RETHINKING CONVERGENCE TO THE TRUTH*

Convergence to the truth is viewed with ambivalence in philosophy of science. On the one hand, methods of inquiry that lead to the truth in the limit are prized as marks of scientific rationality.[1] But someone who, by using a particular method, expects to always converge to the truth seems to fail a minimum standard of epistemic modesty.

This last point was brought home by Gordon Belot in his critique of Bayesian epistemology.[2] Belot uses the staple example of flipping a coin infinitely often. A famous result in standard probability theory, the martingale convergence theorem, implies that, with probability one, a Bayesian agent expects to know the truth for any hypothesis about coin flips as new data accumulates.

The problem, Belot argues, is that this result is true for arbitrary hypotheses, including those that are compatible with any finite stream of data, like the set of sequences of coin flips that are eventually constant or the set of sequences that are periodic. Some of these hypotheses might be such that their truth value could never reasonably be approximated by any finite stream of data. It seems implausible that one should in such a case be *required* to assign probability one to converging to the truth. Belot backs this up by topological considerations, demonstrating that under certain conditions failure to converge to the truth constitutes a topologically large event; he takes this to be a good reason for being modest as regards convergence to the truth. But by the martingale convergence theorem that kind of modesty is

*I am grateful for helpful feedback from audiences at the University of Washington, Columbia University, the Munich Center for Mathematical Philosophy, and the Formal Rationality Forum (Bristol-LSE-Michigan-Irvine). I am especially thankful for comments from Nikhil Addleman, Jeff Barrett, Gordon Belot, Persi Diaconis, Haim Gaifman, Daniel Herrmann, Frederik Herzberg, Conor Mayo-Wilson, Louis Narens, Michael Nielsen, Jan-Willem Romeijn, Teddy Seidenfeld, Brian Skyrms, Rush Stewart, Francesca Zaffora Blando, Kevin Zollman, and to an anonymous referee for this JOURNAL.

[1] Such views are, for instance, expressed (in different ways) by C. S. Peirce, "The Fixation of Belief," *Popular Science Monthly*, XII (November 1877): 1–15; and L. J. Savage, "Implications of Personal Probability for Induction," this JOURNAL, LXIV, 19 (October 1967): 593–607. They are a mainstay of Bayesian epistemology; see, for example, James M. Joyce, "The Development of Subjective Bayesianism," in D. M. Gabbay, S. Hartmann, and J. Woods, eds., *Handbook of the History of Logic, Vol. 10: Inductive Logic* (Amsterdam: Elsevier, 2011), pp. 415–76.

[2] Gordon Belot, "Bayesian Orgulity," *Philosophy of Science*, LXXX, 4 (October 2013): 483–503.

0022-362X/22/1907/380–403

impossible for the Bayesian agent. Bayesian epistemology forces an agent into the straightjacket of having to believe in convergence to the truth. Belot concludes that something is wrong with Bayesian epistemology. (I consider his argument in more detail in section I.)

Belot's argument has engendered a variety of responses. One can show that convergence to the truth is not guaranteed for imprecise probabilities and finitely additive probabilities: imprecise Bayesians and Bayesians who reject countable additivity escape the charge of immodesty.[3] Others argue that the topological considerations used by Belot are not relevant for probability theory.[4]

One might, however, be worried about Bayesian convergence to the truth even if one puts aside Belot's topological argument. I share the concern expressed by Belot and some of his commentators that there is something troubling about applying the traditional convergence to the truth result to certain infinite hypotheses. But I do not think that this requires a radical revision of Bayesian epistemology. It rather calls for a more careful study of infinite hypotheses.

I am going to suggest that standard probability theory is not well equipped for a sufficiently fine-grained analysis of this kind. There is, more specifically, only a rough sense of when an infinite hypothesis can be approximated by finite evidence. I will explore an alternative to the standard theory that is based on nonstandard models of probability theory. These models involve arbitrarily large and arbitrarily small numbers—numbers that are, respectively, larger than any finite number or infinitesimally close to, but distinct from, zero. Nonstandard models have been studied in mathematical logic, especially in the seminal work of Abraham Robinson.[5] I will provide some background on nonstandard analysis in section II before introducing nonstandard probability theory as developed by Edward Nelson in section III.[6] The main result, stated in section IV, shows that within the nonstandard framework, convergence to the truth fails with (non-infinitesimal) positive probability for certain hypotheses: those that

[3] On imprecise probabilities, see Brian Weatherson, "For Bayesians, Rational Modesty Requires Imprecision," *Ergo*, II (2015). On finite additivity, see Adam Elga, "Bayesian Humility," *Philosophy of Science*, LXXXIII, 3 (July 2016): 305–23.

[4] See Jessi Cisewski et al., "Standards for Modest Bayesian Credences," *Philosophy of Science*, LXXXV, 1 (January 2018): 53–78. A similar view is discussed in Simon Huttegger, "Bayesian Convergence to the Truth and the Metaphysics of Possible Worlds," *Philosophy of Science*, LXXXII, 4 (October 2015): 587–601.

[5] Abraham Robinson, *Non-standard Analysis* (Amsterdam: North-Holland, 1966).

[6] Edward Nelson, *Radically Elementary Probability Theory* (Princeton, NJ: Princeton University Press, 1987).

one expects to be beyond the reach of finite streams of evidence. Finally, in section v I argue that this creates a space for modesty within Bayesian epistemology.

I. CONVERGENCE TO THE TRUTH

Martingales are part and parcel of probability theory. One can think of a martingale as the probabilistic analog of a constant sequence of numbers: a martingale is *constant on average*, suggesting that it, too, does converge. *Doob's martingale convergence theorem* shows that this is indeed the case.[7] Martingales almost surely converge to a limit under significantly broader sufficient conditions than those assumed by the strong law of large numbers.

A special case of the martingale convergence theorem speaks to convergence to the truth.[8] It holds quite generally but can be explained in terms of repeatedly flipping a coin without much loss of generality. By associating heads with one and tails with zero, repeated coin flips can be associated with *Cantor space*: the set of all infinite zero-one sequences with an appropriate set of propositions (sets of infinite binary sequences) that describe many events of interest.[9] Many of these events are finite, such as observing at least one zero in the first ten trials, or having exactly 52 one's between the 100th and the 200th trial. Other events are of an infinite nature, such as the event that the limiting relative frequency of ones converges.

A *probability measure* assigns a numerical probability to every proposition of Cantor space. In Bayesian epistemology, the probability measure represents an agent's prior degrees of belief: the probability of a proposition A is the degree to which the agent believes A to be true. The *conditional probability* given a finite sequence of zeroes and ones of length n represents the agent's posterior degree of belief after having made n observations.[10] As n grows, the agent's evidence increases and the agent's posteriors become better informed.

Within this experimental setup, the martingale convergence theorem implies the following. Take any proposition, A, of Cantor space and consider the conditional probability of A. With prior probability

[7] See Joseph L. Doob, *Stochastic Processes* (New York: Wiley, 1953).

[8] The special result was proved by Paul Lévy, *Théorie de l'addition des variables aléatoires* (Paris: Gauthier-Villars, 1937).

[9] The propositions form a σ-algebra, the Borel σ-algebra generated by the open sets of the topology of weak convergence.

[10] In the present setting, conditional probability can itself be taken as a probability measure.

one, as the evidence increases the conditional probability of A converges to one if A is true and to zero if A is false: posterior probabilities of propositions converge almost surely to the truth.

More generally, one often encounters not just propositions but random variables.[11] Random variables allow us to express events involving numerical quantities. For example, consider the random variable that gives the relative frequency of ones in the first n segments of a binary sequence. That random variable may assume any of the values $0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1$. Another important random variable is the limiting relative frequency of ones as n goes to infinity. Propositions can be thought of as special types of random variables.[12]

Expectations and *conditional expectations* are for random variables what probabilities and conditional probabilities are for propositions. The expectation of a random variable, X, represents the best estimate of X prior to making any observations. The conditional expectation of X given the initial n digits of a binary sequence represents the agent's posterior best estimate given the evidence available at stage n. The martingale convergence theorem implies that, with prior probability one, the conditional expectations of X converge to the true value of X as n increases.[13] Since propositions can be represented as random variables, the result about conditional probabilities stated above is a special case. In what follows I will refer to this as the *convergence to the truth theorem*.

The convergence to the truth theorem is of great significance for Bayesian epistemology. One's initial opinions may be based on little or no evidence; they are thus open to the charge of arbitrariness. Convergence to the truth says that for certain hypotheses one expects the starting point to not matter very much in the long run. This illustrates why subjective Bayesians emphasize rational opinion change and put little weight on the choice of a "rational" prior.[14]

[11] A random variable is a measurable function from Cantor space to the reals.

