



# Timothy Williamson's Coin-Flipping Argument: Refuted Prior to Publication?

Colin Howson<sup>1</sup>

Received: 30 May 2018 / Accepted: 29 March 2019  
© Springer Nature B.V. 2019

## Abstract

In a well-known paper, Timothy Williamson (Analysis 67:173–180, 2007) claimed to prove with a coin-flipping example that infinitesimal-valued probabilities cannot save the principle of Regularity, because on pain of inconsistency the event ‘all tosses land heads’ must be assigned probability 0, whether the probability function is hyperreal-valued or not. A premise of Williamson’s argument is that two infinitary events in that example must be assigned the same probability because they are isomorphic. It was argued by Howson (Eur J Philos Sci 7:97–100, 2017) that the claim of isomorphism fails, but a more radical objection to Williamson’s argument is that it had been, in effect, refuted long before it was published.

## 1 Introduction

Since its emergence in Abraham Robinson’s epoch-making work halfway through the twentieth century, nonstandard analysis has enriched many branches of mathematics and science: measure theory (including probability theory), analysis, physics and economics among them.<sup>1</sup> It is also a remarkable tribute to the power and conceptual fertility of modern mathematical logic: Robinson used the Compactness Theorem of first-order logic to prove the existence a suitable enlargement of the universe of sets to a so-called superstructure based on the hyperreal numbers (i.e. the members of a suitable elementary extension of the real number field containing infinitesimal and reciprocally infinite numbers), while Łos’s Theorem proves the

---

<sup>1</sup> A seminal paper by Loeb (1975) inaugurated a flourishing research program of using the tool of what are now known as Loeb spaces to investigate standard problems in applied probability.

---

I am grateful to two anonymous reviewers for their very helpful comments on this paper.

---

✉ Colin Howson  
howson.colin@gmail.com

<sup>1</sup> Professor Emeritus, University of Toronto, 16 Tranby Avenue, Toronto, ON M5R 1N5, Canada

Transfer Principle<sup>2</sup> for extensions using an ultrapower on a non-principal ultrafilter on  $\mathbb{N}$ . Concerns about the apparent lack of a single canonical extension have been allayed by Kanovei and Shelah's proof that a suitable non-standard model of the hyperreals is explicitly definable in ZFC (2004), a result extended by Herzberg to the definability of a nonstandard superstructure enlargement (2008), while Herzberg et al. have also shown that for Kanovei and Shelah's proof only Countable Choice for sets of real numbers need be assumed. Moreover, assuming the existence of an inaccessible cardinal (an assumption arguably underwriting the existence of a standard model of set theory) there is a superstructure enlargement of cardinality  $\kappa$ , where  $\kappa$  is the least inaccessible cardinal, which is  $\kappa$ -saturated, satisfies the Transfer Principle and is unique up to isomorphism (Keisler 2007, 197).<sup>3</sup> This is all model theory, but Edward Nelson (1977) showed that adding a new unary predicate *st* (interpreted as 'standard') and three new axioms added to those of Zermelo–Fraenkel set theory plus the Axiom of Choice (usually abbreviated to ZFC) gives a conservative extension of the latter.

Besides its fruitfulness in applied mathematics research, nonstandard models have also been enlisted in the search for solutions to problems of a more philosophical nature. One of these concerns the doctrine of Regularity, the view that only necessary truths and falsehoods merit the probabilities 1 and 0 respectively, at any rate where the probability in question is regarded as a measure of chance.<sup>4</sup> Yet standard probability functions are forced to assign 0 to some contingent events: for example, to all but a countable number of cells of an uncountable partition, and, if the distribution is uniform, to all the cells in a countably infinite partition. If probability functions are allowed to take values in the nonstandard unit interval, however, Regularity seems achievable by the strategy of assigning such events infinitesimal probabilities. Indeed, in an early paper (1996), Bernstein and Wattenberg showed that there is a finitely additive probability measure that assigns equal infinitesimal values to the points of the standard unit interval and differs by an infinitesimal from Lebesgue measure on the Lebesgue-measurable sets,<sup>5</sup> and more recently Wenmackers and Horsten (2013) have shown that it is possible to define a uniform infinitesimal probability distribution over the partition of  $\mathbb{N}$  into singletons, which sums hyperfinitely to 1.

