

UNDERDETERMINATION OF INFINITESIMAL PROBABILITIES

ALEXANDER R. PRUSS

ABSTRACT. A number of philosophers have attempted to solve the problem of null-probability possible events in Bayesian epistemology by proposing that there are infinitesimal probabilities. Hájek (2003) (more tentatively) and Easwaran (2014) (more definitively) have argued that because there is no way to specify a particular hyperreal extension of the real numbers, solutions to the regularity problem involving infinitesimals, or at least hyperreal infinitesimals, involve an unsatisfactory infidelity or arbitrariness. The arguments depend on the alleged impossibility of picking out a particular hyperreal extension of the real numbers and/or of a particular value within such an extension due to the use of the Axiom of Choice. However, it is false that the Axiom of Choice precludes a specification of a hyperreal extension—such an extension can indeed be specified. Moreover, for all we know, it is possible to explicitly specify particular infinitesimals within such an extension. Nonetheless, I prove that because any regular probability measure that has infinitesimal values can be replaced by one that has all the same intuitive features but other infinitesimal values, the heart of the arbitrariness objection remains.

1. INFINITESIMAL PROBABILITIES AND UNDERDETERMINATION

Bayesian epistemology faces a difficulty with the possibility of events that seem to have zero probability, such as a randomly aimed dart hitting the exact center of a target or a particular ticket in a countably infinite fair lottery being chosen. For updating one's credence for an event A on a zero-probability event B requires calculating the conditional probability $P(A | B)$ which is classically defined as $P(A \& B)/P(B)$, which in turn requires dividing by zero. One family of solutions to the problem involves Popper functions—i.e., taking conditional probabilities to be fundamental and not defined as ratios of unconditional probabilities. The other family of solutions involves extending the real number line which is classically used to provide the values of the probability function to a totally ordered field that includes infinitesimals—quantities α such that $0 < |\alpha| < r$ for every strictly positive real r —and then saying that the events that seem to have zero probability but nonetheless to be possible in fact have infinitesimal probability, and dividing by an infinitesimal to define a conditional probability will be a perfectly well-defined process.

Date: December 2018. Forthcoming in *Synthese*.

In this paper, I will argue that all infinitesimal solutions to the problem of zero-probability events suffer from an underdetermination problem: the extended-real probability functions carry more information than is determined by the plausible kinds of constraints on these probabilities.

As is usual in probability theory, we can identify events with subsets of a sample space Ω —the set of all options. A set \mathcal{F} of subsets of Ω is an *algebra* provided that it is non-empty and closed under complements and finite unions (and hence finite intersections). If \mathbb{R}^* is a totally ordered field extending the real numbers, then an \mathbb{R}^* -valued finitely-additive probability measure (“measure” for short) on \mathcal{F} is a function from \mathcal{F} to \mathbb{R}^* satisfying the axioms:

- (i) $P(A) \geq 0$,
- (ii) $P(\Omega) = 1$, and
- (iii) $P(A \cup B) = P(A) + P(B)$ if A and B are disjoint.

In mathematics, one normally extends (iii) to countable additivity. However, the most popular solutions to the zero probability problem that involve extending the reals do not have a well-defined notion of summing a countably infinite sequence of probabilities, and hence I am limiting the discussion to finite additivity.

The measure P is *regular* provided that $P(A) > 0$ whenever A is non-empty. If there are uncountably many disjoint outcomes (as in the dart case), there will be no regular real-valued measure on an algebra of subsets of the target that includes all these outcomes.¹ And if there are infinitely many disjoint *equal probability* outcomes, then finite additivity will force a real-valued measure to assign probability zero to each outcome.²

Thus with real-valued measures, we lose regularity in at least two kinds of situations: cases where there are uncountably many disjoint outcomes and cases where there are infinitely many (countably or not) equal probability outcomes.

However, for any real-valued measure P on \mathcal{F} , there is a field of hyperreals \mathbb{R}^* and an \mathbb{R}^* -valued regular measure P^* that is infinitesimally close to P , i.e., $P^*(A) - P(A)$ is always zero or infinitesimal.³ For convenience, I will

¹Let the E_i for $i \in I$ be the disjoint outcomes where I is uncountable. For each integer $n > 0$, let $I_n = \{i \in I : 2^{-n} < P(E_i) \leq 2^{-n+1}\}$. If the set I_n has at least 2^n members, then there will be 2^n disjoint outcomes each with probability bigger than 2^{-n} , and their union by finite additivity will then have a probability greater than $2^n \cdot 2^{-n} = 1$, which is impossible. So, each set I_n has less than 2^n members, and in particular is finite. But I is the union of the I_n , since for any $i \in I$, there is an integer $n > 0$ such that $2^{-n} < P(E_i) \leq 2^{-n+1}$, as $P(E_i) > 0$. Hence I is a union of countably many finite sets, and hence is countable, contrary to the assumption.

²For suppose that the probability of each one of these disjoint equal probability outcomes is $\alpha > 0$. Then for any finite n , the probability of the disjunction of n of them will be $n\alpha$ by finite additivity. But for any positive real number α we will have $n\alpha > 1$ if n is sufficiently large, whereas all probabilities must be between 0 and 1.

³E.g., Benci, Horsten and Wenmackers (2018) show that there is a regular measure Q on \mathcal{F} taking values in a field of hyperreals. Let α be any infinitesimal in that field, and

write $x \approx y$ to mean that $x - y$ is zero or an infinitesimal. Note that hyperreals are always an extension of the reals.

Moreover, there is a sense in which anything you can do with Popper functions, you can do with infinitesimals, in the sense that if $P(\cdot | \cdot)$ is a Popper conditional probability function (see McGee 1994 for definitions) then there is a field of hyperreals \mathbb{R}^* and an \mathbb{R}^* -valued measure P^* such that $P(A | B) \approx P(A \cap B)/P(B)$ whenever B is non-empty (see Krauss 1968, McGee 1994 and Brickhill and Horsten 2016). Thus, any assignment of conditional probabilities that is consistent in the sense that it can be extended to a Popper function can be modeled by hyperreal-valued probabilities.

There are a number of problems with the infinitesimal probability solution to the non-regularity problem. Regular infinitesimal probabilities fail to preserve intuitive symmetries (Williamson 2007⁴; Pruss 2013b). The same problem occurs for Popper functions, however (Pruss 2015). Failures of countable additivity in applications to countably infinite fair lotteries lead to paradoxes of non-conglomerability for non-conditional probabilities, whether real- or hyperreal-valued (Pruss 2018), and hence also for Popper functions which agree with non-conditional probabilities when one conditions on the whole space. But of course failures of non-conglomerability just go with the territory of finitely additive functions, so perhaps we need not worry about them here.⁵

There is, however, one problem where infinitesimal probabilities seem to do worse than Popper functions, and it has led some authors to prefer Popper functions. The problem is that infinitesimal probabilities require a field of hyperreals. But hyperreals are obtained by an argument involving using the Axiom of Choice from set theory to prove the existence of an ultrafilter satisfying needed constraints, however without picking out a unique such ultrafilter. There are many ultrafilters, yielding different hyperreal extensions of the reals that can be used to define infinitesimal probabilities, with many choices of value within an extension (cf. Easwaran 2014). All this allows for a multiplicity of assignments of extensions of a real-valued measure to a hyperreal-valued measure that yield, to within an infinitesimal, the same conditional probabilities for Bayesian epistemology. Since it is the conditional probabilities that have epistemological significance, Easwaran (2014) has argued that hyperreal-valued probabilities carry more information than is there in the epistemology.⁶

let $P^*(A) = (1 - \alpha)P(A) + \alpha Q(A)$. Then P^* is a regular measure infinitesimally close to P .

⁴Weintraub (2008) observes that Williamson's symmetries or isomorphism do not preserve a certain global property. This shortcoming of Williamson's example is fixed in Section 3.2 of the present paper.

⁵See also DiBella (2018) on how the phenomenon of non-conglomerability goes beyond numerical probabilities.

⁶Hájek (2003 p. 292) made a somewhat weaker claim.

Even more worryingly, the choice of a hyperreal extension appears to be not only arbitrary but ineffable (Hájek 2003)—we cannot successfully refer to a particular extension, and so a particular extension cannot reflect our credences.⁷

In the next section, I will argue that the ineffability argument does not apply to all extensions of the reals, and even as restricted to the hyperreals it is unsuccessful. But then I will note that there are two levels of underdetermination, and even if there is no underdetermination at the level of choice of extension of the reals, I will prove a result showing that there is underdetermination—but not necessarily ineffability—at the level of the choice of particular infinitesimals in the extension field, no matter how one extends the field of reals. Next, I will argue that adding a certain limit axiom does not solve the uniqueness problem. Finally, I will argue against three defenses of underdetermination offered by an anonymous reader.

