha : X_\alpha \in \mathcal{U}_\alpha\} \in \tilde{\mathcal{U}} \Rightarrow \{\alpha : Y_\alpha \in \mathcal{U}_\alpha\} \in \tilde{\mathcal{U}} \Rightarrow Y \in \mathcal{U},$$
   where the first and third implications are simply by definition, and the second as each $\mathcal{U}_\alpha$ is an ultrafilter.
2. **Intersection**
   Let $X, Y \in \mathcal{U}$. We have $(X \cap Y)_\alpha = X_\alpha \cap Y_\alpha$ so:
   $$\{\alpha : X_\alpha \in \mathcal{U}_\alpha\}, \{\alpha : Y_\alpha \in \mathcal{U}_\alpha\} \in \tilde{\mathcal{U}} \Rightarrow \{\alpha : X_\alpha \in \mathcal{U}_\alpha \wedge Y_\alpha \in \mathcal{U}_\alpha\} \in \tilde{\mathcal{U}}$$
   $$\Rightarrow \{\alpha : X_\alpha \cap Y_\alpha \in \mathcal{U}_\alpha\} \in \tilde{\mathcal{U}} \Rightarrow X \cap Y \in \mathcal{U},$$

[17] This is possible as such sets have the *finite intersection property* and so can be extended to an ultrafilter Goldblatt (1998).

https://doi.org/10.1017/S1755020318000060 Published online by Cambridge University Press

where the first implication is because $\tilde{\mathcal{U}}$ is an ultrafilter and the second because all the $\mathcal{U}_\alpha$'s are.

3. **Ultra**
First note that $\Omega_\alpha \setminus X_\alpha = (\Omega \setminus X)_\alpha$. Thus
$$X \in \mathcal{U} \Leftrightarrow \{\alpha : X_\alpha \in \mathcal{U}_\alpha\} \in \tilde{\mathcal{U}} \Leftrightarrow \{\alpha : X_\alpha^c \notin \mathcal{U}_i\} \in \tilde{\mathcal{U}}$$
$$\Leftrightarrow \{\alpha : X_\alpha^c \in \mathcal{U}_\alpha\} \notin \tilde{\mathcal{U}} \Leftrightarrow X^c \notin \mathcal{U}.$$
4. **Non-principal**
It is enough to show that no finite set is in $\mathcal{U}$ (see Goldblatt (1998), p. 38). But this is clear as X finite $\Rightarrow \{\alpha : X_\alpha \neq \emptyset\}$ is finite $\Rightarrow X \notin \mathcal{U}$.
5. **Fine**. Let $x \in \Omega$. Then there is some $\alpha \in \kappa$ such that $x \in \Omega_{i_\alpha}$, so then $x \in \Omega_\beta$ for all $\beta \geq \alpha$. Let $X = \{y \in \mathcal{P}_{fin}(\Omega) : x \in y\}$. Then $X_\beta = \{y \in \mathcal{P}_{fin}(\Omega_\beta) : x \in y\}$ so as each $\mathcal{U}_\beta$ is fine, for all $\beta \geq \alpha$ we have $X_\beta \in \mathcal{U}_\beta$ and thus $X \in \mathcal{U}$. □

Now we have defined an $PANAP$ space $(\Omega, w, \mathcal{U})$, so we set $P : \mathcal{P}_{fin}(\Omega) \to \mathbb{R}^*$ to be the corresponding $PANAP$ function. It remains to show that for any a, b from $\mathcal{L}$ we have $st(P(a|b)) = C(a, b)$.

To do this we exploit an alternative interpretation of the $PANAP$ function generated from above. In the section on $PANAP$ we defined $P(a)$ to be the ultrafilter equivalence class of sequence $\langle P(a|\lambda) : \lambda \in \mathcal{P}_{fin}(\Omega)\rangle$, but we can also think of $P(a)$ as the ultraproduct of $P_\alpha(a)$ under the ultrafilter $\tilde{\mathcal{U}}$. To see this is equivalent, first we need to see that the hyperreal fields generated by the two processes are the same. But this is straightforward: For any functions $\varphi, \psi : \mathcal{P}_{fin}(\Omega) \to \mathbb{R}$ we have
$$\begin{aligned} \varphi \approx_{\mathcal{U}} \psi &\Leftrightarrow \exists Q \in \mathcal{U}, \forall \lambda \in Q,\ \varphi(\lambda) = \psi(\lambda) \\ &\Leftrightarrow \exists \tilde{Q} \in \tilde{\mathcal{U}}, \forall \alpha \in \tilde{Q}, \exists X \in \mathcal{U}_\alpha, \forall \lambda \in X,\ \varphi(\lambda) = \psi(\lambda) \\ &\Leftrightarrow \exists \tilde{Q} \in \tilde{\mathcal{U}}, \forall \alpha \in Q, \varphi \approx_{\mathcal{U}_\alpha} \psi \Leftrightarrow [\varphi] \approx_{\tilde{\mathcal{U}}} [\psi], \end{aligned}$$
where we take $[\varphi](\alpha) = [\varphi]_{\mathcal{U}_\alpha}$, so $[\varphi]$ is the function taking α to the equivalence class of φ under $\mathcal{U}_\alpha$. It is not hard to see that this reasoning shows the hyperreal fields generated by the two processes are isomorphic under the obvious mapping. Thus it makes sense to ask whether
$$P(a) = [P_\alpha(a)]_{\tilde{\mathcal{U}}}.$$
This does in fact hold as by definition of $\tilde{\mathcal{U}}$ we have $\{\alpha : a \in \mathcal{L}_{i_\alpha}\} \in \tilde{\mathcal{U}}$ and for any α with a in $\mathcal{L}_\alpha$ the definition of the $PANAP$ function gives $P_\alpha(a|\lambda) = P(a|\lambda) = \frac{\Sigma_{x \in a \cap \lambda}(w(x))}{\Sigma_{x \in \lambda}(w(x))}$, so these are the same function on a set in the ultrafilter $\tilde{\mathcal{U}}$.