Nevertheless, Timothy Williamson has claimed to show that appealing to hyperreal probabilities cannot save Regularity (2007), arguing that some events must be assigned probability 0 on pain of contradiction even if the probability function takes values in the nonstandard unit interval. In what follows I will present Williamson's argument and show that it fails, and what is more that its failure is implicit not only

<sup>2</sup> For the meaning of this and other key terms of nonstandard analysis, like internal vs. external, Transfer Principle, hyperfinite etc., which will crop up in this paper see Albeverio et al. (2009, Chapter 1).

<sup>3</sup> A nonstandard extension is  $\kappa$ -saturated if every family of fewer than  $\kappa$  internal sets with the finite intersection property has a nonempty intersection.

<sup>4</sup> Thus David Lewis: 'Zero chance is *no* chance, and nothing with zero chance ever happens' (1976, 176). Probability functions with this property are also called *strictly positive*.

<sup>5</sup> The actual infinitesimal is however rather arbitrary: it depends on the choice of a hyperfinite subset of the nonstandard unit interval which the authors call a 'sample'.

in Bernstein and Wattenberg's results but also in more recent work that still however long antedates Williamson's paper.

## 2 Williamson's Argument

Williamson sets the stage thus:

A fair coin will be tossed infinitely many times at one second intervals. The tosses are independent. ... Let  $H(1)$  be the event that the first toss comes up heads and  $H(2\dots)$  the event that every toss after the first comes up heads (2007, 4).

Then, according to Williamson,

$H(1\dots)$  and  $H(2\dots)$  are isomorphic events. More precisely, we can map the constituent single-toss events of  $H(1\dots)$  one–one onto the constituent single-toss events of  $H(2\dots)$  in a natural way that preserves the physical structure of the set-up just by mapping each toss to its successor.  $H(1\dots)$  and  $H(2\dots)$  are events of exactly the same qualitative type; they differ only in the inconsequential respect that  $H(2\dots)$  starts one second after  $H(1\dots)$ . That  $H(2\dots)$  is preceded by another toss is irrelevant, given the independence of the tosses. Thus  $H(1\dots)$  and  $H(2\dots)$  should have the same probability. (2007, 5)

The argument now proceeds thus: by the probability calculus,  $P(H(1\dots)) = P(H(2\dots)|H(1))P(H(1))$ , where  $H(1)$  is the event 'the first toss lands heads'. So by independence  $P(H(1\dots)) = P(H(2\dots))P(H(1))$ , i.e.  $P(H(1\dots)) = P(H(2\dots))/2$ . But  $P(H(1\dots)) = P(H(2\dots))$  by the isomorphism assumption, whence  $2P(H(1\dots)) = P(H(1\dots))$ . Hence  $P(H(1\dots)) = 0$ , contradicting Regularity.

One problem with this argument is that it had, in effect, already been refuted. To see why we need a little more detail. Formally we can represent the outcome-space of Williamson's coin-tossing 'experiment' by the uncountable set  $X = \{0,1\}^{\mathbb{N}}$  (usually written  $2^{\mathbb{N}}$ ,  $\mathbb{N} = \{1, 2, 3, \dots\}$ ,<sup>6</sup> of all infinite sequences of 1 s and 0 s (1 for a head, 0 for a tail), and the class  $F$  of events in that space is the  $\sigma$ -algebra generated by the cylinder sets (these are the subsets of  $X$  defined by fixing finitely many coordinates). The unit interval  $[0, 1]$ , under the dyadic expansions of its members,  $\sum_{n>0} a_n 2^{-n}$  where  $a_n = 0$  or  $1$ , is the set of all infinite sequences of zeros and ones (two such sequences, e.g. .0111 ... and .1000 ..., may determine the same real, in this case  $1/2$ ). Williamson assumes that the coin-flips are independent with constant probability of  $1/2$  (iid in probabilists' jargon, standing for 'independent, identically distributed'), where in his example that probability is  $1/2$ . It is well known that this probability function, were it assumed to take standard values only, can be extended to a countably additive measure on the Borel subsets of  $[0, 1]$  and hence to Lebesgue measure on the Lebesgue-measurable sets of  $[0, 1]$ . Bernstein and Wattenberg's