Throughout the paper, I am assuming that we are assigning a *single* probability function. An approach in terms of imprecise probabilities, where a *family* of hyperreal-valued probability functions is assigned as in Section 6.1.2 of Benci, Horsten and Wenmackers (2018), may well be able to escape the underdetermination worries.⁸ As such, the paper can also be considered as adding motivation for an imprecise probability approach.⁹

2. SPECIFYING AN EXTENSION

The standard construction of hyperreals makes use of a mathematical object called a free ultrafilter. An *ultrafilter* on an algebra \mathcal{F} of sets can be thought of as classifying which members of \mathcal{F} count as relevant, subject to the axioms that the intersection of a pair of relevant sets is relevant; that a superset of a relevant set is relevant; and that for every set A in \mathcal{F} , either A or its complement is relevant. One trivial way to get an ultrafilter is to distinguish some element a_0 and say that a set is relevant if and only if it contains a_0 . A *free* ultrafilter is any ultrafilter *not* constructed in this way. Any free ultrafilter defines a hyperreal extension of the reals.¹⁰

The existence of hyperreal extensions thus depends on the existence of free ultrafilters. The existence of free ultrafilters on any infinite algebra follows from the Axiom of Choice in set theory, but the proof does not construct a particular ultrafilter—it just shows that there is at least one. And there

⁷Easwaran (2014) gives an elegant argument based on supervenience and the Church-Turing thesis.

⁸I am grateful to an anonymous reader for observing this.

⁹But interestingly perhaps not an imprecise probability approach in terms of *intervals* of hyperreal probabilities. For the transformation in our Theorem 1 may imply arbitrariness in some interval-valued hyperreal accounts.

¹⁰The hyperreals can be taken to be equivalence classes of \mathcal{F} -measurable functions from I to the reals under the equivalence relation \sim where $f \sim g$ if and only if $\{x \in I : f(x) \neq g(x)\}$ is not relevant. We then identify each real number r with the constant function that is equal to r everywhere.

isn't just one, but infinitely many. Thus, we have an underdetermination of the particular extension of the reals.

It is popularly thought that where an existence proof depends essentially on the Axiom of Choice, we cannot uniquely specify a particular one of the items proved to exist. Moreover, assuming that Zermelo-Fraenkel set theory is consistent, the proof of the existence of an ultrafilter on an infinite set is known to *essentially* use some version of the Axiom of Choice (Pincus and Solovay 1977). This leads to the thought that we cannot uniquely specify a hyperreal extension of the reals, and the multiplicity of extensions thus is an objectionable form of underdetermination, since there is no reason to take one or another extension as the home of the correct model of an agent's credences.

This line of argument, however, has multiple difficulties. First, the popular thought about the unspecifiability of objects proved to exist with the Axiom of Choice is known to be mistaken. For instance, say that a set A is a xyzyzy on \mathcal{F} if and only if (a) A is non-empty and (b) every member of A is a free ultrafilter on \mathcal{F} . Then there is a xyzyzy on \mathcal{F} if and only if there is a free ultrafilter on \mathcal{F} . Thus, just as we need the Axiom of Choice to prove the existence of a free ultrafilter on \mathcal{F} , we need it to prove the existence of a xyzyzy on \mathcal{F} . But if the Axiom of Choice is true, then set X of *all* free ultrafilters on \mathcal{F} is non-empty, and hence is a xyzyzy—a xyzyzy that we have explicitly specified. Of course, a xyzyzy is not an ultrafilter. But by leveraging the idea that even when it is difficult to specify a particular ultrafilter, one can specify sets of ultrafilters, Kanovei and Shelah (2004) explicitly defined a particular free ultrafilter on a particular infinite set. The proof that their specification succeeds depends on the Axiom of Choice, but nonetheless the specification is explicit. Furthermore, Kanovei and Shelah used their construction to make an explicitly specified extension of the reals (an iteration of the hyperreal extension using this ultrafilter) having further desirable properties.

Second, one should be satisfied with specifying an *isomorphism class* of hyperreals. After all, Bayesian epistemologists who use classical probability need not be concerned about the fact that there are infinitely many sets isomorphic to the reals that could just as well be used for values of the probability function—the Benacerraf (1965) dilemma is a serious problem in the philosophy of mathematics, but does not seem germane to Bayesian epistemology. The values of the probability function are already just a model of credences, and models need only be picked out up to isomorphism. But it turns out that we can specify a set of hyperreals up to isomorphism. For some cardinals κ , there is a unique-up-to-isomorphism κ -saturated non-standard real line of cardinality κ (Keisler 1994, Section 11). And there might be some non-arbitrary way to choose the cardinal κ , perhaps a way matching the particular problem under discussion.

Third, hyperreals are not the only way to get infinitesimals. There are multiple methods that do not make use of anything like the arbitrary choice

of an ultrafilter. One way is to go for a very large extension of the reals: John Conway's surreals. There are so many surreals that they are a proper class rather than a set, but the class of surreals, as a class, is explicitly specified. Whatever you can do with hyperreals, you can do with surreals, since any field of hyperreals can be embedded in the surreals (Ehrlich 2012).

Or one can go a more modest route and go for formal Laurent series (Laugwitz 1968 and Pedersen MS) or the Levi-Civita field (Shamseddine and Berz 2010). The field of formal Laurent series is the field of all formal sums of the form $\sum_{n \in \mathbb{Z}} a_n \varepsilon^n$, where \mathbb{Z} is the set of integers, each a_n is a real number, only finitely many of the a_n that have $n < 0$ are non-zero, and ε is a formal symbol for an infinitesimal. Similarly, the Levi-Civita field is the field of all formal sums of the form $\sum_{q \in \mathbb{Q}} a_q \varepsilon^q$, where \mathbb{Q} is the rationals, each a_q is a real number, only finitely many of the a_q are non-zero, and ε is a formal symbol for an infinitesimal. These fields are explicitly specified. Here, one does not worry about the meaning of ε , since the members of the fields can just be identified with sequences (a_n) or (a_q) .

Now, admittedly, even if one can specify particular fields containing infinitesimals, there is still the problem of choosing among the rival specified constructions. But here there might be some hope. The surreals have the advantage of being exhaustively large, large enough that they escape the cardinality arguments against regularity of Pruss (2013a). The fields of formal Laurent series and the Levi-Civita field, on the other hand, have the advantage of being elegantly small. (On the other hand, the Kanovei-Shelah field, while mathematically fascinating, probably has little going for it in this context.) So the friend of infinitesimal probabilities has a real hope of non-arbitrarily specifying a particular field of infinitesimals.

3. CHOICE OF PARTICULAR INFINITESIMALS

3.1. Ineffability. But while at the level of the choice of an extension of reals that includes infinitesimals we can escape the arbitrariness objection, there is a lower level where arbitrariness enters in: the choice of particular infinitesimals within the extension to be particular values of the probability function. There are two ways one can raise a concern about this. The stronger way is an ineffability worry. Even if we can refer to a particular extension of the reals, perhaps we cannot refer to a particular infinitesimal member of that extension. The weaker is an underdetermination worry: there is no rational reason to choose a particular infinitesimal member of an extension to be a value for the probability of the function. In the remainder of this subsection I will argue that the ineffability worry has not been successfully established, but in the rest of the paper I will argue that the underdetermination worry is very real.

First, note that if we choose formal Laurent series or the Levi-Civita field, there is no ineffability worry at all. In both cases, ε (say) is a completely specified infinitesimal.

Second, if we insist on the richer mathematical structure of the hyperreals, things become more murky. Gödel proved that if Zermelo-Fraenkel set theory is consistent, then it is also consistent with the Axiom of Constructibility $V = L$. A consequence of the Axiom of Constructibility is that the language of set theory can be used to explicitly define a well-ordering \prec on all sets (Jech, 1997, p. 108). But given this well-ordering, we can define hyperreal extensions *and* particular infinitesimal members of them with great ease. For instance, start with the power set $\mathcal{P}\mathbb{N}$ of the naturals \mathbb{N} . By the Axiom of Choice, there is a non-empty set U of ultrafilters on $\mathcal{P}\mathbb{N}$. But any non-empty well-ordered set has a unique least element, so there is a \prec -least element \mathcal{U}_0 of U , where \prec is our definable well-ordering on the universe of sets. We can then define a hyperreal extension $\mathbb{R}^{\mathcal{U}_0}$ with respect to the ultrafilter \mathcal{U}_0 , and then let α be the \prec -least positive infinitesimal in $\mathbb{R}^{\mathcal{U}_0}$. We have now specified an extension of the hyperreals, $\mathbb{R}^{\mathcal{U}_0}$, and a particular positive infinitesimal in that extension, α . There is no ineffability.

Now, the Axiom of Constructibility may well be false. But it may also be true. If it is true, then even if one insists on hyperreal infinitesimals, ineffability is false: we can refer to a particular infinitesimal. And even if the Axiom of Constructibility is false, it might still be the case that some weaker version of it is sufficient to imply the existence of a definable hyperreal infinitesimal is true. Or at least no argument to the contrary has been offered.

3.2. Intuitions. Prior to giving a theorem embodying a precise underdetermination claim, I will offer some intuitive reasons to think that there is underdetermination.