We have gone to all this trouble because in order to show P agrees with the Popper function C we will use the fact that the functions P_α agree with C, and then, using the latter presentation, apply Łos' Theorem, which states that any first-order sentence ϕ that holds in ultrafilter-many models will also be true in the ultraproduct model. Now for the detail.

Fix a, b in $\mathcal{L}$ and set $r = C(a, b)$. Let the first order (not propositional) language $\mathcal{L}'$ include the language of real analysis plus constant terms '$p_{a \wedge b}$', 'p_b' and 't_r', and the unary predicate N. We define models $\mathcal{M}_\alpha$: set the domain D_α to be the nonstandard reals produced by the ultrafilter $\mathcal{U}_\alpha$. Give the language of real analysis its normal interpretation over the hypereals, and set $\mathfrak{I}_\alpha(t_r) = r$, $\mathfrak{I}_\alpha(N) = \mathbb{N}$ (the true natural numbers, not their nonstandard extension under $\mathcal{U}_\alpha$) and where $a, b \in \mathcal{L}_{i_\alpha}$ set $\mathfrak{I}_\alpha(p_{a \wedge b}) = P_\alpha(a \cap b)$ and

https://doi.org/10.1017/S1755020318000060 Published online by Cambridge University Press

$\mathfrak{I}_\alpha(p_b) = P_\alpha(b)$ and be arbitrary otherwise. Now for each α such that $a, b \in \mathcal{L}_{i_\alpha}$ we have from our construction of P_α that $st(P_\alpha(a|b)) = r$, so for such α's we have

$$\mathcal{M}_\alpha \models \text{“}\forall n \in N \left| \frac{p_{a\wedge b}}{p_b} - t_r \right| < 1/n\text{”}.$$

Now set $\mathcal{M}$ to be the model attained by taking the ultraproduct of the models M_α under the ultrafilter $\tilde{\mathcal{U}}$. As $\{\alpha < \kappa : a, b \in \mathcal{L}_{i_\alpha}\}$ is in $\tilde{\mathcal{U}}$, Łos' Theorem tells us that the above sentence interpreted in the ultraproduct model will also hold, i.e.,

$$\mathcal{M} \models \text{“}\forall n \in N \left| \frac{p_{a\wedge b}}{p_b} - t_r \right| < 1/n\text{”}.$$

What is the ultraproduct model? The members of the domain are the $\tilde{\mathcal{U}}$ equivalence classes of sequences of objects $\langle a : a \in D_\alpha\rangle$, so by the earlier discussion these are exactly the hyperreals generated by $\mathcal{U}$. The interpretation on constants in $\mathcal{L}'$ is straightforward, $\mathfrak{I}(t_r) = [\langle \mathfrak{I}_\alpha(t_r) : \alpha < \kappa\rangle]_{\tilde{\mathcal{U}}}$ etc. So again by the earlier discussion we see that $\mathfrak{I}(p_b) = [P_\alpha(b)]_{\tilde{\mathcal{U}}} = P(b)$ and $\mathfrak{I}(p_{a\wedge b}) = [P_\alpha(a \cap b)]_{\tilde{\mathcal{U}}} = P(a \cap b)$. The interpretation of N in the ultraproduct model will be a nonstandard extension of the natural numbers, but importantly it will include all of $\mathbb{N}$ so we can conclude that

$$\forall n \in \mathbb{N}, \left| \frac{P(a \wedge b)}{P(b)} - r \right| < 1/n.$$

In other words,

$$st\left(\frac{P(a \wedge b)}{P(b)}\right) = r = C(a, b).$$

So we're done.