<sup>6</sup> Some authorities take  $\mathbb{N}$  to be the set  $\{0, 1, 2, \dots\}$ .

discussion therefore implies that all the singleton sequences in the coin-flipping experiment can be consistently assigned the same positive infinitesimal probability value, *subject to exactly the same iid + probability 1/2 of heads assumption Williamson himself makes.*

That the fair-coin sequences in  $X$ , like  $H(1 \dots)$ , can consistently receive equal infinitesimal probabilities is also implicit in a simple example of the use of Loeb spaces in Loeb (1975). The example models in a hyperfinite probability space  $(\Omega, A, \mu)$  the tossing of a fair coin infinitely many times, where  $\Omega = \{0,1\}^\eta$ , the set of all internal sequences of 0 s and 1 s of length  $\eta$  for  $\eta$  some arbitrary hyperfinite number in  ${}^*\mathbb{N}$ ,  $A$  is the set of all internal subsets of  $\Omega$ , and  $\mu$  is the counting measure which assigns internal probability  $|A|/|\Omega|$  to each  $A$  in  $A$ , where the bars signify internal cardinality.<sup>7</sup> Thus each singleton sequence in  $\Omega$  has internal probability  $2^{-\eta}$ . Where  $L(A)$ , the Loeb algebra, is the completion of the external  $\sigma$ -algebra generated by  $A$ , there are events in  $L(A)$  corresponding to the standard events defined in  $\{0,1\}^\mathbb{N}$ , with the fair-coin product measure on the latter differing from the measure  $\mu$  on their counterparts in  $L(A)$  by an infinitesimal (1975, 119). Standard events having standard probabilities can thus be recovered from such a nonstandard model by taking standard parts. In a paper published coincidentally in the same year as Williamson's, Frederik Herzberg used the same type of hyperfinite space to model a coin tossed  $\eta$  times but with an arbitrary probability  $p$  in  ${}^*[0,1]$  of landing tails, and showed using the Transfer Principle how an infinitesimal 'Bernoulli' probability is assigned to each sequence  ${}^*a$  in  $2^{*\mathbb{N}}$  and then by restriction to its counterpart  $a$  in  $2^\mathbb{N}$ .<sup>8</sup>

### 3 Where Does It Go Wrong?

In the light of the foregoing discussion it would seem that something is wrong with Williamson's argument. Apart from the claim that  $H(1 \dots)$  and  $H(2 \dots)$  are isomorphic events, only elementary probabilistic reasoning was involved, so any fault must presumably lie with that claim. Indeed, however plausible it might initially seem, a little reflection shows it to be false. As Howson (2017) pointed out, the extension of  $H(1 \dots)$  is a singleton sequence, and that of  $H(2 \dots)$  is two such sequences,<sup>9</sup> and it therefore makes no sense to say that these are isomorphic events. It seems that Williamson simply overlooked his own description of  $H(2 \dots)$  as 'all tosses after the first land heads', and hence as a disjunctive event defined in the same sample space as  $H(1 \dots)$ , and took it to be a (singleton) *sequence* in its own right containing only 1 s. That certainly seems to be implicit in his claim that

<sup>7</sup> \* Is the embedding of the standard into the nonstandard universe.

<sup>8</sup> Herzberg (2007), Appendix B.

<sup>9</sup> Thus the extension of  $H(1 \dots)$  in the outcome space  $X$  represents the singleton real  $\{1\}$  under the dyadic expansions of the reals, while the extension of  $H(2 \dots)$  represents the pair  $\{1, 1/2\}$ .

we can map the constituent single-toss events of  $H(1\dots)$  one–one onto the constituent single-toss events of  $H(2\dots)$  in a natural way that preserves the physical structure of the set-up just by mapping each toss to its successor. (2007, 5)

Possibly sensing that the isomorphism claim might be unconvincing, Williamson presented a different version omitting it, at any rate explicitly:

To make the point vivid, suppose that another fair coin, qualitatively identical with the first, will also be tossed infinitely many times at one second intervals, starting at the same time as the second toss of the first coin, all tosses being independent. Let  $H^*(1\dots)$  be the event that every toss of the second coin comes up heads, and  $H^*(2\dots)$  the event that every toss after the first of the second coin comes up heads. Then  $H(1\dots)$  and  $H^*(1\dots)$  should be equiprobable, because the probability that a coin comes up heads on every toss does not depend on when one starts tossing, and there is no qualitative difference between the coins. But for the same reason  $H^*(1\dots)$  and  $H(2\dots)$  should also be equiprobable. These two infinite sequences of tosses proceed in parallel, synchronically, and there is no qualitative difference between the coins; in particular, that the first coin will be tossed once before the  $H(2\dots)$  sequence begins is irrelevant. By transitivity,  $H(1\dots)$  and  $H(2\dots)$  should be equiprobable: [hence]  $\text{Prob}(H(1\dots)) = \text{Prob}(H(2\dots))$  (2007, 6).

But this is no improvement on Williamson's first argument.  $H(2\dots)$  is not simply a copy of  $H^*(1\dots)$ : it is, to repeat, the disjunctive event "all the tosses land heads *or* the first is a tail followed by all heads".<sup>10</sup>

My dismissal of Williamson's argument as based, in effect, on an elementary mistake might be objected to on the ground that formal probability itself endorses his claim that  $H(2\dots)$  is just as much a sequence of events as  $H(1\dots)$  itself, *and is indeed isomorphic to it*. Where  $X_i$  is a two-valued random variable defined on  $X$ , with  $X_i(x) = x_i$ ,  $i = 1, 2, \dots$ ,  $H(2\dots)$  is the sequence  $\langle X_2 = 1, X_3 = 1, \dots, X_{n+1} = 1, \dots \rangle$ , which is clearly a subsequence of  $\langle X_1 = 1, X_2 = 1, \dots, X_n = 1, \dots \rangle$ , i.e. of  $H(1\dots)$ . Thus the mapping which sends ' $X_{i+1} = 1$ ' to ' $X_i = 1$ ', does appear to be an isomorphic embedding of  $H(2\dots)$  into  $H(1\dots)$ . So is Williamson right after all in his isomorphism claim? *No*. In the usual formalism of mathematical probability  $\langle X_1 = 1, X_2 = 1, \dots, X_n = 1, \dots \rangle$  is just a *set*,<sup>11</sup> the countable intersection of all

<sup>10</sup> Benci et al. (2018) see Williamson's error as conflating two models, which they call A and B, where A is one infinitely repeated coin-toss and B another starting at the second toss of A, and they claim that in equating  $P(H(1\dots))$  and  $P(H(2\dots))$  Williamson is illegitimately equating  $P_A(H(1\dots))$  and  $P_B(H(2\dots))$  referred to model B,  $H(2\dots)$  is 'all tosses [in that model] land heads' (2018, 29)). But Williamson's second version does in effect distinguish A and B, with his  $P(H^*(1\dots))$  their  $P_B(H(2\dots))$ ; his error arises simply from his belief that  $H^*(1\dots)$  and  $H(2\dots)$  are qualitatively identical events.

<sup>11</sup> ' $X_1 = 1, X_2 = 1, \dots, X_n = 1, \dots$ ' is actually an infinite *conjunction* ' $\bigwedge_{i \in \omega} X_i = 1$ ', not a sequence. Scott and Krauss (1966) argue that it is more natural to view probabilities as defined directly on the *sentences* of an infinitary language than on the sets in the orthodox Kolmogorov formalism, whose excess structure, according to them, is redundant. Accordingly they define a countably additive probability on the sentences of a language of type  $L(\omega_1, \omega)$ , which is to first-order logic what a  $\sigma$ -field is to a field, allowing countably infinite conjunctions and disjunctions (hence  $\omega_1$ ) but only finite strings of quantifiers (hence  $\omega$ ). It has a proof-theory complete for countable sets of sentences but it is not compact.

sets  $\{x \in X: X_i(x) = 1\}$  and hence the singleton sequence consisting of 1 s only, and  $H(2 \dots)$  is the union of *two* sets, the singleton just mentioned and the singleton whose first member is 0 and all the rest 1 s. In terms of the elementary outcomes of Williamson's experiment, *two of them will win a bet on  $H(2 \dots)$  and only one a bet on  $H(1 \dots)$*  (that said, I am not implying that betting at infinitesimal odds on infinite sequences of trials is a practical policy).