Fix an extension \mathbb{R}^* of the reals, and suppose Alice assigns an infinitesimal α in \mathbb{R}^* to the hypothesis that a fair coin flipped countably often always landed heads, while Bob assigns 2α to that hypothesis. Both assign consistent finitely additive probabilities. On what grounds could we say that Alice's assignment of α rather than 2α was rationally right—or even better reflective of Alice's doxastic state?

Here is perhaps a more vivid way to see the point. Suppose a fair coin is tossed every day from eternity to eternity. Number the days with integers n , let H_n be the hypothesis that starting with day n , it's all heads for the rest of time, and write $p(n) = P(H_n)$. Then $p(n) = (1/2)p(n+1)$: to get H_n we need to have heads on day n as well as to have H_{n+1} , while the toss on day n is independent of H_{n+1} . It follows that more generally $p(n) = 2^{n-m}p(m)$ for any integers n and m , and hence that if any one of the $p(n)$ is non-zero, they are all non-zero, and yet they are all different.¹¹ (Note that the present example gets around a problem with Williamson's example noted

¹¹Williamson (2007) uses this to argue against infinitesimal probabilities, since by symmetry considerations he considers it absurd that the $P(H_n)$ be all different. But the friend of infinitesimal probabilities rejects this reasoning (see Hofweber 2014, though that paper does concern chances rather than epistemic probabilities).

by Weintraub 2008.) But now let n be the number corresponding to the day you were born. Then $p(n)$ is some infinitesimal α . But why is that particular infinitesimal α assigned to the day you were born rather than, say, to the day Napoleon died? There seems to be no reason to assign α to that day rather than to some other day.

Other problems of choosing infinitesimals arise for uniform distributions. For while in classical probability, once we say that a distribution is uniform, that tends to fix the probability, while this is not the case when we allow infinitesimal probabilities. Consider a spinner designed to stop uniformly at some angle θ_1 from 0° (inclusive) to 360° (exclusive). Now consider a second spinner which functions in exactly the same way, except that once it stops, we double its angle. Thus, if it stops at 15° , we move it to 30° , and if it stops at 350° , we double the angle to 700° , which is equivalent to 340° . Call the initial angle of the second spinner θ_{21} and the final doubled angle θ_{22} . Observe that θ_{22} is just as uniformly distributed over all the angles as θ_{21} is, and that θ_{21} has the same distribution as θ_1 . Let $\alpha = P(\theta_1 = 0^\circ)$. By uniformity we also have $P(\theta_1 = 180^\circ) = \alpha$, and because the two spinners initially function the same way, $P(\theta_{21} = 0^\circ) = P(\theta_{21} = 180^\circ) = \alpha$. But $\theta_{22} = 0^\circ$ if and only if either $\theta_{21} = 0^\circ$ or $\theta_{21} = 180^\circ$. Thus, $P(\theta_{22} = 0^\circ) = 2\alpha$. And by uniformity 2α is the probability of every outcome for θ_{22} . Hence two uniform processes with the same outcome space—angles from 0° to 360° —would have different infinitesimal probabilities for singletons.

This means that we cannot simply specify an infinitesimal by saying that it is whatever is the probability of a uniform process hitting a particular point. For what that probability is—assuming it is infinitesimal and not zero—depends on details of the process that go beyond uniformity. And we have no idea how that dependence would go, and little reason to think there is a fact of the matter.

3.3. Some constraints. Still there are potential types of constraints on the choice of infinitesimals that may help narrow down the choice of infinitesimal probabilities.

One obvious constraint is formal: the axioms of finitely additive probability.

A second type of constraint is match between the extended real probabilities for a problem and the corresponding classical real-valued probabilities. This match, of course, cannot be equality, because we are after all trying to make events whose probability is classically zero have non-zero probability. But we can require the following ordinal-match constraint: if the classical probability of A is strictly less than the classical probability of B , then the extended real probability of A is strictly less than that of B . We can also require a formally stronger match condition, that if the classical probability

is x and the extended real probability is x' , then x and x' differ at most by an infinitesimal.¹²

One way to get the classical real-valued probabilities for applying the second constraint is to make use of the agent's rational preferences and a classical representation theorem. But for particular problems there may be other ways to get to appropriate classical probabilities, such as a uniform probability distribution over a target for a random dart throwing case, or an infinite product of probabilities for an infinite sequence of coin tosses. And there might be some direct access to agent's credences based on their own probabilistic claims.

A third constraint consists in symmetry conditions. For instance, in the case of a symmetric spinner selecting an angle from 0 to 360 degrees, we would like to insist on rotational symmetry. Unfortunately, we know that there are limits on how much symmetry can be imposed in a context with infinitesimals. In the spinner case, for instance, let r be any irrational number, and consider the sets of angles $A_1 = \{r, 2r, 3r, \dots\}$ and $A_2 = \{2r, 3r, \dots\}$ in degrees (with angles wrapping around, so that 361° is the same as 1°). If each singleton has non-zero probability, then by additivity $P(A_1) = P(A_2) + P(\{r\}) > P(A_2)$. But the set A_2 is just a rotation of A_1 by r degrees. Thus there is no way to assign rotation-symmetric probabilities here with non-zero probability being assigned to each singleton (cf. Pruss 2013a). Thus symmetry conditions often need to be approximate, i.e., probabilities being preserved under symmetries *modulo* infinitesimal differences.

Finally, we may wish to require that not only are the above kinds of constraints satisfied by unconditional probabilities, but also by conditional probabilities defined by the ratio formula. Thus, we may make use of the agent's hypothetical or conditional rational preferences to try to tease out real-valued conditional probabilities that our ratio-defined probabilities should approximately match.

3.4. The underdetermination theorem. Unfortunately, no constraints like the above are sufficient to determine infinitesimal probabilities. It turns out that if we have any extended real probability that satisfies constraints like the above, one can construct a different extended real probability that does so as well.

The construction is pretty simple. Given a totally ordered field extension \mathbb{R}^* of the reals and an $x \in \mathbb{R}^*$, if x is finite (i.e., $-y < x < y$ for some real y), we can uniquely express x as $x_1 + x_2$ where x_1 is a real number and x_2 is infinitesimal. Denote the standard and infinitesimal parts of x as $\text{Std } x = x_1$ and $\text{Inf } x = x_2$ (we will prove the existence of these parts in the Appendix).

¹²The ordinal-match constraint follows from this, because if x and y are real numbers such that $x < y$, and $x \approx x'$ and $y \approx y'$, then $x' < y'$, since the distance between two distinct real numbers must be more than infinitesimal.

Now, suppose P is our favorite \mathbb{R}^* -valued probability assignment. For any positive real α , define $P_\alpha(A) = \text{Std } P(A) + \alpha \text{Inf } P(A)$. Observe that if $\alpha \neq 1$, then $P_\alpha(A)$ is different from $P(A)$ whenever $P(A)$ is not a real number. Thus, if P ever takes on a non-real value, P_α and P are different.

But they are also very similar in respect of our constraints. First, note that $P(A)$ and $P_\alpha(A)$ differ by at most an infinitesimal (namely by $(1 - \alpha)\text{Inf } P(A)$). Furthermore:

Theorem 1. *If P is an \mathbb{R}^* -valued probability function and $\alpha > 0$ is real, then:*

- (i) P_α is an \mathbb{R}^* -valued probability function,
- (ii) $P(A) < P(B)$ if and only if $P_\alpha(A) < P_\alpha(B)$,
- (iii) if $P(C) > 0$, then $P(A | C) < P(B | C)$ if and only if $P_\alpha(A | C) < P_\alpha(B | C)$
- (iv) if $a_1, \dots, a_n, b_1, \dots, b_m$ are real, then for any events $A_1, \dots, A_n, B_1, \dots, B_m$ we have $\sum_i a_i P(A_i) < \sum_i b_i P(B_i)$ if and only if $\sum_i a_i P_\alpha(A_i) < \sum_i b_i P_\alpha(B_i)$
- (v) if $P(C) > 0$, and $a_1, \dots, a_n, b_1, \dots, b_m$ are real, then for any events $A_1, \dots, A_n, B_1, \dots, B_m$ we have $\sum_i a_i P(A_i | C) < \sum_i b_i P(B_i | C)$ if and only if $\sum_i a_i P_\alpha(A_i | C) < \sum_i b_i P_\alpha(B_i | C)$
- (vi) if $P(C) > 0$, then $P(A | C) \approx P_\alpha(A | C)$.

By (i), the formal probabilistic constraints are satisfied by P_α .

By (ii), any ordinal constraints that P satisfies will also be satisfied by P_α . Moreover, it follows from (ii) that $P(A) = P(B)$ if and only if $P_\alpha(A) = P_\alpha(B)$. Thus, any symmetry conditions exactly satisfied by P will be exactly satisfied by P_α . Moreover, since P and P_α are always infinitesimally close, any symmetry conditions satisfied by P modulo infinitesimal differences will also be satisfied by P_α modulo infinitesimal differences.

Next, observe that (iv) guarantees that P and P_α yield the same decision-theoretic comparisons between finite gambling portfolios with real-valued utilities, since they yield the same comparisons between expected utilities.