[12] The proposition A can be identified with its indicator variable that takes on the value 1 if A is true and 0 otherwise.

[13] I gloss over a number of important technical details in the statement of these results, such as the existence and nature of conditional probabilities and expectations or assumptions about random variables. For details, see, for example, David K. Williams, *Probability with Martingales* (Cambridge, UK: Cambridge University Press, 1990).

[14] There are other results that can be used for this line of reasoning, such as the Blackwell-Dubins theorem on merging of opinions; see David Blackwell and Lester Dubins, "Merging of Opinions with Increasing Information," *The Annals of Mathematical Statistics*, XXXIII, 3 (September 1962): 882–86. Merging of opinions says that, under certain conditions, the conditional probabilities of two agents who update on the same evidence get closer. An analogous result was proved independently in a different framework based on rich formal languages by Haim Gaifman and Marc Snir, "Probabilities

Some commentators have put forward a less flattering take on convergence to the truth. It has been pointed out—correctly—that the theorem does not establish *actual* convergence to the truth, but only an agent's *expectation* to converge to the truth.[15] Others have aimed their criticism at the role that countable additivity plays in the proof of the theorem.[16] More recently, Belot has challenged the status of convergence to the truth as one of the pillars of Bayesian epistemology, arguing that it is instead a genuine liability for Bayesians since it reveals them as irredeemably immodest.[17]

Let us take a closer look at Belot's argument, which is set in Cantor space. Conditional probabilities almost trivially converge to the truth for finite events, such as observing at least one zero in the first ten periods. But convergence to the truth also applies to events whose truth is not determined in a finite number of steps, and it is with

---

over Rich Languages, Testing and Randomness," *The Journal of Symbolic Logic*, XLVII, 3 (September 1982): 495–548. Gaifman and Snir's theorem combines merging of opinions, which is about a community of Bayesian agents, with the convergence to the truth result. For further discussions of the Blackwell-Dubins theorem, see Mark J. Schervish and Teddy Seidenfeld, "An Approach to Consensus and Uncertainty with Increasing Information," *Journal of Statistical Planning and Inference*, XXV, 3 (July 1990): 401–14; Simon Huttegger, "Merging of Opinions and Probability Kinematics," *The Review of Symbolic Logic*, VIII, 4 (December 2015): 611–48; Simon Huttegger, *The Probabilistic Foundations of Rational Learning* (Cambridge, UK: Cambridge University Press, 2017); and Rush T. Stewart and Michael Nielsen, "Another Approach to Consensus and Maximally Informed Opinions with Increasing Evidence," *Philosophy of Science*, LXXXVI, 2 (March 2019): 236–54. Belot's argument does not refer to merging of opinions. But since merging of opinions also involves a probability one qualification, it is not difficult to see that the issue of immodesty arises again. As an antidote, the nonstandard model of convergence to the truth can also be applied to merging of opinions, with similar results.

[15] See Clark Glymour, *Theory and Evidence* (Princeton, NJ: Princeton University Press); and John Earman, *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory* (Cambridge, MA: MIT Press, 1992). Michael Nielsen, "Deterministic Convergence and Strong Regularity," *The British Journal for the Philosophy of Science*, LXXI (December 2020): 1461–91, studies convergence to the truth without the probability one qualification.

[16] Joseph B. Kadane, Mark J. Schervish, and Teddy Seidenfeld, "Reasoning to a Foregone Conclusion," *Journal of the American Statistical Association*, XCI, 435 (September 1996): 1228–35; and Cory Juhl and Kevin T. Kelly, "Realism, Convergence, and Additivity," *PSA: Proceedings of the Philosophy of Science Association* (1994): 181–89. There are, however, finitely additive versions of the martingale convergence theorem; see Roger A. Purves and William D. Sudderth, "Some Finitely Additive Probability," *The Annals of Probability*, IV, 2 (April 1976): 259–76; and Sandy L. Zabell, "It All Adds Up: The Dynamic Coherence of Radical Probabilism," *Philosophy of Science*, LXIX (September 2002): 98–103. More specifically, the result holds for a natural class of finitely additive measures, called *strategic measures*.

[17] See Belot, "Bayesian Orgulity," *op. cit.*; and Gordon Belot, "Objectivity and Bias," *Mind*, CXXVI, 501 (July 2017): 655–95. Precursors of this argument can be found in A. P. Dawid, "The Well-Calibrated Bayesian," *Journal of the American Statistical Association*, LXXVII, 379 (September 1982): 605–10; and Juhl and Kelly, "Realism, Convergence, and Additivity," *op. cit.*

some of these that Belot takes issue. He invites us to think of *countable dense hypotheses* in Cantor space. A countable dense hypothesis H is a proposition that is compatible with any finite piece of evidence: at any finite stage it is impossible to rule out H. But since H is countable and dense, its complement is uncountable and dense, and so the complement of H is also compatible with any finite piece of evidence. Hence, at no finite stage are we able to distinguish between H and its complement. As examples, think of the hypothesis that the binary sequence you are observing is eventually constant or the hypothesis that it is periodic; both are countable and dense in Cantor space.[18]

The convergence to the truth theorem implies that for *every* prior probability measure the conditional probabilities of *any* dense hypothesis converges to the truth with prior probability one. Belot considers probability measures with a special kind of open-mindedness property, which says that conditional probabilities of H never fully settle down at finite stages. Based on this assumption, he shows that the failure set—the subset of infinite binary sequences for which conditional probabilities do not converge to the truth—is *comeager*; this is a topological way of saying that the failure set is large, or typical.[19] Thus, the failure set is large in a topological sense, but small in a probabilistic sense (since it has probability zero). Belot takes this to show that Bayesian agents ignore the myriad ways in which their conditional probabilities might fail to converge to the truth, revealing a problematic kind of epistemic arrogance.

I am not going to respond to Belot's topological argument here. I think that worries about convergence to the truth go beyond it. Standard probability theory identifies infinite hypotheses—like countable dense hypotheses in Cantor space—with events whose truth values are unaffected by any finite number of observations.[20] Otherwise these events are all treated on a par. In particular, there seems to be no straightforward way to draw a distinction between hypotheses that can be *approximated* by finite streams of evidence and those about which one can, loosely speaking, always be misled. Belot tries to capture this

[18] Belot, "Bayesian Orgulity," *op. cit.* A binary sequence is eventually constant if it consists exclusively of ones, or exclusively of zeroes, from some point onward. Since the sequence becoming constant can happen at any stage, no finite number of observations is enough to determine the truth of the hypothesis or its complement. Likewise, a periodic sequence can cycle through arbitrarily long patterns, which can never be confirmed or refuted with certainty.

[19] A set is meager if it is the countable union of nowhere dense sets. A comeager set is the complement of a meager set.

[20] Formally, infinite hypotheses are in the tail-σ-algebra of a probability space, which captures events in the remote future.

in terms of his open-mindedness condition, with no effect on how infinite hypotheses are treated within standard probability theory. But one does not need to conclude from this that Bayesian agents are immodest. One can instead try to enrich standard probability theory to account for finite approximations of infinite hypotheses *without changing its basic assumptions*. Leaving the basic assumptions in place is important in order to give a response to Belot's argument from the standpoint of orthodox probability theory instead of revising the standard theory.[21] This strategy, if successful, can lead to a better understanding of convergence to the truth for infinite hypotheses and its connection to epistemic modesty.

There is more than one way of doing this. One approach involves *measure algebras*.[22] A measure algebra identifies infinite hypotheses with those events that cannot be distinguished from it in a finite number of steps (with respect to a suitable background probability measure). This approach provides a couple of insights, but it deals with the concerns about convergence to the truth by explaining away (although in a principled manner) as empirically meaningless distinctions between hypotheses that can only be drawn at the infinite limit.

The approach I take here goes in the opposite direction. The measure algebra approach takes structure away from probability spaces. *Nonstandard models* add structure by way of introducing arbitrarily large and small numbers. We shall see below that this leads to a new understanding of when infinite hypotheses can, and cannot, be approximated by finite streams of evidence. Before diving into the details, I will put some context to nonstandard models.

II. NONSTANDARD MODELS

The familiar objects of analysis—real numbers—are finite (bounded above), and no real number is infinitely small. Yet, starting with Leibniz and lasting into the early nineteenth century, mathematicians and physicists made free use of infinitely large and small numbers.[23]

[21] Weatherson, "For Bayesians, Rational Modesty Requires Imprecision," *op. cit.*; and Elga, "Bayesian Humility," *op. cit.*, are revisionary responses. Weatherson gives up the assumption of precise probability assignments, and Elga the assumption of countable additivity.