§4. Non-Archimedean probabilities and lexicographic probabilities. In §3.1 we noted that for finitely generated languages $\mathcal{L}_i$, any value $Pr_i(A)$ of a non-Archimedean probability function can be given a lexicographic representation as a finite sequence of classical real-valued probability values. In the literature on lexicographical probabilities, generalised probabilities are sometimes considered that represent probabilities as *ω-sequences* of classical real-valued probability values.[18] The naive conjecture that non-Archimedean probability functions for (*countably*) *infinitely* generated languages can be represented as an ω-sequence of classical real-valued probability values has been shown to be incorrect Examples 4.8, 4.10.

The problem is, roughly, one of nonwellfoundedness.[19] We have seen in §3.1 that the terms in the polynomial expression of $Pr_i(A)$ represents a 'level of infinitesimality'. But (as we shall shortly see), the collection of ranks for an infinitely generated language does not in general form a well-ordering. So we will represent $PANAP$ probability values lexicographically as *non-well-ordered* sequences of real numbers. We will concentrate on the simplified case of $PANAP$ functions that have no associated weight function, or, equivalently, for which the associated weight function is constant 1.

[18] For a discussion of the theory of lexicographic probabilities, see Blume *et al.*, (1991).

[19] Halpern briefly discusses the idea of nonwellfounded lexicographical probability functions in (Halpern, 2010, p. 165), and dismisses it.

https://doi.org/10.1017/S1755020318000060 Published online by Cambridge University Press

4.1. Extending van Fraassen's notion of rank. First we extend van Fraassen's definition of the notion of rank to $PANAP$ functions P.

Let F be the non-Archimedean field of which $ran(P)$ is a substructure. Then:

DEFINITION 4.1. $\forall a, b, \in F : a \approx_{rk} b \equiv \exists r \in F : \infty > |st[r]| \neq 0$ *and* $a = r \cdot b$.

So numbers in the field F are of the same rank if they are not infinitely small or infinitely large with respect to each other.

It is immediate that $\approx_{rk}$ is an equivalence relation. So we define *ranks* as equivalence classes of $\approx_{rk}$:

DEFINITION 4.2. $\forall a \in F : rk(a) \equiv [a]_{\approx_{rk}}$, *and* $\mathcal{R} \equiv \{rk(a) : a \in F\}$.

Thus ranks can be seen as *locally Archimedean* substructures of the non-Archimedean field F.

$\mathcal{R}$ is a generalisation of van Fraassen's notion of rank to infinitely generated languages. The elements of $\mathcal{R}$ are linearly ordered in a natural way (induced by the linear ordering on F): the *higher* the rank of a number, the larger its 'degree of infinitesimality', and the probability value $0 \in F$ can be seen as the unique element of F of *maximal rank* $\underline{0}$. But this natural ordering on $\mathcal{R}$ is not in general a well-ordering.

We define the rank $rk(A)$ of an *event* $A \subseteq \Omega$ as $rk(P(A))$.

Now we arbitrarily choose, for each $\underline{\alpha} \in \mathcal{R}$, a positive *rank unit value* $1_{\underline{\alpha}}$ of rank $\underline{\alpha}$.

Let an $PANAP$ function P be given. We want to define a lexicographical representation of P. For every $A \subseteq \Omega$, we want to define a lexicographial ordering $<_L$ such that

$$\forall A, B \subseteq \Omega : P(A) < P(B) \Leftrightarrow P(A) <_L P(B).$$

4.2. Transfinite sums of elements of F. The idea is to approximate $P(A)$ by means of a well-ordered (and generally transfinite) sequence of *approximations*. The definition of these approximations involves transfinite sums of elements of F. So for our construction we will need a notion of sum of elements of F that makes sense also for all transfinite α. We define an appropriate notion of sum using a second ultrafilter construction (recognising that F itself was already generated by an ultrafilter construction).

Let μ be an ordinal that is chosen (with foresight) to be large enough to enumerate the stages of approximation of elements of F.

DEFINITION 4.3.

$$S_\alpha \equiv \{S \in [\mu]^{<\omega} : \alpha \leq min(S)\}$$

$$\mathbb{S} \equiv \{S_\alpha : \alpha < \mu\}.$$

Clearly $\mathbb{S}$ has the finite intersection property. So let U^* be an ultrafilter on $[\mu]^{<\omega}$ extending $\mathbb{S}$. Then U^* can be taken to determine an appropriate notion of sum in the following way.

We define sums $\sum_{\alpha<\beta} f(\alpha)$, with $\beta \leq \mu$, inductively. So we assume $\sum_{\alpha<\beta_0} f(\alpha)$ to be defined already for all $\beta_0 < \beta$, and define $\sum_{\alpha<\beta} f(\alpha)$.