Moreover, (iii) and (v), respectively, show that the equivalences in (ii) and (iv) remain true even if we conditionalize on some event with non-zero \mathbb{R}^* -valued probability. Furthermore, (iii) shows that $P(A | C) = P(B | C)$ if and only if $P_\alpha(A | C) = P_\alpha(B | C)$, which shows that any symmetry constraints that P exactly satisfies conditionally on some subset C will also be satisfied in the same way by P_α .

Finally, (vi) implies that P and P_α generate the same Popper conditional probability function, and that conditional or unconditional symmetry constraints satisfied by P modulo infinitesimal differences are also satisfied by P_α modulo infinitesimal differences.

The proof will be given in the appendix.

Thus the kinds of constraints that we have discussed in the case of P will also be satisfied by an uncountable infinity of other measures taking values in the extended-real field.

3.5. Other putative constraints. One might, however, hope that there are some further constraints one can put on P that will determine it more finely.

We saw that P and P_α yield the same decisions between finite gambling portfolios with real-valued utilities. But what about infinite gambling portfolios or utilities that are not real-valued?

The case of infinite gambling portfolios is difficult to handle for probabilities taking their value in an extension of the reals. A field extension of the reals is guaranteed to have sensible finite sums, but there is in general no well-defined notion of a sum of an infinite sequence of values in an extension \mathbb{R}^* of \mathbb{R} . Hence, if we insist on handling infinite gambling portfolios, we run into technical problems independent of the underdetermination objection.

However, non-real-valued utilities in a finite portfolio could in principle still impose decision-theoretic constraints on probabilities that would distinguish between P_α and P . For instance, suppose that an agent is indifferent between (a) getting one unit of utility with certainty and (b) getting N units of utility, where N is some positive infinite member of \mathbb{R}^* , if a randomly thrown dart lands in the center of the target. Then one can attribute to the agent an infinitesimal credence of $1/N$ that the dart will land there.

However, it is difficult to see how one would go about assigning a particular infinite value in \mathbb{R}^* for a particular infinitely valuable scenario. It is difficult, indeed, to imagine infinitely valuable scenarios, and to assign particular values—say, N rather than $N + 3.4$ —to them seems to be entirely arbitrary.

About the only way we can imagine a particular infinitely valuable scenario seems to be the countably infinite repetition of a finitely valuable scenario.¹³ Perhaps we can imagine eating a particular donut every day for eternity, without ever getting tired of it, and then we might assign to the eternal donut deal a utility equal to the number of days in eternity times the utility of a single donut consumption. But to do that, we would need to assign an infinite quantity, M , in \mathbb{R}^* equal to the number of days in eternity. This, however, faces a surprising technical difficulty. Counts have to be something like natural numbers. We might call them “hypernaturals” when we allow for infinities. But it turns out that for any positive infinite number M in \mathbb{R}^* , if \mathbb{R}^* has a collection of hypernaturals, then there will be *uncountably many* (in external cardinality) hypernaturals between 1 and M (Pruss 2014, Appendix). And so the countable number of future days that we’ve imagined is not what is counted by M : instead, M counts the number of members of the uncountable set $\{1, 2, \dots, M\}$ of hypernaturals. The countable number of future days does not have a hypernatural in \mathbb{R}^* that we can assign, and so in the case of the only infinitely valuable scenario that we can imagine, we cannot assign an appropriate hyperreal value.

¹³Some religions, of course, do claim that union with God is infinitely valuable, even over a short period of time. But religions that hold this are apt to also say that this union is unimaginable to us.

Alternately, we might imagine that an agent has credences that are derived from bets that involve infinitesimal differences in utilities. Thus, if I rationally price a bet on A and a bet on not- A equally, but my utility for A is 10 and my utility for B is $10 + \alpha$ for an infinitesimal α , then my probabilities for A and not- A must respectively be $(10 + \alpha)/(20 + \alpha)$ and $10/(20 + \alpha)$. But again it seems that the choice of the particular infinitesimal α is unconstrained by any realistic rational psychology. Granted, one might reasonably wish to model the fact that the agent slightly, maybe even infinitesimally, prefers B over A . But there seems to be no reason to assign one particular infinitesimal difference in utility rather than another.

3.6. The Ω -limit axiom. Benci, Horsten and Wenmackers (2018) propose an additional axiom for probability functions taking their values in an extension of the reals. According to Hofweber (2014), this axiom is meant to capture the intuition that global probabilities are determined by the probabilities of individual events, in a way that provides a replacement for the countable additivity axiom of classical probability theory (Hofweber is talking of chances, but one may think something similar should apply to epistemic probabilities). The alternate probability function P_α in our underdetermination theorem probability has not been shown to satisfy the Ω -limit axiom.¹⁴

I shall argue that (a) there is a weakening of the axiom that captures the intuition about the determination of global probabilities by individual probabilities to no lesser degree than the original axiom, but the weakened axiom is preserved by the move from P to P_α and (b) there is an underdetermination theorem for the special and unique-up-to-isomorphism extension of the reals considered by Benci, Horsten and Wenmackers' (2013) that preserves the full Ω -limit axiom.

Consider our probability space Ω and suppose that our probability function P is defined for all subsets of Ω . Let Λ be the collection of all the finite non-empty¹⁵ subsets of Ω . Consider the set \mathbb{R}^Λ of all functions from Λ to \mathbb{R} . Then an Ω -limit is a function $L : \mathbb{R}^\Lambda \rightarrow \mathbb{R}^*$ (where \mathbb{R}^* is the extended real field that P takes its values in) such that:

- (i) $L(f + g) = L(f) + L(g)$
- (ii) $L(f \cdot g) = L(f) \cdot L(g)$
- (iii) $L(f) = c$ if f is eventually equal to c in the sense that there is a $\lambda_0 \in \Lambda$ such that if $\lambda \supseteq \lambda_0$, then $f(\lambda) = c$,

where $f + g$ and $f \cdot g$ are the pointwise sum and product of the functions f and g . The Ω -limit axiom then says that our probability function P is such that:

- (a) $P(A \mid \lambda)$ is a real number for all $\lambda \in \Lambda$, and

¹⁴I am grateful to an anonymous referee for pointing out the relevance of the axiom to the project of this paper.

¹⁵Benci, Horsten and Wenmackers (2018) allow for empty subsets, but it simplifies the definitions not to do so, without making any substantive difference.

- (b) there exists an Ω -limit L such that $P(A) = L(P(A | \cdot))$ for all $A \subseteq \Omega$, where $P(A | \cdot)$ is the function in \mathbb{R}^Λ whose value at $\lambda \in \Lambda$ is $P(A | \lambda)$.

Notice that the Ω -limit axiom ensures a way for the probabilities of arbitrary subsets of Ω to depend on the probabilities of finite subsets of Ω , since $P(A | \lambda) = P(A \cap \lambda) / P(\lambda)$ and both $A \cap \lambda$ and λ are finite if $\lambda \in \Lambda$. Indeed it makes the probabilities of arbitrary subsets depend on the probabilities of individual events, since if $\lambda = \{\alpha_1, \dots, \alpha_n\}$ is a finite subset with the α_i distinct, then $P(\lambda) = P(\{\alpha_1\}) + \dots + P(\{\alpha_n\})$ by finite additivity. This is somewhat similar to the way that in a classical countably infinite probability space Ω , for any enumeration $\alpha_1, \alpha_2, \dots$ of Ω , the probability of an arbitrary subset A is equal to $\lim_{n \rightarrow \infty} P(A \cap \Omega_n)$ where $\Omega_n = \{\alpha_1, \dots, \alpha_n\}$, and $P(A \cap \Omega_n)$ is determined by the probabilities of individual events since $P(A \cap \Omega_n) = P(\{\beta_1\}) + \dots + P(\{\beta_k\})$ where $A \cap \Omega_n = \{\beta_1, \dots, \beta_k\}$ for distinct β_i .¹⁶

Note that a consequence of (i)–(iii) is that if $f \geq g$ pointwise (i.e., $f(\lambda) \geq g(\lambda)$ for all $\lambda \in \Lambda$), then $L(f) \geq L(g)$. For let $h(\lambda) = \sqrt{f(\lambda) - g(\lambda)}$ so that $f - g = h \cdot h$. Let $n(\lambda) = -1$ for all $\lambda \in \Lambda$ and observe that by (iii) we have $L(n) = -1$. Then by (i) and (ii):

$$\begin{aligned} L(f) - L(g) &= L(f) + L(n)L(g) \\ &= L(f) + L(n \cdot g) \\ &= L(f + n \cdot g) \\ &= L(f - g) \\ &= L(h \cdot h) \\ &= (L(h))^2 \geq 0, \end{aligned}$$

so $L(f) \geq L(g)$. This gives us an order-preservation property we intuitively want in our Ω -limit axiom, in that we want to make sure that if $P(A | \lambda) \geq P(B | \lambda)$ for all $\lambda \in \Lambda$, then $P(A) \geq P(B)$.

