Probability, Regularity, and Cardinality

Alexander R. Pruss*†

Regularity is the thesis that all contingent propositions should be assigned probabilities strictly between zero and one. I will prove on cardinality grounds that if the domain is large enough, a regular probability assignment is impossible, even if we expand the range of values that probabilities can take, including, for instance, hyperreal values, and significantly weaken the axioms of probability.

1. Introduction. Regularity is usually taken to be the thesis that all contingent propositions should be given probabilities strictly between zero and one (advocates of regularity conditions include Kemeny 1955; Shimony 1955; Jeffreys 1961; Stalnaker 1970; Lewis 1980; Appiah 1985). There are some well-known problems for regularity. For instance, if you throw a dart with a perfectly defined tip at a dartboard, on the standard probability calculus, the probability that it hits a particular point is going to be zero, even though it is possible for the dart to hit that point.[1] Another standard example is that a fair coin that will be flipped an infinite number of times has a zero probability of landing heads each time but nonetheless can do so (after all, the all-heads sequence is just as possible as any other infinite sequence of throw outcomes).

However, such cases may simply be taken to show that we need to expand the range of values that probability assignments take beyond real numbers, to include, for instance, hyperreal values such as infinitesimals in a coherent

Received January 2012; revised July 2012.

*To contact the author, please write to: Department of Philosophy, Baylor University, One Bear Place no. 97273, Waco, TX 76798; e-mail: alexander_pruss@baylor.edu.

†I would like to thank Trent Dougherty, Alan Hájek, Jonathan Kvanvig, and A. Paul Pedersen for discussions, comments, and encouragement and the anonymous readers for feedback that led to significant improvement of this article.

1. The tip could be perfectly sharp, or else it could have perfect symmetries that define its center.

Philosophy of Science, 80 (April 2013) pp. 231–240. 0031-8248/2013/8002-0004$10.00
Copyright 2013 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/670299 Published online by Cambridge University Press

nonstandard analysis setting. Thus, one might assign an infinitesimal probability to any particular outcome of the dart throw experiment (cf. Bernstein and Wattenberg 1969; Skyrms 1980, 187) or to any particular sequence of coin throws. Recently, Williamson (2007) has argued that in the coin-toss case this cannot be done, but his argument relies on a number of assumptions from classical probability theory, such as how independence is to be handled. The regularist can still hope that some extension of the classical probability calculus will save regularity or can deny one of Williamson's intuitions about the probabilities to be assigned in the cases he constructs (e.g., Weintraub 2008).

Alan Hájek (2011) suggested, however, that once one extends the set of probability values, say to the hyperreals, one should be able to extend the probability domains—we may think, say, of darts thrown in a space-time with a hyperreal-based continuum—to make the problem of zero-probability events recur. He conjectured that the probability-value-extension "arms race" is a contest the defender of regularity cannot win: no matter how far she extends the set of probability values, there will be a domain where regularity fails even given that set of probability values. I shall prove Hájek's conjecture under very mild assumptions about probability values, via a counting argument. In the case in which the probability values are partially ordered with an appropriate addition operator and we have a generalization of the Kolmogorov axioms with finite additivity, a no-go result holds that if the sample space has sufficiently large cardinality as compared to the set of probability values, there is no regular probability assignment. The formal results will be presented in section 2, and then in section 3 I will argue on philosophical grounds that the kinds of high-cardinality sample spaces that the formal results discuss are in fact possible. An appendix with some proofs follows. While the results become somewhat neater given the axiom of choice (AC), our main results require only the standard Zermelo-Fraenkel axioms.[2]

2. The Formal Results. Formal Kolmogorovian probability theory involves a sample space Ω, a σ-algebra F on Ω, and a probability measure P on F. A σ-algebra on Ω is a nonempty set of subsets of Ω closed under complementation and countable unions. The members of F are known as the *measurable* subsets of Ω. The probability measure, then, is a function from F to the reals satisfying the following conditions:

(**K1**) $0 \leq P(A)$ for all A in F.

2. It has also come to my attention that a version of the main result, albeit using AC, was independently discovered by Arthur Paul Pedersen around the same time as I discovered it (early 2012).

https://doi.org/10.1086/670299 Published online by Cambridge University Press

(**K2**) $P(\Omega) = 1$.
(**K3**) $P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$ for any pairwise disjoint sets $A_1, A_2, \ldots$ in F (where a collection of sets is pairwise disjoint, provided that no two have any elements in common).