## 4 Intimations of Invariance

It might be claimed that there is nevertheless a valid intuition underwriting Williamson's argument, namely that any subsequence, finite or infinite, *wherever it occurs in the original sequence of coin tosses*, should have exactly the same probability. The French mathematician Joseph Bertrand quipped that a roulette wheel has neither conscience nor memory ('elle n'a ni conscience ni mémoire' (1880, XXII)), but he could have said the same about a coin: your expectation of a bet on the event '0110001011' occurring immediately after the  $n$ th toss is the same for all  $n$ . Moreover, the intuition seems vindicated by a fundamental theorem of elementary ergodic theory for Bernoulli processes (iid processes; the coin-tossing experiment is a *one-sided* Bernoulli process), going under the name of *measure-invariance*. Under the left-shift mapping  $T$  of  $X$  into  $X$  where  $(Tx)_i = x_{i+1}$ , measure-invariance is the condition that  $P(T^{-1}A) = P(A)$  for all  $A$  in  $F$ , and it is easy to see that  $T^{-1}H(1 \dots) = H(2 \dots)$  and so  $P(H(2 \dots)) = P(H(1 \dots))$ . But those probabilities are the same only because they are *standard* probabilities both equal to 0, while as Williamson himself in effect showed, a hyperreal-valued probability function assigning positive infinitesimal values to the singleton sequences in  $2^{\mathbb{N}}$  will *necessarily* assign different infinitesimal probabilities to  $H(1 \dots)$  and  $H(2 \dots)$ , since  $P(H(2 \dots)) = 2P(H(1 \dots))$ .<sup>12</sup>

What this discussion does point to, however, is that the invariance theorems of applied probability and physics depend on the choice of standard-valued probabilities. We have seen this to be true with the measure-invariance of Bernoulli processes, and the translation and rotation-invariance properties of Lebesgue measure on the Borel sets of  $\mathbb{R}^n$  seem similarly dependent.<sup>13</sup> In Bernstein and Wattenberg's example assigning nonzero infinitesimal probabilities to certain sets of Lebesgue measure zero causes translation-invariance to be violated, and correspondingly rotational invariance when the half-open unit interval  $[1, 0)$  is wrapped around the circumference of a circle though the violations are only infinitesimal.<sup>14</sup> Seeing

<sup>12</sup> If instead of  $H(2 \dots)$  we consider  $H(n \dots)$  for arbitrary  $n$  we will of course find increasingly 'large' infinitesimal differences between their probabilities. But a 'large' infinitesimal, even a 'very large' infinitesimal, is still an infinitesimal.

<sup>13</sup> Lebesgue measure uniquely among standard measures has these properties, and for that reason is taken as the model of situations where such invariance seems implicit in the physical structure and dynamics of the relevant systems, for example a person throwing a dart 'at random' at a dartboard, frictionless spinners stopping 'at random', and so forth.

<sup>14</sup> This example is taken from Barrett (2009, 77). With addition and multiplication mod 1, let  $S$  be the set  $\{a, 2a, 3a, \dots\}$ , where  $a$  is an irrational in  $[0, 1)$  (ensuring that all the members of  $S$  are distinct), and

infinitesimals as measuring real differences in fine-structure might therefore seem to conflict with the lesson contemporary physics appears to teach, that the symmetries expressed in the classic invariance results are a fundamental feature of nature. Barrett (2010) blames the practice of using external models for the conflict:

In substantive applications of the theory of chance, there will typically be found constraints on a model arising from physical properties or symmetries in the physical situation. Because we cannot apply the Transfer Principle to an external nonstandard model, the task of showing that an external nonstandard model conforms to these constraints may become difficult or impossible. The lack of true translation invariance in Bernstein's and Wattenberg's model is an example of this. (2010, 73; for an internal model the Transfer Principle shows that Lebesgue measure is indeed translation-invariant).