[22] Huttegger, "Convergence to the Truth and the Metaphysics of Possible Worlds," *op. cit*. For measure algebras, see Andrei N. Kolmogorov, "Algèbres de Boole métriques complètes," *Zjazd Matematyków Polskich*, XX (1948): 21–30, translated by R. C. Jeffrey as "Complete Metric Boolean Algebras," *Philosophical Studies*, LXVII, 1 (January 1995): 57–66.

[23] The existence of such numbers violates the Archimedean property, which says that for any two real numbers there exists a multiple of the first that is greater than the second.

Infinitely small numbers—or *infinitesimals*—were employed by Leibniz in his development of the calculus. The differential calculus is concerned with how the values of a function change as its arguments vary continuously. Leibniz took the relevant quantity to be a ratio of infinitesimals—the infinitesimal change in function value over the infinitesimal change in its argument. Under certain assumptions, the resulting equations can be treated algebraically, that is, like equations between ordinary real numbers.

This is part of a larger package. Infinitesimals are smaller than any ordinary real number: they are squeezed between zero and the reciprocal of any standard number (regardless how large). Each ordinary real number can be pictured as being surrounded by a cloud of infinitesimals that can only be seen through an "infinitesimal" microscope.[24] Infinitesimals often can be treated like ordinary numbers. In particular, they can be added and multiplied in much the same way.

The reciprocal of an infinitesimal is a *hyperfinite* or *unlimited* number—a number that surpasses any finite bound.[25] Hyperfinite numbers can be pictured as visible through an "infinite" telescope; they extend the real line beyond what is viewable in ordinary analysis.[26] With some care one can once again perform many elementary calculations with hyperfinite numbers in the usual way.

While in practice it is usually quite clear how to do calculations with infinitesimals and unlimited numbers, the trouble with Leibniz's approach is its reliance on an intuitive understanding of infinitesimals. In particular, the principles governing when exactly infinitesimals behave like ordinary real numbers remained unclear. The now-orthodox approach to calculus, developed in the nineteenth century by mathematicians like Weierstrass, did away with infinitesimals altogether and replaced them with limits of sequences and functions. While this gave rise to a rigorous foundation of analysis, certain developments in twentieth-century mathematical logic, especially in model theory, led to a reappraisal of infinitesimals. Skolem demonstrated the existence of nonstandard models of arithmetic in the 1930s. These models obey all first-order laws of the standard theory of arithmetic—loosely, all laws that govern the ordering of, and operations on, ordinary natural numbers—but can contain infinitely large elements.[27]

[24] As described in Jerome H. Keisler, *Elementary Calculus: An Infinitesimal Approach* (Boston: Weber and Schmidt, 1986).

[25] Since a positive infinitesimal is smaller than the reciprocal of any ordinary positive number, its reciprocal is larger than any ordinary positive number.

[26] Keisler, *Elementary Calculus*, *op. cit.*

[27] Thoralf Skolem, "Über die Nicht-charakteriesierbarkeit der unendlichen Zahlenreihe mittels endlich oder abzählbar unendlich vieler Assagen mit ausschliesslich

Abraham Robinson used model theory to put nonstandard analysis on a firm footing.[28] Robinson distinguished between *standard* and *nonstandard* numbers, the latter being rigorous formal counterparts to what I called infinitesimals and hyperfinite numbers, and he proved the existence of nonstandard models of the first-order theory of analysis.

There are several ways to construct nonstandard numbers.[29] They all follow the trail blazed by Robinson, who introduced the basic principles that govern standard and nonstandard numbers. These principles guarantee that nonstandard analysis is a *conservative extension* of standard analysis. The conservation has several aspects. First, the set of standard real numbers is contained in the new structure, and standard ordering and functions have natural nonstandard extensions (that is, the nonstandard extensions coincide with their standard counterparts when applied to standard reals). Second, every finite nonstandard number is infinitesimally close to a unique standard real number. Finally, the *transfer principle* says that all first-order laws about standard real numbers also govern nonstandard numbers. First-order laws speak about numbers only (not about sets of numbers, or sets of sets of numbers, and so on). The transfer principle explains what was left vague in the old infinitesimal calculus—namely, what exactly infinitesimals and hyperfinite numbers have in common

---

Zahlenvariablen," *Fundamenta Mathematicae*, XXIII, 1 (1934): 150–61. Skolem's earlier work on nonstandard models dealt with set theory; see Thoralf Skolem, "Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre," *Proceedings of the 5th Scandinavian Mathematical Congress* (1922): 217–32. It follows from the Löwenheim-Skolem theorem that if there is a structure that satisfies the axioms of set theory, there exists a countable structure that does so as well: set theory has a countable model if it has a model at all. The contrast between this and the fact that set theory asserts the existence of uncountable sets is known as *Skolem's paradox*. See Haim Gaifman, "Nonstandard Models in a Broader Perspective," in A. Enayat and R. Kossak, eds., *Nonstandard Models of Arithmetic and Set Theory* (Providence, RI: American Mathematical Society, 2004), pp. 1–22, for a discussion of Skolem's paradox, nonstandard models of arithmetic, and the broader ramifications of nonstandard models.

[28] Robinson, *Non-standard Analysis*, *op. cit.*

[29] Robinson used an ultrapower construction, in which standard and nonstandard numbers are associated with infinite sequences of real numbers; see, for example, chapter 3 of Robert Goldblatt, *Lectures on the Hyperreals: An Introduction to Nonstandard Analysis* (Berlin: Springer, 1998). Another approach applies the compactness property of first-order theories; see, for example, chapter 4 of Sean Walsh and Tim Button, *Philosophy and Model Theory* (Oxford: Oxford University Press, 2018). Other developments are based on nonstandard set theories; see Edward Nelson, "Internal Set Theory: A New Approach to Nonstandard Analysis," *Bulletin of the American Mathematical Society*, LXXXIII, 6 (November 1977): 1165–98; Karel Hrbáček, "Axiomatic Foundations of Nonstandard Analysis," *Fundamenta Mathematicae*, XCVIII, 1 (1978): 1–19; and Vladimir G. Kanovei, "Undecidable Hypotheses in Edward Nelson's Internal Set Theory," *Russian Mathematical Surveys*, XCVI, 6 (1991): 1–54.

with standard reals. Every true statement about real numbers is also true in the nonstandard extension. Yet true statements about sets of real numbers may fail for sets of nonstandard numbers.[30] That nonstandard analysis is a conservative extension of ordinary analysis thus means that a copy of the old theory is part of the new one and that the new elements are treated to some extent in the same way as the old ones; in particular, from a first-order perspective infinitesimals behave like small real numbers and unlimited numbers like large finite ones.

The conservative nature of nonstandard analysis may also help with certain philosophical puzzlements as to the nature of nonstandard numbers. Are nonstandard numbers genuinely new and distinct from standard numbers? Or were they there "all along"? Different positions are possible.[31] For what follows (and for other mathematical applications) adopting one or another position makes no difference. Following Robinson, one may follow a nonstandard methodology simply for instrumental reasons. On this view, nonstandard models of analysis are useful tools for proving results about finite real numbers. As long as the nonstandard model conserves the standard model, no harm is done by using the tool.[32]

III. NONSTANDARD PROBABILITY

We are going to work with a type of nonstandard probability space that was introduced by Nelson, whose theory is known as *radically elementary probability theory*.[33] According to Nelson, a probability space $(\Omega, \mathbb{P})$ satisfies the following two conditions:

[30] As an example, consider the *completeness property*, which says that a set, S, of (standard) real numbers is an interval if S contains any number between x and y whenever both x and y are in S. The set of all infinitesimals has the latter property: whenever two nonstandard numbers are infinitesimally close to zero, so is any number between them. But that set fails to be an interval. See Keisler, *Elementary Calculus*, *op. cit.*

[31] Leibniz arguably viewed infinitesimals as idealizations; see Mikhail G. Katz and David Sherry, "Leibniz's Infinitesimals: Their Fictionality, Their Modern Implementations, and Their Foes from Berkeley to Russell and Beyond," *Erkenntnis*, LXXVIII, 3 (June 2013): 571–625. Other positions are outlined in sections 5.6 and 5.7 of Peter Fletcher et al., "Approaches to Analysis with Infinitesimals following Robinson, Nelson, and Others," *Real Analysis Exchange*, XLII, 2 (Fall 2017): 193–252.

[32] See Abraham Robinson, "Formalism 64," in Y. Bar-Hillel, ed., *Logic, Methodology and Philosophy of Science* (Amsterdam: North-Holland, 1965), pp. 228–46. Robinson's position is close in spirit to Hilbert's formalism—of course, without Hilbert's commitment to eliminating reasoning that involves the actual infinite; see the discussions in Gaifman, "Non-standard Models in a Broader Perspective," *op. cit.*, and Walsh and Button, *Philosophy and Model Theory*, *op. cit.* Walsh and Button also present a precise characterization of the way in which elementary extensions of the field of real numbers conserve the standard system of analysis (their Proposition 4.18).