DEFINITION 4.4. *Let f be any function from β to F. For any $S \in [\beta]^{<\omega}$ with $\beta_0 = min(S)$, let*

$$f(S) \equiv \sum_{\alpha<\beta_0} f(\alpha) + \sum_{\alpha \in S \setminus \{\beta_0\}} f(\alpha).$$

https://doi.org/10.1017/S1755020318000060 Published online by Cambridge University Press

Then $f(S)$ is a *finite* sum of elements of F, which is of course well-defined because F is a field.

Now we identify modulo agreement on the ultrafilter U^*:

DEFINITION 4.5. *For any functions f, g from β to F:*

$$f \sim g \equiv \{S \in [\mu]^{<\omega} : f(S) = g(S)\} \in \mathcal{U}^*.$$

This partitions the functions f from μ to F into equivalence classes $[f]_{\mathcal{U}^*}$, and they form a non-Archimedean field F^* into which F is canonically embedded in the same way as $\mathbb{R}$ is embedded in F.

Now we set:

DEFINITION 4.6.

$$\sum_{\alpha<\beta} f(\beta) \equiv \lceil [f]_{\mathcal{U}^*} \rceil_F,$$

where $\lceil a^* \rceil_F$ is the unique element $a \in F$ that is closest to the element $a^* \in F^*$.

Not all such sums will be well-defined. Intuitively, it may be the case that $\mid [f]_{\mathcal{U}^*} \mid$ is "infinitely large" with respect to all elements of F, in the same way that some elements of F are "infinitely large" with respect to all elements of $\mathbb{R}$. But if $[f]_{\mathcal{U}^*}$ is bounded from below and from above by elements of F as canonically embedded in F^*, then the sum *is* well-defined, because of transfer. The argument goes as follows. Using the $*$-notation from nonstandard analysis, we move from the $\mathbb{R}$ to the non-Archimedean field $\mathbb{R}^* = F$ and then to $\mathbb{R}^{**} = F^*$. Then using this $*$-notation, we know from the completeness of the real field that

$$\forall r^* \in \mathbb{R}^* \exists r \in \mathbb{R} \forall s \in \mathbb{R} : s \neq r \rightarrow \mid r^* - s \mid > \mid r^* - r \mid .$$

By $*$-transfer, this yields

$$\forall r^{**} \in \mathbb{R}^{**} \exists r^* \in \mathbb{R}^* \forall s^* \in \mathbb{R}^* : s^* \neq r^* \rightarrow \mid r^{**} - s^* \mid > \mid r^{**} - r^* \mid .$$

4.3. Approximations. Now we are ready to define the approximations. Let a $PANAP$ function P and an event $A \subseteq \Omega$ be given.

The idea is to approximate $P(A)$ as follows. $P(A)$ is of a certain rank $rk(A)$: call this rank $\underline{0}^A$ ("the rank of the 0th approximation of A"). So we will in a first stage approximate $P(A)$ by the element $A_0 \cdot 1_{\underline{0}^A}$ where $A_0 \in \mathbb{R}$ is such that $A_0 \cdot 1_{\underline{0}^A}$ is closest to $P(A)$. But then it is likely that there is a nonzero *remainder* $P(A) - A_0 \cdot 1_{\underline{0}^A}$. This remainder will then be of a *higher* rank $\underline{1}^A$ than $\underline{0}^A$. So we will approximate the remainder by the element of the form $A_{\underline{1}^A} \cdot 1_{\underline{1}^A}$, with $A_{\underline{1}^A}$ the unique real number such that $A_{\underline{1}^A} \cdot 1_{\underline{1}^A}$ is closest to the remainder. This will leave us with a remainder of a still higher rank. Thus we continue into the transfinite.

The details of the construction go as follows.

We define the remainders and the approximating real numbers inductively. We take them to be defined for all $\alpha < \beta$, and then first define the remainder of stage β:

DEFINITION 4.7.

$$A^r_\beta \equiv P(A) - \sum_{\alpha<\beta} A_\alpha \cdot 1_{\underline{\alpha}^A}.$$

This remainder will be of a certain rank:

https://doi.org/10.1017/S1755020318000060 Published online by Cambridge University Press

DEFINITION 4.8.

$$1_{\underline{\beta}^A} \equiv rk(A^r_\beta).$$

On the basis of this, we then define the *approximation* of $P(A)$ at stage β:

DEFINITION 4.9. *A_β is the unique $r \in \mathbb{R}$ such that $r \cdot 1_{\underline{\beta}^A}$ is closer to A^r_β than any number of the form $s \cdot 1_{\underline{\beta}^A}$ for $s \in \mathbb{R}$ such that $s \neq r$.*

Of course these definitions are only well-formed if the sums involved (Definition 4.7) are well-defined. But this is the case:

PROPOSITION 4.10. *For all $\beta < \mu$, A^r_β is well-defined.*

Proof. By transfinite induction on β. We use boundedness considerations that hold for ultrafilter-large families of finite sets and are then globally preserved. □

PROPOSITION 4.11. $A^r_{\beta+1} = A^r_\beta - A_\beta \cdot 1_{\underline{\beta}^A}$.