Let us now say that a weak Ω -limit is a function $L : \mathbb{R}^\Lambda \rightarrow \mathbb{R}^*$ that satisfies (i), (iii) and:

- (ii*) If $f \geq g$ pointwise, then $L(f) \geq L(g)$.

And say that P satisfies the weak Ω -limit axiom provided that in the statement of the Ω -limit axiom condition (a) above holds and that (b) holds with “ Ω -limit” replaced by “weak Ω -limit”. The weak Ω -limit axiom makes probabilities of arbitrary subsets be determined by the probabilities of finite subsets just as much as the original Ω -limit axiom did: in both cases there

¹⁶It is however worth noting a difference between the Ω -limit axiom and the limit condition that classical countably infinite probability spaces satisfy. In the classical case, the limit operation $\lim_{n \rightarrow \infty}$ is defined independently of P , and the limit condition holds for any enumeration $\alpha_1, \alpha_2, \dots$ of Ω . But in the Ω -limit axiom, it is only claimed that the limit function L exists for any given P : different probability functions P may have different limit functions L that witness to their satisfaction of the Ω -limit axiom. To me, this seems to make the Ω -limit axiom not very satisfying.

is a function that takes us from the values of $P(A \mid \lambda)$ for $\lambda \in \Lambda$ to the value of $P(A)$. In the weak case, the function may lack the nice property (ii), though it has the weaker replacement (ii*). Recall that according to Hofweber (2014), we needed the Ω -limit axiom to ensure the determination of probabilities of arbitrary subsets by probabilities of individual events. The weak Ω -limit axiom yields such a determination, too. It lacks the nice multiplicative property (ii), but it is still a determination of global probabilities by probabilities of individual events.

And now note that in the setting of our Theorem 1, if P satisfies the weak Ω -limit axiom (but not necessarily the Ω -limit axiom), then P_α satisfies the weak Ω -limit axiom. To see this, let $\phi_\alpha(x) = \text{Std } x + \alpha \text{ Inf } x$ for $x \in \mathbb{R}^*$ and $\alpha \in \mathbb{R}$. Let $L_\alpha(f) = \phi_\alpha(L(f))$. That L_α satisfies (i), (ii*) and (iii) follows immediately from these two lemmas that are proved in the Appendix.

Lemma 1. *The function ϕ_α is \mathbb{R} -linear: $\phi_\alpha(\alpha x + \beta y) = \alpha \phi_\alpha(x) + \beta \phi_\alpha(y)$ for all α and β in \mathbb{R} and x and y in \mathbb{R}^* .*

Lemma 2. *The function ϕ_α is strictly increasing: if $x < y$ are in \mathbb{R}^* , then $\phi_\alpha(x) < \phi_\alpha(y)$.*

Note that $P(A \mid \lambda) = P(A \cap \lambda)/P(\lambda)$ is the ratio of two real numbers by condition (a) in the Ω -limit axiom (which holds unchanged in the weak axiom). Since $P(B) = P_\alpha(B)$ whenever $P(B)$ is real (as $\text{Inf } P(B)$ is zero then), it follows that $P(A \mid \lambda) = P_\alpha(A \mid \lambda)$. For the same reason, P_α satisfies condition (a) of the axiom. Then by the weak Ω -limit axiom:

$$L_\alpha(P_\alpha(A \mid \cdot)) = L_\alpha(P(A \mid \cdot)) = \phi_\alpha(L(P(A \mid \cdot))) = \phi_\alpha(P(A)) = P_\alpha(A),$$

and so P_α satisfies both conditions in the axiom.

Next, observe that Benci, Horsten and Wenmackers' (2013, Section 4.2) initial constructions of spaces satisfying the full Ω -limit axiom have not been shown to involve a unique extension of the real numbers, as they depend on an arbitrary choice of a fine ideal. Thus, these constructions do not provide a solution to the first underdetermination problem, the one concerning the choice of the extended real field. Benci, Horsten and Wenmackers (2013) do, however, suggest that one could work with a κ -saturated¹⁷ hyperreal line of cardinality κ for a certain large cardinal κ (Section 4.5), which as noted earlier is unique up to isomorphism (Keisler 1994, Section 11), and which they say would be sufficient for any case "one may wish to consider in applied mathematics". Unfortunately, if one moves to such a field, the second underdetermination problem—the choice of specific non-standard real probabilities—returns. The following is well-known to those in the field, but I have not been able to find an explicit proof in the literature, so a proof is sketched in the Appendix:

Proposition 1. *Suppose \mathbb{R}^* is a κ -saturated hyperreal line with cardinality κ and suppose that $\kappa > \|\mathbb{R}\|$. Then for any $z \in \mathbb{R}^* - \mathbb{R}$, there is a field*

¹⁷See Sacks (2011).

isomorphism ϕ of \mathbb{R}^* onto itself such that $\phi(x) = x$ for all real x and $\phi(z) \neq z$.

But then we have this (also proved in the Appendix):

Theorem 2. *Suppose \mathbb{R}^* is a hyperreal line and let P be an \mathbb{R}^* -valued probability function satisfying the Ω -limit axiom. Let ϕ be a field isomorphism of \mathbb{R}^* onto itself such that $\phi(x) = x$ for all real x . Let $P_\phi(A) = \phi(P(A))$. Then P_ϕ is an \mathbb{R}^* -valued probability function satisfying the Ω -limit axiom and conditions (i)–(vi) of Theorem 1 hold with P_ϕ in place of P_α .*

Thus, if $P(A)$ is a non-standard real z , we can choose ϕ such that $\phi(z) \neq z$, and then P_ϕ will differ from P at least at A , but will nonetheless satisfy all the same axioms, including the Ω -limit axiom, and will satisfy the conditions (i)–(vi) of Theorem 1.

Hence, in the one case Benci, Horsten and Wenmackers (2013) consider where the extended field is uniquely determined, the choice of probabilities is not uniquely determined, even given the strong Ω -limit axiom.

4. THREE DEFENSES OF UNDERDETERMINATION

An anonymous reader has offered three arguments that the kind of underdetermination that I identify might be quite innocent.

4.1. Colors and mathematical constructions. Models have semantically irrelevant features. The reader offered the example a subway map that could indicate the same route with different colors. Perhaps even more tellingly, a mathematical model of some aspect of reality could use any of a host of different set-theoretic constructions of numbers—the natural number 2 could be identified with the set $\{\emptyset, \{\emptyset\}\}$, or the set $\{\{\emptyset\}\}$, etc.¹⁸—without this making any difference to how well the model represents what is being modeled. Indeed, in this way, even classical real-valued probabilities are subject to underdetermination, as it is underdetermined which of the equivalent constructions of real numbers (say, Dedekind cuts, lower parts of Dedekind cuts, upper parts of Dedekind cuts, equivalence classes of Cauchy sequences, infinitely long decimals, infinitely long binary numbers, etc., which in turn can be layered on different constructions of natural numbers, sequences, etc.) is to be used for the values of the probabilities. The typical scientist does not care about such questions: as long as the maps or constructions are isomorphic, the specific choice makes no difference.

The arbitrariness argument in Theorem 1 seems to be more of the same. We have a bijection $\phi_\alpha(x) = \text{Std } x + \alpha \text{Inf } x$ of the non-standard reals between 0 and 1 onto themselves, which bijection can be composed with one assignment P of probabilities to obtain another assignment P_α . The bijection itself is a kind of isomorphism: it preserves sums, multiplication by real numbers and order, as we saw in Lemmas 1 and 2. What difference is

¹⁸Cf. Benacerraf 1965

there between this and the recoloring isomorphism of subway maps or the isomorphism between different equivalent constructions of the reals?

Here is one difference. In the sciences, we do not compare real numbers from different constructions, nor do we in practical life choose a subway route on the basis of the color of the map when that color is semantically insignificant.¹⁹ But it could make important sense to compare different probability functions in situations where infinitesimals are introduced. For instance, a common use for infinitesimals is modeling countably infinite fair lotteries.²⁰ Consider, then, a choice which of two lotteries to play. The prize is the same in each. The lottery is only run if you choose to play it. In each case, you will get a single ticket with the same fixed positive integer, let's say seven. The mechanism for the lotteries works as follows. In the first lottery, there is an infinite urn containing balls numbered $1, 2, 3, \dots$ from which somehow one ball is uniformly chosen, and the winning ticket is the one whose number equals the number printed on the ball. The second lottery involves an exactly similar urn, but the winning ticket number is given by taking the number printed on the ball, dividing by two and rounding up (e.g., if ball 13 comes up, then the winner is given by rounding 6.5 up to 7). If we keep to classical real-valued probabilities, we will say that in each case the chance of winning is zero, so it doesn't matter which lottery you play. But intuitively the second lottery is better, since in the first lottery you win if the ball numbered 7 comes out of the urn, while in the second you win if either 13 or 14 comes out of the urn. Thus intuitively you have a double chance of winning in the second lottery, and hence it's a better choice.