[33] Nelson, *Radically Elementary Probability Theory*, *op. cit.*

(i) Ω is the set of elementary events. The number of elements in Ω is a standard or a nonstandard natural number.
(ii) The probability function $\mathbb{P} : \Omega \to [0, 1]$ assigns real numbers (including infinitesimals) to elementary events such that $\mathbb{P}\{\omega\} > 0$ for all $\omega \in \Omega$ and

$$\sum_{\omega \in \Omega} \mathbb{P}\{\omega\} = 1 \quad \text{as well as} \quad \mathbb{P}A = \sum_{\omega \in A} \mathbb{P}\{\omega\},$$

where the event A may be any subset of Ω.

To illustrate, consider the nonstandard analog of Cantor space: the set of all binary sequences of length ν, where ν is an (arbitrary) nonstandard natural number. The cardinality of this set is 2^ν, another nonstandard number. Since ν exceeds any standard natural number, the nonstandard probability space is a legitimate model of flipping a coin arbitrarily often.

Nelson proved radically elementary analogs of many of the central results of standard probability theory, such as the strong law of large numbers, the martingale convergence theorem, and the central limit theorem.[34] One salient feature of this framework is that the standard concept of convergence branches into two nonstandard concepts, only one of which corresponds to finite approximations. This is at the heart of my treatment of convergence to the truth. I shelve the topic of convergence for now, however, and return to it in the next section in order to introduce a few more concepts first. (Appendix A has more information.)

Certain issues of integration and measurability that arise in standard probability theory do not play a role in the present framework. In standard probability theory there are non-measurable sets,[35] and expectations are in general given by integrals and not by sums. According to the second clause in the definition of a probability space, *every* subset of Ω is assigned a probability (there are no non-measurable sets). And since probabilities of subsets of Ω are given by sums, the expected value of any random variable is also given by a sum (and not by an integral). If $X : \Omega \to \mathbb{R}$ is a random variable, its expected value is

$$\mathbb{E}X = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}\{\omega\}.$$

[34] Nelson, *Radically Elementary Probability Theory*, *op. cit.* See Frederik S. Herzberg, *Stochastic Calculus with Infinitesimals* (Berlin: Springer, 2013), for more results.

[35] Events in a probability space that cannot be assigned a probability.

Here there is a methodological continuity with finite probability spaces that derives from the fact that unlimited numbers are treated in nonstandard models like limited numbers (by the transfer principle).

The second clause says that every element of Ω has a (perhaps infinitesimal) positive probability. The resulting probability function is *regular*: no event other than the empty set is assigned probability zero. Some commentators have argued in favor of regularity, a view which has not escaped critical scrutiny.[36] I will not enter this debate here, but I wish to highlight a technical point: whereas in standard probability theory conditional expectations are only unique up to probability zero, here conditional expectations are unique. To explain what this means we need to introduce the notion of an algebra. An algebra $\mathfrak{A}$ is a subset of the set of all random variables that is closed under addition and multiplication (if two random variables are in $\mathfrak{A}$ then so is their sum and their product). Intuitively, $\mathfrak{A}$ can be thought of as the set of those random variables whose values are fully determined in a particular state of information. For example, the set of random variables whose values only depend on the first n digits of a binary sequence of length ν is an algebra. After having observed the first n digits, each random variable in this algebra has a determinate value. If we let X be a random variable, then the conditional expectation of X given the algebra $\mathfrak{A}$, denoted $\mathbb{E}_{\mathfrak{A}}X$, is the expected value of X when the values of all elements of $\mathfrak{A}$ are known: $\mathbb{E}_{\mathfrak{A}}X$ is the expectation of the true value of X in light of the information provided by $\mathfrak{A}$. If X is the indicator of a proposition B, then its conditional expectation given $\mathfrak{A}$ is the same as the conditional probability of B given $\mathfrak{A}$, $\mathbb{P}_{\mathfrak{A}}B$. Both $\mathbb{P}_{\mathfrak{A}}B$ and $\mathbb{E}_{\mathfrak{A}}X$ are themselves unique members of $\mathfrak{A}$. Otherwise, conditional expectations observe the same laws as their counterparts in ordinary probability theory.[37] In particular, $\mathbb{E}_{\mathfrak{A}}X$ and $\mathbb{P}_{\mathfrak{A}}B$ can be thought of as best estimates of, respectively, X and χ_B among all random variables in $\mathfrak{A}$.

Since algebras represent the evidence an agent may have, they lend themselves to model increasing evidence. If $\mathfrak{A}$ is a subset of another

[36] The advantages have to do with Bayesian conditioning; see, for example, Brian Skyrms, *Causal Necessity: A Pragmatic Investigation of the Necessity of Laws* (New Haven, CT: Yale University Press, 1980); and David Lewis, "A Subjectivist's Guide to Objective Chance," in R. C. Jeffrey, *Studies in Inductive Logic and Probability* (Berkeley: University of California Press, 1980), pp. 263–93. For critical reflections, see, for example, Timothy Williamson, "How Probable Is an Infinite Sequence of Heads?," *Analysis*, LXVII, 3 (July 2007): 173–80; Alan Hájek, "Is Strict Coherence Coherent?," *Dialectica*, LXVI, 3 (September 2012): 411–24; and Kenny Easwaran, "Regularity and Hyperreal Credences," *The Philosophical Review*, CXXIII, 1 (January 2014): 1–41.

[37] I explain this in appendix A.

algebra $\mathfrak{B}$, then every random variable in $\mathfrak{A}$ is also in $\mathfrak{B}$. Hence, if the values of all elements of $\mathfrak{B}$ are known, then everything in $\mathfrak{A}$ is known as well. As a result, an experimental setup that proceeds in stages and reveals increasing information can be represented in the following way. Let $\mathfrak{P}_n$ be the algebra of random variables whose values are known at stage n, and suppose that $\mathfrak{P}_n$ is a subset of $\mathfrak{P}_{n+1}$ for all n. The sequence of algebras $\mathfrak{P}_1, \ldots, \mathfrak{P}_\nu$ is called a *filtration*. If ν is an unlimited number, the filtration represents an experimental process that continues indefinitely. If we write the conditional expectation of a random variable X given $\mathfrak{P}_n$ as $\mathbb{E}_n X$, the sequence $\mathbb{E}_1 X, \ldots, \mathbb{E}_\nu X$ gives an agent's increasingly informed opinions about the random variable X.

In the next section we will study ways in which a sequence of conditional expectations or conditional probabilities may converge to the truth. This requires a notion of when a statement, A, holds almost everywhere (abbreviated "a.e."). For instance, A could mean that a sequence of random variables converges to a certain value. Intuitively, A holding a.e. says that A is true for "almost all" ω in Ω. In standard probability theory, "almost all" is captured by saying that the set of all ω for which A holds has probability one. In nonstandard probability, "almost all" means that A holds with probability infinitesimally close to one. Hence A fails to hold a.e. if the failure of A has *non-infinitesimal positive probability*. If we think of this as an event, then the set of all ω in Ω for which A is false has strictly positive (non-infinitesimal) probability.[38]

IV. TWO TYPES OF CONVERGENCE

Standard convergence has no unique analog in our nonstandard framework. I am going to explore two ways in which a sequence of random variables—in particular, a sequence of conditional expectations of one random variable—can be said to "converge." The first one is *almost sure convergence*; the second one requires a sequence to be of *limited fluctuation*.

To get an intuition for almost sure convergence, consider an unlimited sequence of real numbers $x_1, \ldots, x_\nu$. The sequence $x_1, \ldots, x_\nu$ is said to *converge to* x if, for all unlimited $n \leq \nu$, x_n is infinitesimally close to, and hence indistinguishable from, x, denoted $x_n \simeq x$. Convergence thus defined implies that for all non-infinitesimal positive ε there is a limited number m such that $|x_n - x| \leq \varepsilon$ for all $n \geq m$.

[38] Within Nelson's framework things are more nuanced because not every statement can be used in the formation of sets. This is not relevant for stating the results of the next section, but I explain the details in appendix A.

The nonstandard definition of convergence hence captures the idea of approximating a number.[39]

Almost sure convergence builds upon the idea of convergence but is a bit more contrived.[40] The relevant characterization involves a uniformity criterion: convergence of a sequence of random variables holds uniformly over all unlimited indices. (See appendix A for details.) The important thought to keep in mind is that a sequence of random variables converges almost surely to a given random variable if it approximates the latter with increasing accuracy up to infinitesimal probability. If, for instance, the sequence $\mathbb{E}_1 X, \dots, \mathbb{E}_\nu X$ converges to X almost surely, then $\mathbb{E}_n X$ gets closer to X for sufficiently large limited n a.e. In this case the value of X can be approximated by conditional expectations based on limited evidence. Otherwise, $\mathbb{E}_n X$ is bounded away from X for all large finite n by a (non-infinitesimal) amount with (non-infinitesimal) positive probability.