Proof. The reason is that $S_\beta \in \mathcal{U}^*$ (degenerate case). □

LEMMA 4.12. *If $rk(A^r_\alpha) \neq \underline{0}$, then $\alpha < \beta \leq \mu \Rightarrow rk(A^r_\alpha) < rk(A^r_\beta)$.*

Proof. Induction on β.
1. $\beta = \gamma + 1$. By the previous proposition and the definition of A_γ we have $rk(A^r_{\gamma+1}) > rk(A^r_\gamma)$. The result then follows by the induction hypothesis.
2. $Lim(\beta)$. Let $\alpha < \beta$. Then by the induction hypothesis we have $rk(A^r_\alpha) < rk(A^r_{\alpha+1})$. So it suffices to show that $rk(A^r_\beta) \geq rk(A^r_{\alpha+1})$.

Since $S_{\alpha+1} \in \mathcal{U}^*$, we want to show that for each $S \in S_{\alpha+1}$, we have

$$rk(P(A) - \sum_{\gamma \in S} A_\gamma \cdot 1_{\underline{\gamma}^A}) \geq rk(P(A) - \sum_{\gamma < \alpha+1} A_\gamma \cdot 1_{\underline{\gamma}^A}).$$

We only need to look at those S such that $\min(S) < \beta$, since only they can contribute to the sums. By the induction hypothesis, for such S we have

$$rk(P(A) - \sum_{\gamma < \min(S)} A_\gamma \cdot 1_{\underline{\gamma}^A}) \geq rk(P(A) - \sum_{\gamma < \alpha+1} A_\gamma \cdot 1_{\underline{\gamma}^A}).$$

This entails that indeed

$$rk(P(A) - \sum_{S} A_\gamma \cdot 1_{\underline{\gamma}^A}) \geq rk(P(A) - \sum_{\gamma < \alpha+1} A_\gamma \cdot 1_{\underline{\gamma}^A}).$$

□

So there must for simple cardinality reasons be an ordinal ζ such that $rk(A^r_\zeta) = \underline{0}$. Then ζ is called the *closure ordinal* for A. In a similar vein, the closure ordinal of P, denoted as $cl(P)$, is defined as

$$max\{\alpha : \alpha \text{ is the closure ordinal of some } A \subseteq \Omega\}.^{20}$$

This means that the $PANAP$ functions on Ω yield an *ordinal spectrum* that is determined by their closure ordinals.

[20] In fact, we have a notion of closure ordinal for each $r \in F$. So we could also define the closure of P as $max\{\alpha : \alpha$ is the closure ordinal of some $r \in F\}$.

https://doi.org/10.1017/S1755020318000060 Published online by Cambridge University Press

THEOREM 4.13.

$$P(A) = \sum_{\alpha < cl(A)} A_\alpha \cdot 1_{\underline{\alpha}^A}.$$

Proof. Consider the closure ordinal ζ of A, which must exist by Lemma 4.12. Then we have

$$P(A) - \sum_{\beta < \zeta} A_\beta \cdot 1_{\underline{\beta}^A} = 0.$$

□

Of course we can then also express $P(A)$ as $\sum_{\alpha < cl(P)} A_\alpha \cdot 1_{\underline{\alpha}^A}$.

The infinite sum $\sum_{\alpha < cl(A)} A_\alpha \cdot 1_{\underline{\alpha}^A}$ can be seen as a kind of *Cantor normal form* for $P(A)$.[21] The upshot of our discussion is that even though the structure $\mathcal{R}$ of the ranks is nonwellfounded, every probability value can be expressed as a well-founded infinite sum of components of increasing rank.

4.4. Representation theorem. The infinite sum $\sum_{\alpha < cl(A)} A_\alpha \cdot 1_{\underline{\alpha}^A}$ can be seen as a *lexicographic presentation* of $P(A)$. It can be used to define a lexicographical ordering:

DEFINITION 4.14 (Lexicographic order). $P(A) <_L P(B) \equiv$ *for the smallest* α *such that* $A_\alpha \cdot 1_{\underline{\alpha}^A} \neq B_\alpha \cdot 1_{\underline{\alpha}^B}$, *we have* $A_\alpha \cdot 1_{\underline{\alpha}^A} < B_\alpha \cdot 1_{\underline{\alpha}^B}$.

It is immediate that $<_L$ is a strict linear ordering.

Now we are ready to prove the main theorem of this section:

THEOREM 4.15 (Representation theorem).

$$\forall A, B \subseteq \Omega : P(A) < P(B) \Leftrightarrow P(A) <_L P(B).$$

Proof.