Each lottery's winning number is governed by its own probability distribution. Let's say that the two probability distributions are $P^{(1)}$ and $P^{(2)}$. The difference between $P^{(1)}(\{7\})$ and $P^{(2)}(\{7\})$ (given that you've been told you will get ticket number seven) really matters. But if there is no semantically significant difference between a probability distribution P and the modified distribution P_α , then the difference between $P_\alpha^{(1)}(\{7\})$ and $P^{(1)}(\{7\})$ won't matter. But take $\alpha = 2$. Since ϕ_2 doubles infinitesimal arguments, we then have $P_2^{(1)}(\{7\}) = 2P^{(1)}(\{7\})$, which in turn is intuitively equal to $P^{(2)}(\{7\})$. Hence, if the difference between $P_\alpha^{(1)}(\{7\})$ and $P^{(1)}(\{7\})$ does not matter, neither does the difference between $P^{(2)}(\{7\})$ and $P^{(1)}(\{7\})$. But that difference does actually matter to decision theory: we want to play the lottery where the chance of winning is bigger.

¹⁹Of course, it need not be insignificant. Express routes, for instance, could be marked in red.

²⁰Admittedly, there are serious issues in how to make an infinite lottery work (see Norton 2017, Norton and Pruss 2017 and Pruss 2018). But here I am simply using the lotteries as an intuition pump about the importance of assignments of particular infinitesimals.

The lesson from the lottery case is that it is not only the comparisons between values of the same probability function that matter, but also comparisons between values of different probability functions that matter, and hence the transformation ϕ_α does actually make a difference. The case where different constructions of real numbers are used for probability assignments or different colors are used by different maps is disanalogous, for there is a natural and non-arbitrary way to translate between the constructions or colors.²¹

Moreover, if it turns out that the exact values of the infinitesimals are insignificant artifacts of a hyperreal model of probabilities, we have good reason to prefer a model that lacks such artifacts, such as a Popper function or a comparative (or “qualitative”) probability model. Or at least proponents of hyperreal models owe us a defense of an account of exactly which features of the model correspond to reality (for instance, in a model that uses real numbers, we might reasonably say that the features of the model that correspond to reality are those that are preserved by field isomorphism).

4.2. Benacerraf. The Thomson’s Lamp paradox supposes a lamp that starts (say) off, but its switch is toggled infinitely often during a finite interval of time. The paradox is supposed to be that we have no good answer to the question of whether after the interval of time the lamp is on or off (Thomson 1954). Benacerraf (1962) argues that there is no real paradox, but simply an insufficiency of information. The setup tells us what happens after a finite number of switch togglings, but says nothing about what happens after an infinite number. Both the answer that the lamp is on and that the lamp is off at the end of the interval are logically compatible with the story, but to know which answer is true, we would need more information, say about some law of nature specifying what happens in infinite switching situations. This is a plausible response to the paradox.²² By analogy, one might think that more information is needed about a particular infinite probabilistic setup to determine the specific infinitesimal probabilities, but that this is not a problem.

This response, however, is not plausible in cases where the infinite probabilistic setup is generated by an infinite number of independent classical setups, as in the case of the doubly infinite sequence of coin tosses. We would expect the chancy character of the system to be determined by the classical probabilities of each of the individual setups together with a statement that the setups are independent. There is nothing more to be said about the chanciness of the n th coin toss X_n than the classical statement

²¹For any two constructions of the real numbers, there is only one isomorphism between them that preserves the arithmetical operations. This follows from the fact that the only field automorphism of the reals is the identity (Robert 2000, p. 54).

²²Though Pruss (2018) has recently argued that this response may not be fully satisfactory.

$P(X_n = H) = P(X_n = T) = 1/2$ together with the classical axioms of probability, and there seems to be nothing more to be said about how these coin tosses are combined than that they are causally²³ independent.²⁴ Genuine causal independence seems to be the whole story about the connection of the events.

To make the point vivid, imagine a multiverse consisting of a countable infinity of causally isolated island universes (except perhaps for an initial universe-generation process or creator, if things cannot come into existence *ex nihilo*), numbered $\dots, -2, -1, 0, 1, 2, \dots$, in each of which one of the coin tosses occurs, and nothing else relevant happens. The chances in such a multiverse should supervene on the fact of causal isolation as well as the chances in each universe of the multiverse. It seems absurd to suppose there could be two physically exact duplicates of such a multiverse, with the same chances within each island universe, but with different global infinitesimal probabilities.

4.3. Priors. Standard subjective Bayesianism holds that our credences should be driven by the evidence and the priors, but there are no rational constraints on priors other than regularity, the requirement that no contingent proposition get a prior of 0 or 1. As a result, while the dynamics of our credences in response to evidence may be rationally determined by the evidence, the actual values of our credences are rationally underdetermined as they depend on the priors, which are arbitrary. If we accept this, why can't infinitesimal probabilities involve a similar arbitrariness?

But, first, even if we accept subjective Bayesianism, that will only help with the arbitrariness of credences and not with the arbitrariness of objective chances. Second, there is good reason not to make Bayesianism be *entirely* subjective, but to place constraints on the credences. One plausible constraint would be some version of David Lewis's Principal Principle, that (roughly speaking) credences should match chances when the chances are known. But if the arbitrariness concern forces us to abandon infinitesimal values for chances, then the Principal Principle will force us to likewise abandon them for those credences that need to match these chances.

5. CONCLUSIONS

There might turn out to be some defensible specific choice for an extension of the real numbers that includes infinitesimals for measuring probabilities. But even if there is, the choice of which particular infinitesimals in this extension to assign to which probabilistic situations is highly underdetermined. We can see the underdetermination in two ways. First, we have considerations connected to particular infinitary processes, such as repeated coin tosses, where the infinitesimal needs to appropriately co-vary with the

²³And not merely statistically according to the classic definition (for limitations of the latter, see Fitelson and Hájek 2017).

²⁴I am grateful to Alan Hájek for a sketch of this response.

event under consideration, but there is no reason to attach a particular infinitesimal to a particular event. Second, we have seen that any assignment of infinitesimal probabilities can be replaced by a different assignment that matches the same real-valued conditional and unconditional probabilities and that satisfies all the same plausible decision-theoretic constraints.

Thus using a single hyperreal-valued unconditional probability function to define conditional probabilities, as opposed to either simply taking the conditional probabilities to be primitive as in the Popper function approach or considering a family of hyperreal-valued functions satisfying some constraints, introduces degrees of freedom in the choice that do not seem to match up with anything in reality, and there is good reason to balk at such an introduction.²⁵

APPENDIX

Let us start by reviewing some standard facts about extended reals and their standard and infinitesimal parts.

Suppose that \mathbb{R}^* is an extension of the reals. Say that elements of \mathbb{R}^* are infinitesimally close provided that their difference is zero or infinitesimal. Note that the only way two *real* numbers can be infinitesimally close is if they are equal. For the difference between two reals is real, and hence can't be infinitesimal.

Now suppose $x \in \mathbb{R}^*$ is finite, i.e., $-y < x < y$ for some real y . There can be at most one real number infinitesimally close to x . For suppose there were two. Then they would also be infinitesimally close to each other (the sum of two infinitesimals is infinitesimal). But as we just saw, that's impossible for *two* reals.

But there is at least one real number infinitesimally close to x . For let $S = \{z \in \mathbb{R} : z < x\}$. Then S is a non-empty set, as $-y \in S$ (we assumed $-y < x < y$ for a real y), and it's bounded above because $z < y$ for every $z \in S$. Every non-empty set of reals that is bounded above has a supremum, i.e., a least upper bound (this is a defining property of the reals). Let x_1 be the supremum of S . I now claim that $x_2 = x - x_1$ is infinitesimal or zero. If $x_1 = x$, this is trivial. So suppose $x_1 \neq x$. There are two possibilities. Either $x_1 < x$ or $x < x_1$. Suppose first that $x_1 < x$. If $x - x_1$ is not infinitesimal, there must be a real number $a > 0$ such that $a \leq x - x_1$. Then $x_1 + a \leq x$, and so $x_1 < x_1 + a/2 < x$. Hence, $x_1 + a/2 \in S$, and so x_1 is not an upper bound for S , contrary to the assumption that x_1 is the supremum of S . Hence, $x - x_1 = x_2$ has to be infinitesimal. The other possibility is that $x < x_1$. If $x_1 - x$ is not infinitesimal, there must be a real number $a > 0$ such that $a \leq x_1 - x$. Then $x \leq x_1 - a$, so $x < x_1 - a/2 < x_1$. But then $x_1 - a/2$ is an upper bound for S and yet is less than x_1 , so x_1 is not

²⁵The author is grateful to Alan Hájek as well as to two anonymous readers for a number of comments that have greatly improved this paper.

the least upper bound for S , contrary to x_1 's being the supremum of S . So, $x_1 - x$ must be infinitesimal, and hence $x_2 = -(x_1 - x)$ is infinitesimal.

Thus, x_1 is infinitesimally close to x , and so for every finite extended real x there is a unique real, which we will denote by $\text{Std } x$, infinitesimally close to it. This is called the *standard part* of x . Let $\text{Inf } x = x - \text{Std } x$ be the *infinitesimal part* of x , so $x = \text{Std } x + \text{Inf } x$.