I show below that $\mathbb{E}_1 X, \dots, \mathbb{E}_\nu X$ converges to X almost surely whenever the underlying filtration provides no unlimited evidence, and vice versa. The idea of unlimited evidence can be made precise as follows. Let $\mathfrak{A}$ and $\mathfrak{B}$ be two algebras with $\mathfrak{A} \subseteq \mathfrak{B}$. We say that $X \in \mathfrak{B}$ is *almost an element of* $\mathfrak{A}$ if $\mathbb{E}_{\mathfrak{A}} X \simeq X$ a.e. Furthermore, $\mathfrak{A}$ and $\mathfrak{B}$ are *almost equal* if any $X \in \mathfrak{B}$ is almost an element of $\mathfrak{A}$.

Recall that $\mathbb{E}_{\mathfrak{A}} X$ can be thought of as the best estimate of X among all random variables in $\mathfrak{A}$. In case X is itself a member of $\mathfrak{A}$, then $\mathbb{E}_{\mathfrak{A}} X = X$. The concept of being almost an element extends this thought to cases in which X itself is not in $\mathfrak{A}$ but is nearly indistinguishable from $\mathbb{E}_{\mathfrak{A}} X$: the true value of X is effectively given by its best estimate in $\mathfrak{A}$, and so X can be regarded as a member of $\mathfrak{A}$ for all intents and purposes. If this is the case for all random variables in $\mathfrak{B}$, then $\mathfrak{A}$ and $\mathfrak{B}$ are nearly indistinguishable; knowledge of the values of random variables in $\mathfrak{B}$ does not give more information than knowledge of the values of random variables in $\mathfrak{A}$.

Now, let $\mathfrak{P}_1, \dots, \mathfrak{P}_\nu$ be a filtration. We say that $\mathfrak{P}_1, \dots, \mathfrak{P}_\nu$ *provides no unlimited information* about X if X is almost a member of $\mathfrak{P}_n$ for all unlimited $n \leq \nu$. Since $\mathbb{E}_n X$ is, in effect, equal to the true value of X for all unlimited $n \leq \nu$, unlimited members of the partition reveal nothing new about X. Along the same lines, we say that $\mathfrak{P}_1, \dots, \mathfrak{P}_\nu$ *provides no unlimited information* if it provides no unlimited information about any random variable in $\mathfrak{P}_\nu$.

[39] See Nelson, *Radically Elementary Probability Theory, op. cit.*, p. 20.

[40] Suppose that for the sequence of random variables $X_1, \dots, X_\nu$ we have $X_n \simeq X$ a.e. for all unlimited $n \leq \nu$. Since the a.e. qualification depends on n, the exceptional sequences may have strictly positive probability when taken together.

With these definitions in hand we can prove the following theorem.[41]

Theorem 1. Let $\mathfrak{P}_1, \ldots, \mathfrak{P}_\nu$ be a filtration.

(i) $\mathbb{E}_1 X, \ldots, \mathbb{E}_\nu X$ converges to X a.e. if and only if $\mathfrak{P}_1, \ldots, \mathfrak{P}_\nu$ provides no unlimited information about X.
(ii) Suppose $\mathfrak{P}_1, \ldots, \mathfrak{P}_\nu$ provides no unlimited information. Then for every random variable X, $\mathbb{E}_1 X, \ldots, \mathbb{E}_\nu X$ converges to $\mathbb{E}_\nu X$ a.e.

The first part of the theorem zooms in on a particular random variable X. It asserts that convergence of conditional expectations to X is fully characterized by there not being unlimited information about X. Unless that is the case, there is a non-infinitesimal positive probability that the sequence of conditional expectations fails to converge to the true value of X.

This result specializes seamlessly to propositions. The sequence of conditional probabilities $\mathbb{P}_1 A, \ldots, \mathbb{P}_\nu A$ converges to the indicator of A, χ_A, if and only if χ_A is almost a member of every unlimited algebra of the filtration: convergence to the truth requires that unlimited best estimates of A are almost indistinguishable from its truth value: $\mathbb{P}_n A \simeq \chi_A$ for all unlimited $n \leq \nu$ a.e.

The second part of the theorem speaks to the more restrictive scenario of no unlimited evidence for any random variable in $\mathfrak{P}_\nu$. This is sufficient for a.e. convergence of conditional expectations. Together with the first part of the theorem, this implies a.e. convergence to X for all X whenever $\mathfrak{P}_\nu$ is almost the same as the set of all random variables. The following theorem is a converse.[42]

Theorem 2. Suppose $\mathbb{E}_1 X, \ldots, \mathbb{E}_\nu X$ converges to $\mathbb{E}_\nu X$ for all random variables X. Then $\mathfrak{P}_1, \ldots, \mathfrak{P}_\nu$ provides no unlimited information.

The almost everywhere qualification in the definition of unlimited information is significant. If $\mathbb{E}_{\mathfrak{A}} X \simeq X$ a.e., then X is indistinguishable from its best estimate given $\mathfrak{A}$ *relative to a particular probability* $\mathbb{P}$. From a Bayesian point of view the concepts of almost-membership and almost-equality are not objective but are part of an agent's probability judgments. Two agents may disagree as to when a filtration provides unlimited information.

[41] For the proof, see appendix B. The random variable X in this and the following theorems is assumed to be integrable.
[42] It is also proved in appendix B.

This might lead one to seek a less subjective account of unlimited information. However, the most obvious candidate turns out to be inadequate. Suppose we were to say that a filtration $\mathfrak{P}_1, \ldots, \mathfrak{P}_\nu$ provides no unlimited information about X if $X \in \mathfrak{P}_n$ for all unlimited $n \leq \nu$.[43] Then the first part of Theorem 1 is true without there being any reference to the underlying probability function. Moreover, even if two agents disagree in their beliefs, they will agree on whether there is unlimited information in this alternative sense.

But the alternative is too restrictive. Suppose $X \in \mathfrak{P}_n$ for all unlimited n, and let n be the smallest natural number such that $X \in \mathfrak{P}_n$. Then n must be standard—otherwise X is not in $\mathfrak{P}_{n-1}$, where $n-1$ is unlimited, contrary to our assumption. Consequently, convergence to the truth already happens at a *limited* stage: there is a standard number n at which the true value of X is known. While this mode of arriving at the true value of X clearly falls under the scope of convergence to the truth, it fails to capture the idea that the true value of X might never be known but can be approximated as closely as one wishes. For this reason the more liberal—and more subjective—account of unlimited information is more adequate.

Let me now turn to the second analog for standard convergence. In standard analysis, a sequence of (standard) real numbers $x_1, x_2, \ldots$ *admits k ε-fluctuations* if there exist (standard) natural numbers $n_0 < n_1 < \cdots < n_k$ such that

$$|x_{n_0} - x_{n_1}| \geq \varepsilon, |x_{n_1} - x_{n_2}| \geq \varepsilon, \ldots, |x_{n_{k-1}} - x_{n_k}| \geq \varepsilon$$

(here, ε is not infinitesimal and k is standard). It can be shown that $x_1, x_2, \ldots$ converges[44] if and only if for all positive ε there is a k such that the sequence does not admit k ε-fluctuations. Within the standard framework, convergence and being of limited fluctuation are equivalent.

This is not true in the nonstandard framework.[45] Here, a sequence $x_1, \ldots, x_\nu$ is said to be of *limited fluctuation* if for all non-infinitesimal positive ε and all unlimited k, it does not admit k ε-fluctuations: there is no unlimited number of noticeable fluctuations. Any convergent sequence is also of limited fluctuation.[46] The converse is false. Consider

[43] An argument similar to the following one can be made by assuming that $\mathfrak{P}_n = \mathfrak{P}_\nu$ for all unlimited $n \leq \nu$.

[44] That is, converges to some value x in the standard sense of convergence.

[45] See Nelson, *Radically Elementary Probability Theory*, *op. cit.*, chapter 6.

[46] Recall that convergence of $x_1, \ldots, x_\nu$ implies that for all positive non-infinitesimal ε there is a limited m with $|x_n - x| \leq \varepsilon$ for all $n \geq m$. Thus the sequence can fluctuate by more than ε only before m is reached.

the sequence $x_n = 0$ for $n = 1, \dots, \nu - 1$ and $x_\nu = 1$. This sequence is of limited fluctuation but clearly fails to converge (not all unlimited elements of the sequence are infinitesimally close to the same number). Unlike a convergent sequence, a sequence that is of limited fluctuation can exhibit momentous shifts in the unlimited domain. Thus, in the nonstandard setting convergence and limited fluctuation come apart.