We then say that P is regular, provided that $P(A) > 0$ for every nonempty member A of F (it then follows from the axioms that $P(A) < 1$ for every member A of F other than Ω).

For our formal no-go results, we shall weaken these assumptions. First, we shall loosen the assumption that probabilities are real numbers. Instead, we shall be working with a set V of values with a partial ordering $\leq$, that is, a transitive binary relation such that for all x and y in V if $x \leq y$ and $y \leq x$ then $x = y$, and for all x in V we have $x \leq x$. I shall write "$x < y$" for "$x \leq y$ and $x \neq y$." We shall also assume that V has an addition operation "+" and a distinguished element 0, satisfying the following two axioms:

(**V1**) $x + 0 = x$ for all x in V.
(**V2**) For all x and y in V, if $0 < y$ then $x < x + y$.

Moreover, we shall weaken the countable additivity axiom K3 to finite additivity. If we do that, it will also be natural to weaken the assumption that F is a σ-algebra to the assumption that it is an algebra of subsets of Ω, that is, a nonempty set of subsets of Ω closed under complementation and finite unions. Thus, suppose that we have a V as above. A V-probability assignment will then be a function P from the algebra F to V satisfying:

(**K1***) $0 \leq P(A)$ for all A in F.
(**K3***) $P(A \cup B) = P(A) + P(B)$ for any disjoint sets A and B in F.

The total probability axiom K2 has been dropped entirely because our formal results do not require it. If we really wanted to, we could replace K2 with the nontriviality assumption:

(**K2***) $0 < P(\Omega)$.

But we will not need to. We will, as before, say that P is regular, provided that $0 < P(A)$ for every nonempty A in F. The range R of the V-probability assignment shall be the set of values that P can take: $R = \{P(A) : A \in F\}$.

We will write that $|A| \leq |B|$ and say that the cardinality of A is less than or equal to that of B, provided that there is a one-to-one function from A to B, and write that $|A| = |B|$ and say that A and B have the same cardinality, provided that there is a one-to-one function from A onto all of B. Then $|A| < |B|$ means that $|A| \leq |B|$ and $|A| \neq |B|$. For convenience, I will use $|A|_+$

https://doi.org/10.1086/670299 Published online by Cambridge University Press

as an abbreviation for $|A \cup \{a_0\}|$, where a_0 is not in A; thus, $|A|_+$ is the same cardinality as $|A|$, if A is infinite, and one plus the cardinality of A, if A is finite. I will also write $|A| \not\leq |B|$ for the denial of $|A| \leq |B|$ ($|A| \not\leq |B|$ is equivalent to $|B| < |A|$ if AC holds). At this point, no interpretation is needed for the symbol $|A|$ outside of a context like that above.

Our main formal result follows and is proved in the appendix. A total ordering $<$ on a set A is a transitive binary relation on A such that for all x and y in A exactly one of the following relations holds: $x < y$, $y < x$, or $x = y$.

> **Theorem 1.** Suppose that the sample space Ω has a total ordering $<$ and that for every $x \in \Omega$, both the singleton $\{x\}$ and the set $L_x = \{y \in \Omega : y < x\}$ are members of the algebra F on Ω. Suppose that V satisfies V1 and V2 and that P is a V-probability assignment on F. Let R be the range of P. Suppose that $|\Omega|_+ \not\leq |R|$. Then there is a member x of Ω such that $P(\{x\}) = 0$.

To see how the theorem works, consider as a special case classical Kolmogorovian probabilities on a finite space $\Omega = \{\omega_1, \ldots, \omega_n\}$ with n elements. Let $\Omega_0 = \varnothing$ and let $\Omega_i = \{\omega_1, \ldots, \omega_i\}$ for $0 < i \leq n$. Suppose P is a regular probability assignment on the power set of Ω. Then by K3, we have $P(\Omega_i) = P(\Omega_{i-1}) + P(\{\omega_i\}) > P(\Omega_{i-1})$, and so $P(\Omega_0) < P(\Omega_1) < \ldots < P(\Omega_n)$. Thus, the range of P must contain at least $n + 1$ values. Contrapositively, if $n + 1$ is not less than or equal to the cardinality of the range of P, the assignment P will not be regular, and hence $P(\{x\}) = 0$ for some $x \in \Omega$, which is what the theorem says.