We know from Theorem 4.13 that $P(A) = \sum_{\alpha < cl(P)} A_\alpha \cdot 1_{\underline{\alpha}^A}$ and $P(B) = \sum_{\alpha < cl(P)} B_\alpha \cdot 1_{\underline{\alpha}^B}$.

($\Rightarrow$) We are given that $P(A) < P(B)$. Let α be the first ordinal where $P(A)$ and $P(B)$ differ. Then we want to show that

$$A_\alpha \cdot 1_{\underline{\alpha}^A} < B_\alpha \cdot 1_{\underline{\alpha}^B}.$$

Suppose, for a contradiction, that this is not the case, i.e., that

$$A_\alpha \cdot 1_{\underline{\alpha}^A} > B_\alpha \cdot 1_{\underline{\alpha}^B}.$$

We will show that then, for all $S \in S_\alpha$:

$$\sum_S A_\alpha \cdot 1_{\underline{\alpha}^A} > \sum_S B_\alpha \cdot 1_{\underline{\alpha}^B}.$$

We aim to show this by an induction on $\min(S)$.

1. Suppose $\alpha = \min(S)$. Then the property holds by Lemma 4.12.
2. Suppose the property holds for all $S \in S_\alpha$ such that $\min(S) < \beta$. Then we want to show that the property also holds for all $S \in S_\alpha$ such that $\min(S) = \beta$.

[21] In a somewhat related (but also significantly different) context, the connection between a non-Archimedean notion of size and Cantor normal forms is explored in (Benci, Di Nasso, & Forti, 2006, sec. 1.5).

https://doi.org/10.1017/S1755020318000060 Published online by Cambridge University Press

2a. Suppose $\beta = \gamma + 1$. For A, we know that

$$\sum_{\kappa<\beta} A_\kappa \cdot 1_{\underline{\kappa}^A} = \sum_{\kappa<\gamma} A_\kappa \cdot 1_{\underline{\kappa}^A} + A_\gamma \cdot 1_{\underline{\gamma}^A},$$

and similarly for B. So the property follows by the induction hypothesis and Lemma 4.12.
2b. $Lim(\beta)$. It suffices to show

$$\sum_{\kappa<\beta} A_\kappa \cdot 1_{\underline{\kappa}^A} > \sum_{\kappa<\beta} B_\kappa \cdot 1_{\underline{\kappa}^B}.$$

But by the induction hypothesis, this holds on all $S \in S_\alpha$, and $S_\alpha \in \mathcal{U}^*$, so this indeed holds also.

From this inductive argument we conclude that for all $S \in S_\alpha$,

$$\sum_S A_\alpha \cdot 1_{\underline{\alpha}^A} > \sum_S B_\alpha \cdot 1_{\underline{\alpha}^B}.$$

So, since $S_\alpha \in \mathcal{U}^*$,

$$P(A) = \sum_{\alpha<cl(P)} A_\alpha \cdot 1_{\underline{\alpha}^A} > \sum_{\alpha<cl(P)} B_\alpha \cdot 1_{\underline{\alpha}^B} = P(B),$$

which gives us the required contradiction.
($\Leftarrow$) This follows by a similar argument. It is given that $P(A) <_L P(B)$. Then there is a first α such that

$$A_\alpha \cdot 1_{\underline{\alpha}^A} < B_\alpha \cdot 1_{\underline{\alpha}^B}.$$

Then we argue inductively that

$$\sum_{\alpha<cl(P)} A_\alpha \cdot 1_{\underline{\alpha}^A} < \sum_{\alpha<cl(P)} B_\alpha \cdot 1_{\underline{\alpha}^B},$$

and we are done. □

Note that this implies that even though a choice of ultrafilter $\mathcal{U}^*$ was needed to define $<_L$, the resulting ordering is invariant with respect to this choice of ultrafilter.

4.5. Discussion. In sum, PANAP functions can be represented as (transfinite) *well-founded* sums of terms that represent nonzero contributions of different ranks. Somewhat surprisingly, perhaps, this does not contradict the nonwellfoundedness of the rank ordering.

Let a sample space Ω be given. Then an ultrafilter on $\mathcal{P}_{fin}(\Omega)$ determines a $PANAP$ function P (with 'uniform weight 1'). Different such $PANAP$ functions may have different closure ordinals in the sense of §4.3. Intuitively, $PANAP$ functions with larger closure ordinals may be taken to be 'more complicated' than $PANAP$ functions with smaller closure ordinals. Thus analysing the closure ordinals of $PANAP$ functions would yield a *classification* or *spectral analysis* of $PANAP$ functions (with uniform weight 1, on a fixed sample space). Thus it might be worthwhile to undertake a spectral analysis of $PANAP$ functions.

§5. Closing. The results in this article build on and extend known results.