The next result, proved in appendix C, shows that the sequence of conditional expectations of a random variable is always of limited fluctuation.

Theorem 3. $\mathbb{E}_1 X, \dots, \mathbb{E}_\nu X$ is of limited fluctuation a.e.

The approach via limited fluctuations retains the flavor of the standard proof of the martingale convergence theorem, which relies on the idea that a martingale does not exhibit infinitely many upcrossings.[47]

Altogether, then, the following picture emerges. The best estimates of a random variable are almost always of limited fluctuation. Convergence to the truth is less permissive. If the true value of a random variable goes substantially beyond what is given by limited evidence, best estimates may fail to converge to the truth. In the next section I illustrate this in Cantor space.

V. CONVERGENCE TO THE TRUTH RECONSIDERED

Consider the nonstandard counterpart of Cantor space and define a sequence of random variables $X_1, \dots, X_\nu$ that keep track of binary outcomes: $X_n(\omega) = 1$ if the nth element of the sequence ω is a one and $X_n(\omega) = 0$ if it is a zero. Let $\mathfrak{P}_n$ be the algebra generated by the first n random variables $X_1, \dots, X_n$. Let $\mathbb{P}$ be the radically elementary analog of fair coin flips: $X_1, \dots, X_\nu$ are independent and $\mathbb{P}_m\{X_n = 1\} = \frac{1}{2}$ for all $m < n$. Consider events of the form $A_n = \{X_n = 1\}$ (the nth digit of the coin is a one). For each limited n, the sequence of conditional probabilities $\mathbb{P}_m A_n$, $m = 1, \dots, \nu$ converges to the indicator of A, χ_{A_n} a.e. But this is not true if n is unlimited. For then, $\mathbb{P}_m A_n = \frac{1}{2}$ for $m < n$ and $\mathbb{P}_m A_n = \chi_{A_n}$ a.e. for $m \geq n$; thus $\mathbb{P}_1 A_n, \dots, \mathbb{P}_\nu A_n$ does not converge to χ_{A_n} a.e. (In fact, the sequence fails to converge to the truth with probability one.) Nevertheless, the sequence of conditional probabilities clearly is of limited fluctuation a.e.

Things are similar in the more complex epistemic scenarios Belot has in mind. An atom of $\mathfrak{P}_n$ is a maximal subset of Ω on which all

[47] See Williams, *Probability with Martingales*, *op. cit.*, for a rigorous development of this idea.

random variables in $\mathfrak{P}_n$ are constant. Think of atoms as the most fine-grained pieces of evidence provided by $\mathfrak{P}_n$. A subset H of Ω is *dense* if for all limited n and all atoms A of $\mathfrak{P}_n$ we have $H \cap A \neq \emptyset$. We suppose, in addition, that the complement of H is also dense, so that H and its complement are indistinguishable by limited evidence.

To fix ideas, let H be the dense hypotheses that the "last" element of the binary sequence is one: $H = \{\omega : X_\nu(\omega) = 1\}$. Then the complement of H is the dense hypothesis that the "last" element is zero. There is no real loss in focusing on H; other examples of dense hypotheses with dense complements can be treated in a similar way.

How do the results of the previous section play into the charge of immodesty? According to Theorem 3, the sequence of conditional probabilities $\mathbb{P}_1 H, \ldots, \mathbb{P}_\nu H$ is almost surely of limited fluctuation—for H as well as for any other dense hypothesis—which means that a Bayesian does not change her mind noticeably an unlimited number of times. Yet this does not mean that the truth of H can be approximated by conditional probabilities: the limited fluctuation of $\mathbb{P}_1 H, \ldots, \mathbb{P}_\nu H$ is compatible with the presence of unlimited evidence. The unlimited piece of evidence could be observing the outcome of the "last" element, which is clearly beyond the reach of a finite agent. The radically elementary framework, as we have seen, is broad enough to allow for this. An agent's beliefs can be such that, with (non-infinitesimal) positive probability, $\mathbb{P}_1 H, \ldots, \mathbb{P}_\nu H$ fails to converge to the truth in the presence of unlimited evidence (as asserted in Theorem 1). Such an agent believes that due to her limited observational capacities she might never approximate the truth about H.

As a result, nonstandard probability theory can capture modesty about convergence to the truth. In the most extreme case, an agent may even believe that $\mathbb{P}_1 H, \ldots, \mathbb{P}_\nu H$ fails to converge to χ_H almost surely. But the theory does not mandate modesty. If the sequence $\mathbb{P}_1 H, \ldots, \mathbb{P}_\nu H$ does converge to χ_H almost surely, the agent believes that limited evidence is basically sufficient to approximate the truth about H. Whether or not this is plausible depends on the epistemic situation. Sometimes limited information might exhaust what there is to know about a dense hypothesis. It is perhaps not so easy to see this in the example of the "last" element of a binary sequence. But consider the hypothesis that the limiting relative frequency of ones is $\frac{1}{2}$. Both the hypothesis and its negation are dense, and it is plausible that their truth value can be approximated by limited evidence about relative frequencies.[48]

[48] Nelson, *Radically Elementary Probability Theory*, *op. cit.*, chapter 16, studies convergence and limited fluctuation in the context of relative frequencies.

I submit that convergence to the truth is treated with appropriate modesty within the nonstandard framework. One expects to converge to the truth whenever one thinks that increasing limited evidence is sufficient for approximating the truth. I also wish to point out that this does not contradict the standard theorem of convergence to the truth. The nonstandard theory used here operates within a conservative extension of classical probability theory. The classical martingale convergence theorem is still correct. Rather than refuting it, the results in the foregoing section are *refinements* of their standard counterpart.[49] The nonstandard model provides a fine-grained conceptual framework that allows us to articulate more precisely when convergence to the truth is expected.

Finally, I wish to suggest that these considerations steer us toward a Bayesian reply to Belot's criticism. The Bayesian advocate may maintain that in the standard setting it is unclear what exactly the worry is since it is difficult to distinguish infinite hypotheses that can be approximated by finite evidence from those that cannot. She can continue the defense by outlining a plausible way to state the worry in a nonstandard conservative extension of the classical theory and end by pointing out that, in this setting, the worry about immodesty dissolves.

VI. CONCLUSION

The nonstandard approach gives us an account of when a Bayesian agent considers an infinite hypothesis to be accessible by a finite inquiry. It is then, and only then, that the agent believes in a refined variant of convergence to the truth. Since the refined variant is fully compatible with its standard counterpart, it provides one way to spell out the finite content of convergence to the truth, thereby revealing a hidden nugget of modesty in Bayesian epistemology.

APPENDIX A. THE BASIC FRAMEWORK

I will use the following notation. If x and y are two numbers, they are infinitesimally close, denoted $x \simeq y$, if their difference $x-y$ is infinitesimal. Thus, x may be less than y, $x < y$, but only by an infinitesimal amount. If $x < y$ but it is not the case that $x \simeq y$, then x is significantly smaller than y, which we write as $x \ll y$.

Let $\mathbb{R}^\Omega$ be the set of all random variables on $(\Omega, \mathbb{P})$.[50] An *algebra* $\mathfrak{A}$ is a subset of $\mathbb{R}^\Omega$ that is closed under addition and multiplication (for

[49] In fact, one could follow Robinson's strategy of proving standard results via nonstandard proofs by taking either Theorem 1 or Theorem 3 as a starting point. See the appendix of Nelson, *Radically Elementary Probability Theory*, *op. cit.*, for more on this.

[50] This is the set of all functions from Ω to the set of standard and nonstandard reals. The functions can take on infinitesimal and unlimited values.

every $X, Y \in \mathfrak{A}$, $X + Y \in \mathfrak{A}$ and $XY \in \mathfrak{A}$). An *atom* A of $\mathfrak{A}$ is a maximal subset of Ω such that every element of $\mathfrak{A}$ is constant on A. The atoms of $\mathfrak{A}$ form a partition of Ω. Conversely, every partition gives rise to the algebra of random variables that are constant on its atoms. The conditional expectation of a random variable X given $\mathfrak{A}$ is the random variable defined by

$$\mathbb{E}_{\mathfrak{A}} X(\omega) = \frac{1}{\mathbb{P} A_\omega} \sum_{\eta \in A_\omega} X(\eta)\mathbb{P}(\eta),$$

where A_ω is the atom of $\mathfrak{A}$ that contains ω. Clearly, $\mathbb{E}_{\mathfrak{A}} X$ is constant on the atoms of $\mathfrak{A}$, and thus is itself an element of $\mathfrak{A}$. If the random variable in question is the indicator of some event B, then conditional expectation reduces to conditional probability:

$$\mathbb{P}_{\mathfrak{A}} B(\omega) = \frac{\mathbb{P}(B \cap A_\omega)}{\mathbb{P} A_\omega}.$$

Since $\mathbb{P} A_\omega > 0$ for each ω, both $\mathbb{E}_{\mathfrak{A}} X$ and $\mathbb{P}_{\mathfrak{A}} B$ exist and are unique.