Theorem 1 will follow from theorem 2, below. But first we give two corollaries.

> **Corollary 1.** Assume the axiom of choice. Suppose V satisfies V1 and V2. Then there is no regular V-probability assignment on all the subsets of Ω if Ω has cardinality greater than V.

For by AC, Ω has a well ordering (a well ordering on A is a total ordering on A such that every nonempty subset of A has a minimal element) and hence a total ordering.[3] If $|\Omega| > |V|$, then a fortiori $|\Omega|_+ > |R|$, where R is the range of P, and hence $|\Omega|_+ \not\leq |R|$, where the move from $|A| > |B|$ to $|A| \not\leq |B|$ uses the Cantor-Schröder-Bernstein theorem (Takeuti and Zaring 1971, 71). But then theorem 1 shows that there is no regular V-probability assignment on all the subsets of Ω.

3. AC is equivalent to the claim that every set can be well ordered. The assumption that every set has a total ordering has been proved to be weaker (Gonzáles 1995).

https://doi.org/10.1086/670299 Published online by Cambridge University Press

The above corollary made use of AC. Here is one that does not use AC but merely the standard Zermelo-Fraenkel axioms. Given a set R, let R^+ be the Hartogs number of R, that is, the least ordinal α such that there is no one-to-one function from α to R. An ordinal is a set α with the properties that (a) if $x \in \alpha$, then every member of x is also a member of α, and (b) α is well ordered by the membership relation $\in$. (The finite ordinals are the natural numbers constructed inductively by setting $0 = \varnothing$ and $n + 1 = n \cup \{n\}$.) It is then a theorem of Zermelo-Fraenkel set theory, without any need for AC, that a Hartogs number exists for every set (Hrbacek and Jech 1999, 130).

Corollary 2. Suppose V satisfies V1 and V2. Then there is no regular V-probability assignment P on all the subsets of Ω if $|\Omega| \geq |R^+|$, where R is the range of P.

This corollary could be proved from theorem 1, but we will instead get it as a direct consequence of corollary 4, below.

One may want to generalize the notion of probability and regular probability even further, allowing for a set V of values that has neither a distinguished zero nor an addition operation. Suppose now that V is any partially ordered set. If Ω is a sample space and F an algebra of subsets of Ω, then we say that P is a generalized V-probability, provided that:

(**K3****) If $A \subseteq B$ for sets A and B in F, then $P(A) \leq P(B)$.

Moreover, we say that P is a regular generalized V-probability, provided that whenever A and B are measurable sets such that A is a proper subset of B, then $P(A) < P(B)$. We then have the following analogues of our formal results above.

Theorem 2. Suppose that the sample space Ω has a total ordering $<$ and that for every $x \in \Omega$, the set $L_x = \{y \in \Omega : y < x\}$ is a member of the algebra F on Ω. Suppose that V is a partially ordered set and that P is a generalized V-probability assignment on F. Let R be the range of P. Suppose that $|\Omega|_+ \not\leq |R|$. Then P is not a regular generalized V-probability assignment.

Corollary 3. Assume the axiom of choice. Suppose V is a partially ordered set. Then there is no regular generalized V-probability assignment on all the subsets of Ω if Ω has cardinality greater than V.

Corollary 4. Suppose V is a partially ordered set. Then there is no regular generalized V-probability assignment P on all the subsets of Ω if $|\Omega| \geq |R^+|$, where R is the range of P.

https://doi.org/10.1086/670299 Published online by Cambridge University Press

The following lemma implies that, given V1 and V2, *V*-probabilities are generalized *V*-probabilities and regular *V*-probabilities are regular generalized *V*-probabilities.

> **Lemma 1.** Suppose V satisfies V1 and V2, and P is a V-probability assignment. Let A and B be measurable sets. If A is a subset of B, then $P(A) \leq P(B)$. If P is regular and A is a proper subset of B, then $P(A) < P(B)$.