McGee (1994) and Halpern (2010) relate non-Archimedean probability functions to Popper functions. However, they do not show that all Popper functions, the countably additive ones as well as those that are not countably additive, are approximated by regular

https://doi.org/10.1017/S1755020318000060 Published online by Cambridge University Press

perfectly additive non-Archimedean probability functions. But Theorem 3.1 shows that this is the case, whereby there is a sense in which even Popper functions that are only finitely additive can be taken to be infinitely additive. In this way, Theorem 3.1 goes beyond the results in the literature.

Halpern relates finitely generated non-Archimedean probability functions to lexicographic probability functions Halpern (2010). Halpern observes that the task of finding for every non-Archimedean probability function an equivalent lexicographical probability function, fails (Halpern, 2010, p. 165, Example 4.8). He then asks whether, given a suitable definition of countable additivity for non-Archimedean probability functions,[22] we can find a lexicographical equivalent for every non-Archimedean probability function, and concludes that this is not the case (Halpern, 2010, p. 165, Example 4.10). Nonetheless, Theorem 4.15 shows how any PANAP function can be represented using the (*nonwell-founded*) field of ranks of that probability function. So again we have obtained a stronger result.

Representation Theorem 4.15 is related to a very general representation theorem that is announced in Pedersen (2014, sec. 6). In this article, Pedersen extends De Finetti's fundamental theorem of comparative expectations to expectation orderings that satisfy a version of the *principle of weak dominance* (rather than uniform simple dominance). He then announces that such systems of comparative expectations (finite and infinite) can be represented by an expectation function that takes values in a non-Archimedean field in which in which every number can be written as formal well-founded power series in a single infinitesimal. This means, of course, that the elements in such a field are lexicographically ordered.

Representation theorems like those in McGee (1994), Van Fraassen (1976), Krauss (1968), Leitgeb (2012), Halpern (2010), together with the two representation theorems in the present article give us reasons to think that Popper functions capture a robust concept of probability. Popper functions cannot do anything that $PANAP$ functions and lexicographical probability functions can't do. So if we believe any of these model our intuitive notion of probability well, then Popper functions must too. Though they may not fully cash out all our intuitions, for example about infinite additivity, they cannot go against them to any greater extent than these alternatives.

Also, these representation theorems, which relate all the different models via Popper functions, give us reasons to believe that they all correctly represent our intuitions of probability, for the same reason that the equivalence of the different mathematical attempts at describing what an algorithm is gives us reason to believe we have captured that notion: there are many ways to be wrong about something, so it would seem unlikely that every time we attempted it we were wrong *in the same way*.

So we have better reasons to believe that each of these theories for probability really do model an intuitive concept. $PANAP$ gives us both a notion of how probabilities for events depend on the probabilities of the individual outcomes that make up those events, and also relates the conditional probability to the ratio of absolute probabilities in the familiar, intuitive way. So Popper functions are perhaps closer to our intuitions concerning probability than their initial axiomatic presentation may suggest. But non-Archimedean probability functions in general, and $PANAP$ functions in particular, are themselves not completely intuitive. One source of unintuitiveness is the nonwellfoundedness of the degrees of infinitesimality that such probability functions entail. This seems to open the

[22] See (Halpern, 2010, p. 166).

https://doi.org/10.1017/S1755020318000060 Published online by Cambridge University Press

prospect that the possibility of one event can be smaller than that of another event even though there is no largest degree of infinitesimality at which they differ. Nonetheless, this turns out not to be the case. Our second representation theorem (§4.4) shows that every $PANAP$ function value can be represented as a well-founded power series of which each term represents the contribution of a specific degree of infinitesimality.

Our conclusion is then that Popper functions, non-Archimedean probability functions, and lexicographical functions cohere well and mutually support each other. The representation theorems from the literature and from this article show that most of the uses to which any one of these notions of generalised probability been put, can also be seen as applications of the other two notions of generalised probability.

We do not claim that PANAP is the *only* notion of non-Archimedean probability that can conceptually be supported by relating it via representation theorems to other concepts of probability such as (but not only) lexicographic probability and Popper functions. In particular, an interesting avenue for future research would be to enquire whether p-adic probability can be related via representation theorems to other conceptions of probability.

§6. Acknowledgments. The authors are indebted to Stanislav Speranski, Paul Pedersen, Karim Thebault, Sylvia Wenmackers, and two anonymous referees for extremely helpful comments and suggestions. The first author's research was in part supported by an EPSRC doctoral fellowship.

BIBLIOGRAPHY

Benci, V., Di Nasso, M., & Forti, M. (2006). An aristotelian notion of size. *Annals of Pure and Applied Logic*, **143**, 43–53.

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

Benci, V., Horsten, L., & Wenmackers, S. (2018). Infinitesimal probabilities. *The British Journal for the Philosophy of Science*, **69**(2), 509–552.