In Nelson's probability theory, *external statements* involve the term 'standard', as in 'n is a standard natural number' or 'ν is a nonstandard natural number'. In contrast, *internal* statements can be identified with the statements of ordinary mathematics since they do not refer to the term 'standard'. Now, in Nelson's framework sets based on external statements are not axiomatized: only internal statements are allowed for defining sets. For example, the theory does not assert the existence of sets like the set of all standard natural numbers or the set of all nonstandard numbers, while it does assert the existence of sets like the set of all natural numbers less than some (standard or nonstandard) μ.[51]

Many interesting properties of a nonstandard probability space involve external statements (for example, a sequence of random variables being infinitesimally close to a given random variable at all nonstandard elements of the sequence). Since there is no guarantee in Nelson's theory that the corresponding external sets exist, they cannot be assigned probabilities. Nelson uses approximations to get around this problem. The following definition of a statement being true a.e., and the ensuing definition of almost sure convergence, are two instances of that idea.[52]

[51] This is known as the principle of illegal set formation; see Nelson, *Radically Elementary Probability Theory*, *op. cit.* It corresponds to the well-known distinction between internal and external sets in nonstandard analysis.

[52] Nelson, *Radically Elementary Probability Theory*, *op. cit.*, chapter 7.

Definition 4. The (internal or external) statement $A(\omega)$ holds a.e. if for every $\varepsilon \gg 0$ there is an event N such that $\mathbb{P}N \leq \varepsilon$ and $A(\omega)$ holds for all $\omega \in N^c$.

If $A(\omega)$ is an internal statement, then $A(\omega)$ holds a.e. is equivalent to $\mathbb{P}\{\omega : A(\omega)\} \simeq 1$. External statements are evaluated with approximations by sets that are part of the probability space.

It follows from the definition that if $A(\omega)$ does not hold a.e., then $\mathbb{P}\{\omega : \text{ not } A(\omega)\} \gg 0$ in case $A(\omega)$ is internal. If $A(\omega)$ is external and $A(\omega)$ does not hold a.e., then there is an $\varepsilon \gg 0$ such that for all events N, $\mathbb{P}N > \varepsilon$ whenever $A(\omega)$ is true for all $\omega \in N^c$. Thus, every set that contains all ω for which $A(\omega)$ is false has a non-infinitesimal positive probability. In both cases we may say that $A(\omega)$ holds with strictly (that is, non-infinitesimal) positive probability.

The nonstandard analog of L^1 spaces is of particular importance for proving our results. The L^1 norm $\|X\|$ of a random variable X is given by the expectation of the absolute value of X, $\|X\|_1 = \mathbb{E}|X|$. Let the truncation of a random variable X be given by $X^n = X\ \chi_{\{|X| \leq n\}}$. We say a random variable X *is* L^1 if $\|X - X^\mu\|_1 \simeq 0$ for all unlimited μ. This is equivalent to saying that the sum used for calculating $\mathbb{E}|X|$ converges.[53] Let us note some basic facts for future reference.[54]

Lemma 5.

(i) A random variable X is L^1 if and only if $\mathbb{E}|X| \ll \infty$ and for all M with $\mathbb{P}M \simeq 0$ we have $\mathbb{E}|X|\chi_M \simeq 0$.

(ii) If X and Y are L^1 and $X \simeq Y$ a.e., then $\mathbb{E}X \simeq \mathbb{E}Y$.

(iii) Let $\mathfrak{A}$ be an algebra and X a random variable. If X is L^1, then $\mathbb{E}_{\mathfrak{A}}X$ is L^1.

Let $T = \{1, \ldots, \nu\}$, where ν is nonstandard; T can be thought of as a nonstandard analog of $\mathbb{N}$. A *sequence of random variables* or a *discrete stochastic process* is a function $X : T \to \mathbb{R}^\Omega$. Such a process will be denoted by $X_1, \ldots, X_\nu$.

We are interested in whether a sequence $X_1, \ldots, X_\nu$ converges almost surely to a random variable X. Based on Definition 4 the following can be proved.[55]

Lemma 6. The sequence $X_1, \ldots, X_\nu$ converges to X a.e. if and only if

$$\max_{m \leq n \leq \nu} |X_n - X| \simeq 0 \quad \text{a.e.}$$

for all unlimited $m \leq \nu$.

[53] Nelson, *Radically Elementary Probability Theory, op. cit.*, chapter 8.

[54] Statement (i) is the radically elementary Radon-Nikodým theorem, and (ii) is the radically elementary Lebesgue theorem. Together with (iii) they are proved in Nelson, *Radically Elementary Probability Theory, op. cit.*, pp. 30–31.

[55] Nelson, *Radically Elementary Probability Theory, op. cit.*, Theorem 7.2.

Let $\mathfrak{P}_1, \ldots, \mathfrak{P}_\nu$ be a filtration. The sequence of random variables $X_1, \ldots, X_\nu$ is a $\mathfrak{P}$-process if $X_n \in \mathfrak{P}_n$ for $n = 1, \ldots, \nu$. If $X_1, \ldots, X_\nu$ is a $\mathfrak{P}$-process, then $X_1, \ldots, X_\nu$ is called a *submartingale* if, for all $m \leq \nu$, $\mathbb{E}_n X_m \geq X_n$ whenever $n \leq m$; a *supermartingale* if, for all $m \leq \nu$, $\mathbb{E}_n X_m \leq X_n$ whenever $n \leq m$; and a *martingale* if it is both a submartingale and a supermartingale.

Nelson proved the following result about martingales, which is the foundation for our results below.[56] The theorem is very powerful, since it connects the probabilistic notion of convergence a.e. to the analytic notion of convergence in L^1:

Definition 7. A sequence $X_1, \ldots, X_\nu$ *converges to a random variable X in L^1* if $\|X_n - X\|_1 \simeq 0$ for all unlimited $n \leq \nu$.

Theorem 8. Let $X_1, \ldots, X_\nu$ be a supermartingale or a submartingale that converges to X in L^1. Then it converges to X a.e. Furthermore, if X_ν is L^1, then $X_1, \ldots, X_\nu$ converges to X in L^1 if and only if it converges to X a.e.

APPENDIX B. CONVERGENCE OF CONDITIONAL EXPECTATIONS

Let $(\Omega, \mathbb{P})$ be a probability space, and let X be a random variable. Let $\mathfrak{P}_1, \ldots, \mathfrak{P}_\nu$ be a filtration. Then the sequence $\mathbb{E}_1 X, \ldots, \mathbb{E}_\nu X$ is a martingale since, for all $m \leq \nu$, $\mathbb{E}_n \mathbb{E}_m X = \mathbb{E}_n X$ whenever $n \leq m$. In order to converge, the sequence $\mathbb{E}_1 X, \ldots, \mathbb{E}_\nu X$ has to become approximately constant. This implies that X must not be affected by unlimited information. As mentioned in the main text, this can be made precise by using the following two concepts.

Definition 9. Let $\mathfrak{A}$ and $\mathfrak{B}$ be two algebras with $\mathfrak{A} \subseteq \mathfrak{B}$.

(i) $X \in \mathfrak{B}$ is *almost an element of* $\mathfrak{A}$ if $\mathbb{E}_{\mathfrak{A}} X \simeq X$ a.e.
(ii) $\mathfrak{A}$ and $\mathfrak{B}$ are *almost equal* if any $X \in \mathfrak{B}$ is almost an element of $\mathfrak{A}$.

The following theorem shows that the foregoing concepts are plausible. It says that random variables that are almost elements of an algebra $\mathfrak{A}$ are almost constant on the atoms of $\mathfrak{A}$. This corresponds to the fact that all elements of $\mathfrak{A}$ are constant on the atoms of $\mathfrak{A}$.

Theorem 10. Let X be L^1, and let $\mathfrak{A} \subseteq \mathfrak{B}$ be two algebras of random variables such that $X \in \mathfrak{B}$. If X is almost a member of $\mathfrak{A}$, then for each atom A of $\mathfrak{A}$ there exists a constant c such that $X \simeq c$ a.e. on A.