But I fail to see this argument as warranting the rejection of external models (to be fair Barrett says that he himself does not regard it as conclusive (*ibid.*)). In fact the practice he condemns<sup>15</sup> is the basis of a highly successful field of research in applied probability: external Loeb spaces have found manifold applications in statistics, physics, classical and quantum (see for example the list in Albeverio 2009), and even economics, where hyperfinite populations have proved a very useful analytical tool (Anderson 1991; infinite hyperfinite sets, though formally finite, are actually uncountable). And what if the 'true' translation invariance of Lebesgue measure does fail in an external model? It does so only infinitesimally, which might seem in empirical terms a more-than-acceptable margin of error.<sup>16</sup> That being so, the postulates on which the standard invariance results rest would arguably seem little the worse for being judged as what physicists call 'effective' theories.

It might of course be that some suitable internal model is 'really' fundamental. According to two advocates of Nelson's Internal Set Theory (IST) in which the entire mathematical universe is nonstandard since it is a universe just of internal sets,<sup>17</sup> IST is the mathematical theory actually better suited to describe reality:

---

Footnote 14 (continued)

let  $S_a$  be  $S$  shifted by  $a$ .  $S$  differs from  $S_a$  by  $\{a\}$ , and so differs in probability by an infinitesimal. The Lebesgue measure of all three sets is the same, 0.

<sup>15</sup> It is also condemned by Bascelli et al. (2014), claiming that 'once one decides to use hyperreal infinitesimals, one should also replace the original algebra of propositions in which the agent has credences' with an internal algebra of the hyperreal setting.' (2016, 11). But why 'should' one? The authors claim that the internal hyperfinite setting described in Sect. 3 'allows one to avoid ... the problems raised by Williamson's argument' (*ibid.*), but it is not clear what the alleged problems are; the authors merely point out that in that setting  $H(1 \dots)$  and  $H(2 \dots)$  are shown to be non-isomorphic by the Transfer Principle. But if I am correct one doesn't need to appeal to an internal probability space to refute the isomorphism claim.

<sup>16</sup> A no higher standard of accuracy justifies the use of the real-number continuum itself. Gisin argues that the infinite amount of information contained in irrational numbers (a set of measure 1 in the reals) is not only redundant for the purposes of physical description but in effect renders them essentially random (2018).

<sup>17</sup> In this theory the real line corresponds to the internal set  ${}^*\mathbb{R}$  in the orthodox approach, rather than the standard real line  $\mathbb{R}$  (which is now not a set).

Mathematics is intended to model reality as well as possible. The nonstandard universe gives a richer model of the reals than the standard universe, so it is natural to identify the real line with  $\mathbb{R}$  from IST. From this point of view the standard objects are merely ‘shadows’ of the real objects: taking standard parts loses information. (Diener and Stroyan 2009, 260–261).

The loss is not all on one side though. Internal models do not contain the familiar number systems, the natural numbers, rationals and reals, as *sets*, nor are they closed under countable Boolean operations,<sup>18</sup> which is why external Loeb spaces are so useful. On balance, there seem to be good reasons for adopting a liberal, ‘ecumenical’ approach in which both external and internal nonstandard models are acknowledged to have important roles to play in extending scientific understanding.

## 5 Conclusion

Barrett’s interesting discussion cited in the preceding section concludes that ‘we can have a [nonstandard] model which is regular, but which is non-translation invariant or external or both, or we can have an internal, translation invariant model which is not regular’ (2010, 74). In either case, internal or external, there are as we have seen nonstandard models of infinite flips of a fair coin satisfying the iid conditions, with Williamson’s own argument contesting that possibility refuted *avant la lettre* both by Bernstein and Wattenberg’s results and by Loeb’s.

In general, good arguments *for* Regularity seem thin on the ground, and Lewis’s ‘Zero chance is *no* chance, and nothing with zero chance ever happens’ is simply bluster: few if any probabilists appear to balk at assigning probability 0 to the logical, if not practical, possibility of a fair coin by chance yielding an infinite sequence of heads, as the strong law of large numbers dictates (hence the jargon ‘almost everywhere’, ‘almost surely’ and ‘almost certainly’). For Lewis in effect to claim that such a practice is conducted in ignorance of a simple truth seems to be an example of philosopher’s hubris of a high order.