Theorem 1 then follows from theorem 2. For under the conditions of theorem 1, by theorem 2, the function P is not a regular generalized V-probability assignment, and so by lemma 1, P is not a regular V-probability assignment, and hence $P(A) = 0$ for some nonempty A. But then let x be any member of A and observe that $0 \leq P(\{x\}) \leq P(A) = 0$ by K1* and lemma 1, from which it follows by partial ordering that $P(\{x\}) = 0$, which is the conclusion of theorem 1. Corollary 3 follows from theorem 2 just as corollary 1 followed from theorem 1. Finally, corollary 4, given lemma 1, immediately implies corollary 2. In the appendix, we will prove lemma 1, theorem 2, and corollary 4.

3. The Philosophical Results. Corollary 2 tells us that no matter how rich and nonstandard we make our probability values, no matter how far in a hierarchy of hyperreals we go, as long as our sample space is large enough, we will run into zero-probability possibilities. There is, however, a philosophical question: Are there any such high-cardinality sample spaces of possible situations? I will offer one family of constructions.

Plausibly, there are kinds K of concrete objects such that for any nonzero cardinality k, it is possible that there are exactly k objects that are Ks. For example, we might simply take K to be the kind *concrete object*. While it is controversial whether there could be no concrete objects at all,[4] it is very plausible that for any nonzero cardinality k there could be exactly k concrete objects. Or we might follow Pruss (2001) and take K to be a kind of elementary particle like *photon*, which allows for more than one instance of the kind to be in the same quantum state (you cannot run this argument on the basis of electrons). Moreover, while physicists appear to think of space-time as having at most the cardinality of the continuum, there does not seem to be a principled reason to limit space-time to be Archimedean in nature.[5]

4. Theists who think God is a necessary being deny such a possibility, as do some nontheists on other grounds (Rundle 2004).

5. The mathematician E. E. Rosinger, in a series of preprints (e.g., Rosinger 2007), has tried to encourage physicists to consider space-times with larger cardinalities, but as far as I can tell his appeal has had no effect.

https://doi.org/10.1086/670299 Published online by Cambridge University Press

One can easily mathematically construct large cardinality non-Archimedean space-times (just use the upward Löwenheim-Skolem theorem), and it is plausible that such a space-time could be populated with a large cardinality of entities. At least we should not rule out such possibilities a priori. Or we might take K to be a kind like *universe*, for example, defining a universe as a maximal spatiotemporally connected aggregate of concrete objects. *Pace* David Lewis (1986), it is very plausible indeed that for every nonzero cardinality k (again, the $k = 0$ case will be controversial) there could have been exactly k universes. Many versions of string theory predict a multiverse containing a large finite cardinality (and others predict an infinite cardinality) of universes, so hypotheses about the cardinality of the set of universes are within the realm of science.

Given such a kind K, we can construct a counterexample to regularity as follows. We imagine the regularist to have assigned to all propositions probabilities that take values in some large set V satisfying V1 and V2 and satisfying propositional analogues of K1* and K3*, namely, that $0 \leq P(p)$ for every proposition p and $P(p \vee q) = P(p) + P(q)$ whenever p and q are incompatible. Let V^+ be the Hartogs number of V. Let $\aleph_\alpha$ be the αth element in the aleph sequence (Takeuti and Zaring 1971, 81). Then the number of initial ordinals β such that $\aleph_0 < \beta \leq \aleph_{V^+}$ equals V^+. Now let Ω be the space of all such initial ordinals β (this will be well ordered under the ordinal ordering) and let F be any algebra of measurable sets that contains all singletons $\{\beta_0\}$ and all ordinal intervals of the form $\{\beta \in \Omega : \beta < \beta_0\}$ (we could just let F be the power set of Ω if we think every subset of Ω should be measurable). Then $|\Omega| = |V^+| \not\leq |V|$. Since the range R of P is a subset of V, it follows that $|\Omega|_+ \not\leq |R|$, and, by theorem 1, it follows that P is not regular.[6]

The singletons in F correspond to propositions of the form "there are exactly β_0 Ks," while the ordinal intervals correspond to propositions of the form "the cardinality of Ks is infinite but less than β_0," and it is very plausible that such propositions should have probabilities on a Bayesian program. After all, for certain kinds K, we might well have to engage in scientific reasoning about such propositions.