Let $\phi_\alpha(x) = \text{Std } x + \alpha \text{Inf } x$. We can now prove the lemmas from Section 3.6, which will be useful for proving Theorem 1. First, Lemma 1 follows immediately from:

Lemma 3. *Std and Inf are \mathbb{R} -linear functions.*

Proof of Lemma 3. If x_1 and x are infinitesimally close to each other and if y_1 and y are infinitesimally close to each other, then $ax_1 + by_1$ is infinitesimally close to $ax + by$, if a and b are real, since $ax + by - (ax_1 + by_1) = a \cdot (x - x_1) + b \cdot (y - y_1)$ and if $x - x_1$ and $y - y_1$ are infinitesimal, their product with a real number is still infinitesimal, and the sum of two infinitesimals is infinitesimal. Hence, if $x_1 = \text{Std } x$ and $y_1 = \text{Std } y$, it follows that $\text{Std}(ax + by) = a \text{Std } x + b \text{Std } y$. Hence, Std is an \mathbb{R} -linear function. But $\text{Inf } x = x - \text{Std } x$, so if Std is \mathbb{R} -linear, so is Inf . \square

Second, we prove the strictly increasing character of ϕ_α .

Proof of Lemma 2. Suppose $x < y$ are in \mathbb{R}^* . By linearity $\phi_\alpha(y) - \phi_\alpha(x) = \phi_\alpha(y - x)$. Letting $z = y - x$, all we need to show is that if $z > 0$ is finite, then $\phi_\alpha(z) > \phi_\alpha(0) = 0$. To show this, suppose first that z is infinitesimal. Then $\text{Std } z = 0$ and $\text{Inf } z = z$, so $\phi_\alpha(z) = \alpha z > 0$. Suppose now that z is not infinitesimal. Then $\text{Std } z$ is a positive real, but $\alpha \text{Inf } z$ is infinitesimal. Adding an infinitesimal to a positive real has to result in something positive, so $\phi_\alpha(z) = \text{Std } z + \alpha \text{Inf } z > 0$. \square

Note also that ϕ_α fixes the reals, i.e., $\phi_\alpha(x) = x$ for any real x , as then $\text{Inf } x = 0$. Finally, observe that $\phi_{1/\alpha}(\phi_\alpha(x)) = x$, so ϕ is a bijection.

Lemma 4. *Suppose P is an \mathbb{R}^* -valued probability function, and $\phi : \mathbb{R}^* \rightarrow \mathbb{R}^*$ is an \mathbb{R} -linear and strictly increasing bijection fixing the reals. Let $P_\phi(A) = \phi(P(A))$. Then:*

- (i) P_ϕ is an \mathbb{R}^* -valued probability function,
- (ii) $P(A) < P(B)$ if and only if $P_\phi(A) < P_\phi(B)$,
- (iii) if $P(C) > 0$, then $P(A | C) < P(B | C)$ if and only if $P_\phi(A | C) < P_\phi(B | C)$
- (iv) if $a_1, \dots, a_n, b_1, \dots, b_m$ are real, then for any events $A_1, \dots, A_n, B_1, \dots, B_m$ we have $\sum_i a_i P(A_i) < \sum_i b_i P(B_i)$ if and only if $\sum_i a_i P_\phi(A_i) < \sum_i b_i P_\phi(B_i)$
- (v) if $P(C) > 0$, and $a_1, \dots, a_n, b_1, \dots, b_m$ are real, then for any events $A_1, \dots, A_n, B_1, \dots, B_m$ we have $\sum_i a_i P(A_i | C) < \sum_i b_i P(B_i | C)$ if and only if $\sum_i a_i P_\phi(A_i | C) < \sum_i b_i P_\phi(B_i | C)$.

Proof. We first verify the axioms of probability for P_ϕ . Consider any P -measurable event A . If $P(A) = 0$, then $P_\phi(A) = \phi(P(A)) = 0$ as ϕ preserves reals. If $P(A) > 0$, then $\phi(P(A)) > 0$ by increasingness and real-preservation. Hence, we have $P_\phi(A) \geq 0$ since $P(A) \geq 0$ for all A .

Next, since 1 is real, $P_\phi(\Omega) = \phi(P(\Omega)) = \phi(1) = 1$.

Finally, if A and B are disjoint, then by linearity and the finite additivity of P , we have $P_\phi(A \cup B) = \phi(P(A \cup B)) = \phi(P(A) + P(B)) = \phi(P(A)) + \phi(P(B)) = P_\phi(A) + P_\phi(B)$.

Thus P_ϕ is a finitely additive probability, as condition (i) requires. Conditions (ii)–(iv) are special cases of condition (v) (though (ii) also immediately follows from the strict increasingness of ϕ).

So let's prove (v). Suppose that $a_1, \dots, a_n, b_1, \dots, b_m$ are real and that events $A_1, \dots, A_n, B_1, \dots, B_m$ are such that

$$(1) \quad \sum_i a_i P(A_i | C) < \sum_i b_i P(B_i | C).$$

We will now show that

$$(2) \quad \sum_i a_i P_\phi(A_i | C) < \sum_i b_i P_\phi(B_i | C).$$

We don't need to separately show the other direction of the biconditional in (iv), because we can write $P = (P_\phi)_{\phi^{-1}}$ and apply the first direction with P_ϕ in place of P and ϕ^{-1} in place of ϕ , as ϕ is a bijection and its inverse satisfies the conditions of the lemma as well.

Let $D_i = A_i \cap C$ and $E_i = B_i \cap C$. Then if we multiply both sides of (1) by $P(C)$ and both sides of (2) by $P_\phi(C)$, and apply the ratio definition of conditional probability, we see that the two inequalities are respectively equivalent to:

$$(3) \quad \sum_i a_i P(D_i) < \sum_i b_i P(E_i).$$

and

$$(4) \quad \sum_i a_i P_\phi(D_i) < \sum_i b_i P_\phi(E_i).$$

But since $P_\phi(U) = \phi(P_\phi(U))$ and by the \mathbb{R} -linearity of ϕ , inequality (4) is equivalent to:

$$(5) \quad \phi \left(\sum_i a_i P(D_i) \right) < \phi \left(\sum_i b_i P_\phi(E_i) \right).$$

And by the strict increasingness of ϕ , this is indeed equivalent to (3), thereby completing the proof of (v). \square

Proof of Theorem 1. By our lemmas, all that remains is to prove condition (vi). Suppose first that $P(C) \approx 0$. Then $P_\alpha(C) = \alpha P(C)$ by definition

of P_α . Moreover, $P(A \cap C)$ will also be infinitesimally close to zero, so $P_\alpha(A \cap C) = \alpha P(A \cap C)$. Thus

$$P_\alpha(A | C) = \frac{P_\alpha(A \cap C)}{P_\alpha(C)} = \frac{P(A \cap C)}{P(C)} = P(A | C),$$

as desired.

Now suppose that $P(C) \not\approx 0$. Let $c_1 = \text{Std } P(C)$. This is a non-zero real number. Observe that if x is finite and $a \approx b$ with $a \not\approx 0$, then $x/a \approx x/b$. (For $\frac{x}{a} - \frac{x}{b} = \frac{(b-a)x}{ab}$, which is infinitesimally close to zero since $b - a \approx 0$ while $ab \not\approx 0$.) Now, $P(C) \approx P_\alpha(C) \approx c_1$. Thus,

$$P_\alpha(A | C) = \frac{P_\alpha(A \cap C)}{P(C)} \approx \frac{P_\alpha(A \cap C)}{c_1}.$$

But if $y > 0$ is real and $a \approx b$, then $a/y \approx b/y$, so using the fact that $P(A \cap C) \approx P_\alpha(A \cap C)$:

$$\frac{P_\alpha(A \cap C)}{c_1} \approx \frac{P(A \cap C)}{c_1}.$$

Hence

$$(6) \quad P_\alpha(A | C) \approx \frac{P(A \cap C)}{c_1}.$$

The right hand side of (6) does not depend on α , so all the values of the left hand side as we vary $\alpha > 0$ must be infinitesimally close. In particular, $P_1(A | C) \approx P_\alpha(A | C)$, and since $P_1 = P$, we have (vi) as desired. \square

Proof of Proposition 1. As a convenient and usual abuse of notation, we will consider “ z ” to be a constant denoting z in our preferred interpretation in the model \mathbb{R}^* , and other similar abuses will be perpetrated. Start by noting that z cannot be defined over \mathbb{R} in the first-order language of fields, i.e., there is no formula $F(a_1, \dots, a_n, x)$ in the first order language of fields with a_1, \dots, a_n real-valued constants such that $F(a_1, \dots, a_n, z)$ holds in \mathbb{R}^* and $\forall x(F(a_1, \dots, a_n, x) \rightarrow x = z)$. For if there were such a formula, then $\exists! x F(a_1, \dots, a_n, x)$ would hold in the model \mathbb{R}^* . But by the transfer principle for hyperreals (e.g., Keisler 2007, p. 9), likewise $\exists! x F(a_1, \dots, a_n, x)$ would hold in the model \mathbb{R} . Thus, there would be a real number u such that $F(a_1, \dots, a_n, u)$, and this sentence would also hold in \mathbb{R}^* by the transfer principle, which would imply that $u = z$, contradicting the fact that z is not real.