## References

- Albeverio, S. (2009). Applications of nonstandard analysis in mathematical physics. In Albeverio et al. (Eds.), (pp. 182–220).
- Albeverio, S., Fenstad, J. E., Høegh-Krohn, R., & Lindstrom, T. (2009). *Nonstandard methods in stochastic analysis and mathematical physics*. New York: Dover.
- Anderson, R. M. (1991). Nonstandard analysis with applications to economics. In W. Hildenbrand & H. Sonnenschein (Eds.), *Handbook of mathematical economics* (Vol. 4, pp. 2145–2208). Amsterdam: Elsevier.
- Barrett, M. (2010). The possibility of infinitesimal chances. In E. Eells & J. H. Fetzer (Eds.), *The place of probability in science: In honor of ellery eells (1953–2006)* (pp. 65–79). Dordrecht: Springer.

<sup>18</sup> No countably infinite union of disjoint sets is in any internal algebra.



- Bascelli, T., Bottazzi, E., Herzberg, F., Kanovei, V., Katz, K., Katz, M., et al. (2014). *Fermat, Leibniz, Euler, and the gang: The true history of the concepts of limit and shadow*. [arXiv:1407.0233v1](https://arxiv.org/abs/1407.0233v1). [Math.HO].
- Benci, V., Horsten, L., & Wenmackers, S. (2018). Infinitesimal Probabilities. *British Journal for the Philosophy of Science*, 69, 502–522.
- Bernstein, A. R., & Wattenberg, F. (1969). Nonstandard measure theory. In W. A. J. Luxemburg (Ed.), *Applications of model theory to algebra, analysis and probability* (pp. 171–186). New York: Holt, Rinehart and Winston.
- Bertrand, J. (1880). *Calcul des probabilités*. Paris: Gauthier-Villars.
- Diener, F., & Stroyan, K. D. (2009). Syntactical methods in infinitesimal analysis. In Albeverio et al. (Eds.), (pp. 258–281).
- Gisin, N. (2018). *Indeterminism in physics, classical chaos and Bohmian mechanics. Are real numbers really real?* Group of Applied Physics, University of Geneva. [arXiv:1803.06824v2](https://arxiv.org/abs/1803.06824v2).
- Herzberg, F. (2007). Internal laws of probability, generalized likelihoods and lewis' infinitesimal chances: A response to Adam Elga. *British Journal for the Philosophy of Science*, 58, 25–43.
- Herzberg, F. (2008). A definable nonstandard enlargement. *Mathematical Logic Quarterly*, 54, 167–175.
- Herzberg, F., Kanovei, V., Katz, M., & Lyubetsky, V. (2018). Minimal axiomatic frameworks for definable hyperreals with transfer. *Journal of Symbolic Logic*, 83, 385–391.
- Howson, C. (2017). Regularity and infinitely tossed coins. *European Journal for the Philosophy of Science*, 7, 97–100.
- Kanovei, V., & Shelah, S. (2004). A definable non-standard model of the reals. *Journal of Symbolic Logic*, 69, 159–164.
- Keisler, H. J. (2007). *Foundations of infinitesimal calculus*. Madison: University of Wisconsin-Madison.
- Lewis, D. (1976). *Philosophical papers* (Vol. 2). Oxford: Oxford University Press.
- Loeb, P. A. (1975). Conversion from nonstandard to standard measure spaces and applications in probability theory. *Transactions of the American Mathematical Society*, 211, 113–122.
- Nelson, E. (1977). Internal set theory: A new approach to nonstandard analysis. *Bulletin of the American Mathematical Society*, 83, 1165–1198.
- Scott, D., & Krauss, P. (1966). Assigning probabilities to logical formulas. In J. Hintikka & P. Suppes (Eds.), *Aspects of inductive logic* (pp. 219–264). Amsterdam: North Holland.
- Wenmackers, S., & Horsten, L. (2013). Fair infinite lotteries. *Synthese*, 190, 37–61.
- Williamson, T. (2007). How probable is an infinite sequence of heads? *Analysis*, 67, 173–180.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.