Blume, L., Brandenburger, A., & Dekel, E. (1991). Lexicographical probabilities and choice under uncertainty. *Econometrica*, **59**, 61–79.

Campbell-Moore, C., Horsten, L., & Leitgeb, H. (forthcoming). Probabilities for the revision theory of truth. *Journal of Philosophical Logic*, to appear.

de Finetti, B. (1949). *On the Axiomatization of Probability*. Translated and published as Chapter 5 in Probability, Induction, and Statistics (1972). Wiley Publications.

Goldblatt, R. (1998). *Lectures on the Hyperreals. An Introduction to Nonstandard Analysis*. Graduate Texts in Mathematics, Vol. 188. New York: Springer-Verlag.

Hajek, A. (2003). What conditional probability could not be. *Synthese*, **137**, 273–323.

Halpern, J. (2010). Lexicographic probability, conditional probability, and non-standard probability. *Games and Economic Behavior*, **68**, 155–179.

Harper, W. (1975). Rational belief change, popper functions and counterfactuals. *Synthese*, **30**, 221–262.

Ilic-Stepic, A., Ognjanovic, Z., & Ikodinovic, N. (2014). Conditional p-adic probability logic. *International Journal for Approximate Reasoning*, **55**, 1843–1865.

Khrennikov, A. (1996). p-adic valued probability measures. *Indagationes Mathematicae*, **7**, 311–330.

Khrennikov, A. (2008). Toward theory of p-adic valued probabilities. *Studies in Logic, Grammar, and Rhetoric*, **14**, 137–154.

https://doi.org/10.1017/S1755020318000060 Published online by Cambridge University Press

Khrennikov, A. & Schumann, A. (2006). Logical approach to p-adic probabilities. *Bulletin of Symbolic Logic*, **45**, 49–57.

Kolmogorov, A. (1933). *Grundbegriffe der Wahrscheinlichkeitrechnung* (Ergebnisse Der Mathematik). Translated by Morrison, N. *Foundations of probability.* Chelsea Publishing Company (1956) 2nd English Edition.

Krauss, P. (1968). Representation of conditional probability measures on boolean algebras. *Acta Mathematica Academiae Scientiarum Hungaricae*, **19**, 229–241.

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

Leitgeb, H. (2012). A probabilistic semantics for counterfactuals. *The Review of Symbolic Logic*, **5**, 26–121.

Levi, I. (1989). Possibility and probability. *Erkenntnis*, **31**, 365–386.

Lewis, D. (1980). A subjectivist's guide to objective chance. In Jeffrey, R. C., editor. *Studies in Inductive Logic and Probability*. Berkeley, CA: University of California Press, pp. 263–293.

McGee, V. (1994). Learning the impossible. In Ellery, E. and Brian, S., editors. *Probability and Conditionals: Belief Revision and Rational Decision*. Cambridge: Cambridge University Press, pp. 179–199.

Nelson, E. (1987). *Radically Elementary Probability Theory*. Princeton, NJ: Princeton University Press.

Norton, J. (2018). How to build an infinite lottery machine. *European Journal for Philosophy of Science*, **8**, 71–95.

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

Popper, K. (1959). *The Logic of Scientific Discovery*. New York: Basic Books.

Pruss, A. (2015). Popper functions, uniform distributions, and infinite sequences. *Journal of Philosophical Logic*, **44**, 259–271.

Renyi, A. (1955). On a new axiomatic theory of probability. *Acta Mathematica Academiae Scientiarum Hungaricae*, **6**, 285–335.

Robinson, A. (1961). Non-standard analysis. *Nederlandse Akademie van Wetenschappen, Series A: Mathematical Sciences*, **64** and *Indagationes Mathematicae*, **23**, 432–440.

Schervish, M., Seidenfeld, T., & Kadane, J. (2017). Non-conclomerability for countably additive measures that are not κ-additive. *Review of Symbolic Logic*, **10**, 284–300.

Schumann, A. (2008). Non-Archimedean fuzzy and probability logic. *Journal of Applied Non-Classical Logics*, **18**, 29–48.

Schumann, A. (2014). Probabilities on streams and reflexive games. *Operations Research and Decisions*, **1**, 71–96.

Van Fraassen, B. (1976). Representation of conditional probabilities. *Journal of Philosophical Logic*, **5**, 417–430.

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

GRADUATE SCHOOL OF ENGINEERING
KOBE UNIVERSITY
1-1 ROKKODAI-CHO
KOBE 657-8501, JAPAN
E-mail: brickhill@dragon.kobe-u.ac.jp

DEPARTMENT OF PHILOSOPHY
UNIVERSITY OF BRISTOL
BRISTOL BS6 6JL, UK
E-mail: leon.horsten@bristol.ac.uk

https://doi.org/10.1017/S1755020318000060 Published online by Cambridge University Press