We can also extend such examples to generalized V-probability assignments. While for V-probability on propositions, regularity says that only impossible probabilities p have V-probability 0, for generalized V-probability on propositions, regularity would say that if p entails q and $P(p) = P(q)$, then q also entails p (i.e., p and q are logically equivalent). Extended examples would show that generalized V-probabilities do not save regularity. The basic lesson of this article is that no matter how large our range of probability

6. I am very grateful to a referee who pointed out that my original argument here, which used AC, can be replaced by a simpler argument using Hartogs numbers.

https://doi.org/10.1086/670299 Published online by Cambridge University Press

values is, we can choose a sample space of possible situations that is large enough that we will have to assign probability zero to some possibility or at least (this is the case of generalized *V*-probabilities) assign the same probability to a set as to a proper subset of it, even if we have significantly relaxed the axioms of probability.

Appendix

Proofs of Lemma 1, Theorem 2, and Corollary 4

Proof of Lemma 1. Suppose $A \subseteq B$. Let $C = \{x \in B : x \notin A\}$. Then C is measurable, given that A and B are measurable. (The measurable sets are closed under complements and finite unions, and $C = (A \cup B^c)^c$, where A^c is the complement of A.) Then $B = A \cup C$. Since A and C are distinct, by K3* we have $P(B) = P(A) + P(C)$. Now, by K1*, $0 \leq P(C)$. There are two possibilities: either $0 = P(C)$ or $0 < P(C)$. If $0 = P(C)$, then $P(B) = P(A) + 0 = P(A)$ by V1, and hence $P(A) \leq P(B)$. If $0 < P(C)$, then by V2 we have $P(A) < P(A) + P(C) = P(B)$, and so once again $P(A) \leq P(B)$. This completes the proof of the first half of the lemma. The proof of the second half observes that if A is a proper subset of B, then C is nonempty and by regularity $0 < P(C)$. But then $P(B) = P(A) + P(C) > P(A)$ by K3* and V2. QED

Proof of Theorem 2. Start by observing that $L_x \neq L_y$ if $x \neq y$. To see this, observe that if $x \neq y$, then by total ordering on Ω either $x < y$ or $y < x$. If $x < y$, then $x \in L_y$ but $x \notin L_x$ (since we never have $x < x$ in a totally ordered set), so $L_x \neq L_y$. In the same way, if $y < x$, then $y \in L_x$ but $y \notin L_y$, and so $L_x \neq L_y$.

Next observe that for every x in Ω, we have $L_x \neq \Omega$. This follows from the fact that $x \notin L_x$ but $x \in \Omega$. Let $\Omega^* = \{\Omega\} \cup \{L_x : x \in \Omega\}$. By our observations, all the L_x are distinct from one another and from Ω. Thus, $|\{L_x : x \in \Omega\}| = |\Omega|$. Therefore, $|\Omega^*|$ equals $|\Omega|_+$; that is, it equals $|\Omega| + 1$ if Ω is finite and equals $|\Omega|$ if Ω is infinite (since then the cardinality of the union of $\{L_x : x \in \Omega\}$ with a singleton is the same as the cardinality of $\{L_x : x \in \Omega\}$).

Now assume that $|\Omega|_+ \not\leq |R|$, and for a reductio assume that P is a regular generalized *V*-probability assignment. Observe now that $P(A) \neq P(B)$ if A and B are distinct members of Ω^*. To see this, consider two cases. The first case is where $A = L_x$ and $B = L_y$ for some $x \neq y$. In that case, by total ordering, either $x < y$ or $y < x$. If $x < y$, then A is a proper subset of B. For if $z \in A$, then $z < x$ and so by transitivity $z < y$ and so $z \in B$, and hence A is a subset of B. But $x \in L_y$ and $x \notin L_x$, so A is a proper subset of B. By lemma 1

https://doi.org/10.1086/670299 Published online by Cambridge University Press

and regularity, it follows that $P(A) < P(B)$, and so $P(A) \neq P(B)$. By exactly the same reasoning, if $y < x$, then B is a proper subset of A, and so $P(B) < P(A)$, and hence again $P(A) \neq P(B)$.