Now let Σ_0 be the set of all formulas $F(x)$ in the first-order language of fields with constants from \mathbb{R} and at most one free variable x such that $F(z)$ is true in \mathbb{R}^* and let Σ_1 be Σ_0 with the addition of the formula $z \neq x$. Observe that $\|\Sigma\| = \|\mathbb{R}\|$. Moreover, any finite subset of Σ is satisfiable over \mathbb{R}^* , i.e., if $F_0(x), F_1(x), \dots, F_n(x)$ are in Σ , then there is a u in \mathbb{R}^* that satisfies all the formulas. To see this, without loss of generality, assume that $F_0(x)$ is $x \neq z$ and that $F_1(x), \dots, F_n(x)$ are in Σ_0 . Let $F(x)$ be $F_1(x) \wedge \dots \wedge F_n(x)$. Then z satisfies $F(x)$ in \mathbb{R}^* , and so by the observations of the previous

paragraph it is false that $\exists!xF(x)$ holds in \mathbb{R}^* . Thus in \mathbb{R}^* there must be another element besides z that satisfies $F(x)$, say z' . This other element then satisfies $F_0(x)$, and so it satisfies all of $F_0(x), F_1(x), \dots, F_n(x)$. Hence every finite subset of Σ is satisfiable over \mathbb{R}^* .

Since \mathbb{R}^* is κ -saturated for $\kappa > \|\mathbb{R}\| = \|\Sigma\|$, it follows that Σ is satisfiable over \mathbb{R}^* . Suppose it is satisfied by z' . Then $z' \neq z$, since $x \neq z$ is one of the formulas in Σ . Moreover, $z \neq u$ for every real u , so the formula $z \neq u$ is in Σ_0 , so $z' \neq u$ for every real u , and hence $z' \notin \mathbb{R}$. Now define $\phi(x) = x$ for $x \in \mathbb{R}$ and $\phi(z) = z'$. Because z' satisfies every formula with one free variable that z satisfies, it follows that ϕ is a partial elementary embedding. But a partial elementary embedding in a κ -saturated model of cardinality κ extends to an automorphism of the model (Sacks, 2010, pp. 72–73), which completes the proof. \square

Proof of Theorem 2. By Proposition 1 and Lemma 4, we just need to prove the analogue of (vi), namely that if $P(C) > 0$, then $P(A | C) \approx P_\phi(A | C)$. To see this, observe that because ϕ is a field isomorphism:

$$\begin{aligned} P_\phi(A | C) &= \frac{P_\phi(A \& C)}{P_\phi(C)} \\ &= \frac{\phi(P(A \& C))}{\phi(P(C))} \\ &= \phi\left(\frac{P(A \& C)}{P(C)}\right) = \phi(P(A | C)). \end{aligned}$$

But $\phi(P(A | C)) \approx P(A | C)$ which follows from the fact that $\phi(x) \approx x$ for all finite $x \in \mathbb{R}^*$ (and $P(A | C)$ is finite as it is between 0 and 1). For suppose x is finite so we can write $x = a + b$ for a real a and an infinitesimal b . Since ϕ fixes the reals, $\phi(x) = \phi(a) + \phi(b) = a + \phi(b)$. If we can show that $\phi(b)$ is infinitesimal, it will follow that $\phi(x) \approx a \approx x$. To see that $\phi(b)$ is infinitesimal, fix any real $c > 0$. Then $-c < b < c$. Hence $\phi(-c) < \phi(b) < \phi(c)$ as ϕ is a field isomorphism of \mathbb{R}^* to itself and hence preserves order (for $\alpha \leq \beta$ in a real-line if and only if there is a γ such that $\beta = \alpha + \gamma^2$, and this algebraic characterization of \leq is preserved by field isomorphism). But $\phi(-c) = -c$ and $\phi(c) = c$ as ϕ fixes the reals. Thus $-c < \phi(b) < c$, and since this holds for all real $c > 0$, $\phi(b)$ is infinitesimal. \square

REFERENCES

- [1] Benacerraf, Paul. 1962. “Tasks, super-tasks, and the modern Eleatics.” *Journal of Philosophy* 59:765–784.
- [2] Benacerraf, Paul. 1965. “What numbers could not be”. *Philosophical Review* 74:47–73.
- [3] Benci, Vieri; Horsten, Leon; and Wenmackers, Sylvia. 2013. “Non-archimedean probability”. *Milan Journal of Mathematics* 81:121–151.
- [4] Benci, Vieri; Horsten, Leon; and Wenmackers, Sylvia. 2018. “Infinitesimal probabilities”. *British Journal for the Philosophy of Science* 69:509–552.

- [5] Brickhill, Hazel and Horsten, Leon. 2016. “Popper Functions, lexicographical probability, and non-Archimedean probability”. <https://arxiv.org/abs/1608.02850>.
- [6] DiBella, Nicholas. 2018. “The qualitative paradox of non-conglomerability”. *Synthese* 195:1181–1210.
- [7] Easwaran, Kenny. 2014. “Regularity and hyperreal credences”. *Philosophical Review* 123:1–41.
- [8] Ehrlich, Philip. 2012. “The absolute arithmetic continuum and the unification of all numbers great and small”. *Bulletin of Symbolic Logic* 18:1–45.
- [9] Fitelson, Branden and Hájek, Alan. “Declarations of independence”. *Synthese* 194:3979–3995.
- [10] Hájek, A. 2003. “What conditional probability could not be”. *Synthese* 137:273–323.
- [11] Hofweber, Thomas. 2014. “Infinitesimal chances”. *Philosophers’ Imprint* 14:1–14.
- [12] Jech, Thomas. 1997. *Set Theory*. 2nd edition. Berlin: Springer.
- [13] Kanovei, V. and Shelah, S. 2004. “A definable nonstandard model of the reals”. *Journal of Symbolic Logic* 69:159–164.
- [14] Keisler, H. J. 2007. *Foundations of Infinitesimal Calculus*, Online Edition. 2007. <https://www.math.wisc.edu/~keisler/foundations.html>
- [15] Keisler, H. J. 1994. “The hyperreal line”. In: P. Erlich (ed.), *Real Numbers, Generalizations of Reals, and Theories of Continua*. Kluwer. Pp. 207–237.
- [16] Krauss, P. H. 1968. “Representation of conditional probability measures on Boolean algebras”. *Acta Mathematica Academiae Scientiarum Hungarica* 19:229–241.
- [17] Laugwitz, Detlef. 1968. “Eine nichtarchimedische Erweiterung angeordneter Körper.” *Mathematische Nachrichten* 37:225–236.
- [18] McGee, Vann. 1994. “Learning the impossible.” In: E. Eells and B. Skyrms (eds.), *Probability and Conditionals: Belief Revision and Rational Decision*. Cambridge: Cambridge University Press.
- [19] Norton, John D. 2018. “How to build an infinite lottery machine.” *European Journal for the Philosophy of Science* 8:71–95.
- [20] Norton, John D. and Pruss, Alexander R. 2018. “Correction to John D. Norton ‘How to build an infinite lottery machine’.” *European Journal for the Philosophy of Science* 8:143–144.
- [21] Pedersen, A. Paul. MS. “Strictly coherent preferences, no holds barred.”
- [22] Pincus, David and Solovay, Robert M. 1977. “Definability of measures and ultrafilters”. *Journal of Symbolic Logic* 42:179–190.
- [23] Pruss, Alexander R. 2013a. “Probability, regularity, and cardinality.” *Philosophy of Science* 80:231–240.
- [24] Pruss, Alexander R. 2013b. “Null probability, dominance and rotation”. *Analysis* 73:682–685.
- [25] Pruss, Alexander R. 2014. “Infinitesimals are too small for countably infinite fair lotteries”. *Synthese* 191:1051–1057.
- [26] Pruss, Alexander R. 2015. “Popper functions, uniform distributions and infinite sequences of heads”. *Journal of Philosophical Logic* 44:259–271.
- [27] Pruss, Alexander R. 2018. *Infinity, Causation and Paradox*. Oxford: Oxford University Press.
- [28] Robert, Alain M. 2000. *A Course in p-adic Analysis*. New York: Springer.
- [29] Sacks, Gerald E. 2010. *Saturated Model Theory*. 2nd ed. Singapore: World Scientific Publishing.
- [30] Shamseddine, Khodr and Berz, Martin. 2010. “Analysis on the Levi-Civita field, a brief overview”. *Contemporary Mathematics* 508:215–237.
- [31] Thomson, James F. Thomson. “Tasks and super-tasks.” *Analysis* 15:1–13.
- [32] Weintraub, Ruth. 2008. “How probable is an infinite sequence of heads? A reply to Williamson” *Analysis* 68:247–250.

- [33] Williamson, Timothy. 2007. "How probable is an infinite sequence of heads?" *Analysis* 67:173–80.