Proof. Let A be an atom of $\mathfrak{A}$. Since $\mathbb{E}_{\mathfrak{A}} X$ is an element of $\mathfrak{A}$, it is constant on every atom of $\mathfrak{A}$. Thus there is a constant c such that $\chi_A \mathbb{E}_{\mathfrak{A}} X = c$ on A. Since X is almost in $\mathfrak{A}$, $X \simeq \mathbb{E}_{\mathfrak{A}} X$ a.e. Thus $|X - \mathbb{E}_{\mathfrak{A}} X| \chi_A \leq |X - \mathbb{E}_{\mathfrak{A}} X| \simeq 0$ a.e. It follows that $\chi_A X \simeq \chi_A \mathbb{E}_n X$ a.e., and thus $X \simeq c$ a.e. on A. □

[56] Nelson, *Radically Elementary Probability Theory*, *op. cit.*, Theorem 11.3(i).

The next theorem is our first main result. Recall that a filtration $\mathfrak{P}_1, \ldots, \mathfrak{P}_\nu$ provides no unlimited information about X if X is almost a member of $\mathfrak{P}_n$ for every unlimited $n \leq \nu$. Also recall that $\mathfrak{P}_1, \ldots, \mathfrak{P}_\nu$ provides no unlimited information if $\mathfrak{P}_1, \ldots, \mathfrak{P}_\nu$ provides no unlimited information about any random variable in $\mathfrak{P}_\nu$, that is, if, for any unlimited $n \leq \nu$, $\mathfrak{P}_\nu$ and $\mathfrak{P}_n$ are almost equal.

Theorem 11. Let X be L^1, and let $\mathfrak{P}_1, \ldots, \mathfrak{P}_\nu$ be a filtration.

(i) $\mathbb{E}_1 X, \ldots, \mathbb{E}_\nu X$ converges to X a.e. if and only if $\mathfrak{P}_1, \ldots, \mathfrak{P}_\nu$ provides no unlimited information about X.
(ii) Suppose $\mathfrak{P}_1, \ldots, \mathfrak{P}_\nu$ provides no unlimited information. Then $\mathbb{E}_1 X, \ldots, \mathbb{E}_\nu X$ converges to $\mathbb{E}_\nu X$ a.e.

Proof. (i) Suppose $\mathbb{E}_1 X, \ldots, \mathbb{E}_\nu X$ converges to X a.e. Then by Lemma 6, $\max_{m \leq n \leq \nu} |\mathbb{E}_n X - X| \simeq 0$ a.e. for all unlimited $m \leq \nu$. This implies $\mathbb{E}_n X \simeq X$ a.e. for all unlimited $n \leq \nu$. Hence $\mathfrak{P}_1, \ldots, \mathfrak{P}_\nu$ provides no unlimited information about X.

For the other direction, we first establish the following fact (which will also be used below):

Lemma 12. If X and Y are L^1, then $|X - Y|$ is L^1.

Proof. By Lemma 5(i) it is enough to show that $\mathbb{E}|X - Y|$ is limited and that for all events M with $\mathbb{P}M \simeq 0$ we have $\mathbb{E}|X - Y|\chi_M \simeq 0$. Note that since both X and Y are L^1

$$\mathbb{E}|X - Y| \leq \mathbb{E}|X| + \mathbb{E}|Y| \ll \infty,$$

and so $\mathbb{E}|X - Y|$ is limited. If $\mathbb{P}M \simeq 0$, then

$$\mathbb{E}|X - Y|\chi_M \leq \mathbb{E}|X|\chi_M + \mathbb{E}|Y|\chi_M \simeq 0,$$

again because X and Y are L^1. □

Suppose now that $\mathfrak{P}_1, \ldots, \mathfrak{P}_\nu$ provides no unlimited information about X, that is, $\mathbb{E}_n X \simeq X$ a.e. for all unlimited $n \leq \nu$. Then $|\mathbb{E}_n X - X| \simeq 0$ a.e. for all unlimited $n \leq \nu$. Lemma 5(iii) implies that $\mathbb{E}_n X$ is L^1. Thus $|\mathbb{E}_n X - X|$ is L^1 by Lemma 12. It now follows from Lemma 5(ii) that $\mathbb{E}|\mathbb{E}_n X - X| \simeq \mathbb{E}0 = 0$. Hence for every unlimited $n \leq \nu$ we have $\|\mathbb{E}_n X - X\|_1 \simeq 0$, and so $\mathbb{E}_1 X, \ldots, \mathbb{E}_\nu X$ converges to X a.e. by Theorem 8.

(ii) If $\mathfrak{P}_n$ and $\mathfrak{P}_\nu$ are almost equal for every unlimited $n \leq \nu$, $\mathfrak{P}_n, \ldots, \mathfrak{P}_\nu$ provides no unlimited information about $\mathbb{E}_\nu X$. In addition, since X is L^1, $\mathbb{E}_\nu X$ is L^1 by Lemma 5(iii). Therefore, by part (i) of the theorem, $\mathbb{E}_1 X, \ldots, \mathbb{E}_\nu X$ converges to $\mathbb{E}_\nu X$ a.e. □

The next result is the converse of Theorem 11(ii) referred to in the main text.

Theorem 13. Suppose $\mathbb{E}_1 X, \dots, \mathbb{E}_\nu X$ converges to $\mathbb{E}_\nu X$ for all $X \in \mathbb{R}^\Omega$. Then $\mathfrak{P}_1, \dots, \mathfrak{P}_\nu$ provides no unlimited information.

Proof. Let X be a random variable. If $\mathbb{E}_1 X, \dots, \mathbb{E}_\nu X$ converges to $\mathbb{E}_\nu X$ a.e., then $\mathbb{E}_n X \simeq \mathbb{E}_\nu X$ a.e. for all unlimited $n \leq \nu$ by the same argument as in the proof of Theorem 11(ii). Since X is arbitrary, this implies that for all $Y \in \mathfrak{P}_\nu$, $\mathbb{E}_n Y \simeq Y$ a.e. for all unlimited $n \leq \nu$. Hence, any element of $\mathfrak{P}_\nu$ is almost an element of $\mathfrak{P}_n$ for any unlimited $n \leq \nu$. It follows that $\mathfrak{P}_1, \dots, \mathfrak{P}_\nu$ provides no unlimited information. □

APPENDIX C. FLUCTUATIONS OF CONDITIONAL EXPECTATIONS

The standard convergence to the truth theorem asserts convergence to the truth for all L^1 random variables. Theorem 11 restricts convergence to random variables that are not affected by unlimited information. By relaxing the nonstandard concept of convergence we can get a theorem which is closer to the classical result.

The relevant concept is that of a process that exhibits *limited fluctuations.*

Definition 14. (i) A sequence $x_1, x_2, \dots x_\nu$ *admits* k *ε-fluctuations* if there exist numbers $n_0 < n_1 < \dots < n_k$ such that

$$|x_{n_0} - x_{n_1}| \geq \varepsilon, |x_{n_1} - x_{n_2}| \geq \varepsilon, \dots, |x_{n_{k-1}} - x_{n_k}| \geq \varepsilon.$$

(ii) A sequence $x_1, \dots, x_\nu$ is said to be of *limited fluctuation* if for all $\varepsilon \gg 0$ and all unlimited k, it does not admit k ε-fluctuations.

As mentioned in the main text, any convergent sequence is of limited fluctuation but not vice versa.

The proof that conditional expectations are of limited fluctuation is based on the following result.[57]

Lemma 15. Let $X_1, \dots, X_\nu$ be a supermartingale or a submartingale with $\|X_\nu - X_1\|_1 \ll \infty$. Then $X_1, \dots, X_\nu$ is of limited fluctuation a.e. for all $n \leq \nu$.

Notice that the lemma only requires $\|X_\nu - X_1\|_1$ to be limited, which is weaker than assuming that $X_\nu - X_1$ is L^1 (see Lemma 5(i)).

Theorem 16. Let X be L^1. Then $\mathbb{E}_1 X, \dots, \mathbb{E}_\nu X$ is of limited fluctuation a.e.

Proof. If X is L^1, then both $\mathbb{E}_1 X$ and $\mathbb{E}_\nu X$ are L^1 by Lemma 5(iii). It follows from Lemma 12 that $|\mathbb{E}_\nu X - \mathbb{E}_1 X|$ is L^1. Lemma 5(i) now implies that $\|\mathbb{E}_\nu X - \mathbb{E}_1 X\|_1$ is limited. By Lemma 15, $\mathbb{E}_1 X, \dots, \mathbb{E}_\nu X$ is of limited fluctuation a.e. □

SIMON M. HUTTEGGER

University of California, Irvine

[57] Nelson, *Radically Elementary Probability Theory*, *op. cit.*, Theorem 12.3.