The remaining case is where one of A and B is Ω, and the other is L_x for some x. Since L_x is a proper subset of Ω (x is a member of Ω but not of L_x), again by lemma 1 we can conclude that $P(L_x) \neq P(\Omega)$. So we have shown that P takes distinct values on distinct members of Ω^*. Consequently, P is a one-to-one function from Ω^* to R. Thus, $|\Omega^*| \leq |R|$. But $|\Omega^*| = |\Omega|_+$, and so we have a contradiction to our assumption that $|\Omega|_+ \not\leq |R|$. Absurdity has been reached, and so we must reject the assumption of regularity. QED

Proof of Corollary 4. Let $\Omega' = R^+$. For a reductio, suppose $|\Omega| \geq |\Omega'|$ and that P is a regular generalized V-probability assignment on the power set of Ω with range R. Since $|\Omega| \geq |\Omega'|$, let f be a one-to-one function from Ω' to Ω and define the V-probability function P' on the power set of Ω' by the formula $P'(A) = P(f[A])$, where $f[A] = \{f(x) : x \in A\}$.

Observe that since P is regular, so is P'. For suppose that A is a proper subset of B. Then $f[A]$ is a proper subset of $f[B]$ because f is one to one, and so by regularity $P(f[A]) < P(f[B])$, and thus $P'(A) < P'(B)$.

But now Ω' is well ordered (since it is an ordinal) and hence totally ordered. Moreover $|\Omega'| \not\leq |R|$ by definition of the Hartogs number, and so $|\Omega'|_+ \not\leq |R|$. Since R is a superset of the range of P', by theorem 2 we conclude that P' is not regular, which contradicts our regularity claim above. By reductio, the proof is complete. QED

REFERENCES

Appiah, Kwame Anthony. 1985. *Assertion and Conditionals*. Cambridge: Cambridge University Press.

Bernstein, Allen R., and Frank Wattenberg. 1969. "Non-standard Measure Theory." In *Applications of Model Theory of Algebra, Analysis, and Probability*, ed. W. A. J. Luxemberg, 171–85. New York: Holt, Rinehart & Winston.

Gonzáles, Carlos G. 1995. "Dense Orderings, Partitions and Weak Forms of Choice." *Fundamenta Mathematicae* 147:11–25.

Hájek, Alan. 2011. "Staying Regular?" Paper presented at the Australasian Association of Philosophy Conference, July.

Hrbacek, Karel, and Thomas Jech. 1999. *Introduction to Set Theory*. 3rd ed. New York: Dekker.

Jeffreys, Harold. 1961. *Theory of Probability*. 3rd ed. Oxford: Oxford University Press.

Kemeny, John G. 1955. "Fair Bets and Inductive Probabilities." *Journal of Symbolic Logic* 20: 263–73.

Lewis, David. 1980. "A Subjectivist's Guide to Objective Chance." In *Studies in Inductive Logic and Probability*, vol. 2, ed. R. C. Jeffrey, 263–93. Berkeley: University of California Press.

———. 1986. *On the Plurality of Worlds*. Oxford: Blackwell.

Pruss, Alexander R. 2001. "The Cardinality Objection to David Lewis's Modal Realism." *Philosophical Studies* 104:169–78.

Rosinger, Elemer E. 2007. "String Theory: A Mere Prelude to Non-Archimedean Space-Time Structures?" arXiv.org, Cornell University. arXiv:physics/0703154v1.

https://doi.org/10.1086/670299 Published online by Cambridge University Press

Rundle, Bede. 2004. *Why There Is Something Rather than Nothing*. Oxford: Oxford University Press.
Shimony, Abner. 1955. "Coherence and the Axioms of Confirmation." *Journal of Symbolic Logic* 20:1–28.
Skyrms, Brian. 1980. *Causal Necessity: A Pragmatic Investigation of the Necessity of Laws*. New Haven, CT: Yale University Press.
Stalnaker, Robert C. 1970. "Probability and Conditionals." *Philosophy of Science* 37:64–80.
Takeuti, Gaisi, and Wilson M. Zaring. 1971. *Introduction to Axiomatic Set Theory*. New York: Springer.
Weintraub, Ruth. 2008. "How Probable Is an Infinite Sequence of Heads? A Reply to Williamson." *Analysis* 68:247–50.
Williamson, Timothy. 2007. "How Probable Is an Infinite Sequence of Heads?" *Analysis* 67: 173–80.

https://doi.org/10.1086/670299 Published online by Cambridge University Press