The Review of Symbolic Logic
Volume 17, Number 2, June 2024

UNPRINCIPLED

GORDON BELOT
Department of Philosophy, University of Michigan

Abstract. It is widely thought that chance should be understood in reductionist terms: claims about chance should be understood as claims that certain patterns of events are instantiated. There are many possible reductionist theories of chance, differing as to which possible pattern of events they take to be chance-making. It is also widely taken to be a norm of rationality that credence should defer to chance: special cases aside, rationality requires that one's credence function, when conditionalized on the chance-making facts, should coincide with the objective chance function. It is a shortcoming of a theory of chance if it implies that this norm of rationality is unsatisfiable. The primary goal of this paper is to show, on the basis of considerations concerning computability and inductive learning, that this shortcoming is more common than one would have hoped.

§1. Introduction. This paper is concerned with the rational relation between credence (degrees of belief) and chance (objective probability) and with reductionist theories of chance (according to which facts about chance supervene on non-chance-y facts). There is, of course, already a large and flourishing literature on these topics. Its dual centrepieces are: Lewis's Best System account of chance, on which laws of chance are scientifically ideal summaries of patterns of events; and his Principal Principle, which says that the degrees of belief of rational agents must defer, in a certain sense, to facts about chance.[1] In giving us a relation between chance and rational credence, the Principal Principle gives us interesting constraints on both relata: Lewis thought of the Principal Principle as telling us something about what it means to be rational and at the same time as encapsulating important truths about the concept of chance.[2] Relative to a theory of chance, the Principal Principle picks out those prior credence functions that are rationally permitted. At the same time it gives us possible grounds for criticizing theories of chance. I suggest the following example: if we are thinking of theories of chance as something like summaries meant to be used by beings like us, there is something pathological about a theory of chance relative to which *no* priors count as rationally permitted.

Here I aim to make a contribution that is orthogonal to the directions pursued in most of the literature, making elementary use of some tools from the theory of computation and from the theory of inductive learning to clarify questions about which theories of chance are pathological in the sense just mentioned. It will emerge

Received: February 4, 2022.
2020 *Mathematics Subject Classification*: Primary 03B48, Secondary 03A10, 60A05, 62A01.
Key words and phrases: chance, credence, Principal Principle, computable Bayesianism.

[1] See [45, 46].

[2] In addition to the papers cited in the preceding note, see [7, Letters 659 and 660].

doi:10.1017/S1755020323000151

below that it is much more difficult than one might expect to find reductionist theories of chance that do not face difficulties of this sort.

Very roughly speaking: on Lewis's Best System account, the chance laws at a world are just good scientific summaries of the patterns of events at that world; and his Principal Principle says that if you are rational, then to the extent that you are confident in a chance hypothesis it should guide your estimates of probability. So, taken together, the package seems to imply that in order to be rational, you need to be a certain sort of universal learner: whatever the chance laws at your world are, if you are a good scientist and you see enough data, you should become confident in something like those laws and so you should eventually mimic those laws in your estimates of probability. But in many contexts we know that there can be no universal learner.[3]

Almost all of the discussion below will take place in the special setting in which possible worlds can be modelled by infinite binary sequences (simple enough to be tractable, complex enough to provide a picture of inquiry in which typical questions are never definitively settled by a finite number of observations). The framework we will be using is introduced in Sections 1.1–1.3.

Section 2 is concerned with reductionist theories of chance. For present purposes, such a theory of chance for a given space of possible worlds is a map (satisfying certain mild technical conditions) that assigns to (some) worlds of that space a probability measure on that space—for a world assigned such a probability measure, the measure plays the familiar role of chance laws (telling us the chance for various events to happen, unconditionally or conditionally on other events having happened). This section also introduces five very simple paradigm examples of reductionist theories of chance for our chosen set of worlds. These will be used throughout as illustrations.

Section 3 is concerned with a weak version of the Principal Principle, the Chance–Credence Principle (CCP). We will see that it is straightforward to characterize the priors that satisfy CCP with respect to any given theory of chance—and that some of our paradigm examples do not admit any priors that satisfy this principle.

There would be something unsatisfactory about an account of rationality that required rational agents to be able to perform some absolutely impossible task—squaring the circle, say. In the same way, if we are interested in the rationality of finite agents—e.g., human beings and the machines that they build—then there is something unsatisfactory about any account of rationality that requires agents to be able to perform tasks that are impossible for finite agents—solving the Halting Problem, say.[4] We will see in Section 4 that *each of our paradigm examples of theories of chance is unsatisfactory in this way.* So there is a challenge here for those attracted to reductionist theories of chance and to the Principal Principle: to develop and to defend examples of reductionist theories of chance that are CCP-satisfiable by computable priors (or to explain how norms of rationality can require agents to perform tasks that are impossible for computable agents).

Motivated in part by the picture sketched above of the relationship between the Lewisian package (Best System account of chance + the Principal Principle) and learning, Section 5 considers weaker analogs of CCP (and versions of them that require priors to be computable). The upshot is that if CCP is replaced by a weak enough principle along these lines, then some more reductionist theories of chance are

[3] For early results of this kind, see [54, 55]. For elaborations, see [36, 41].

[4] For a dissenting perspective, see [19, sec. 9.8].

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press

consistent with Bayesian rationality. But many obstacles remain: for those interested in the combination of a reductionist theory of chance and a Principal Principle-style account of the relation between chance and credence there is interesting work to be done.

Appendix A consists of some remarks concerning the relation between the Principal Principle, some of its principal rivals, and CCP. Appendix B plays some Whac-A-Mole against the complaint that at various points, my discussion shows that I am being stiff-necked, closed-minded, naive, or unimaginative in my handling of conditional probabilities and null sets.

Ground-rules: all probability measures are real-valued and countably additive; the Axiom of Choice is taken for granted; and, for the most part, the notational conventions of [18] are followed. Non-reductionist theories of chance are neglected throughout—not because they are somehow immune from the problems encountered below, but because their treatment would require an extension of the present framework.

1.1. Main characters.

a. The set of bits, $\mathsf{2} := \{0, 1\}$ (note the sans serif font—$\mathsf{2}$ denotes a set, 2 a number).
b. For each $n \in \mathbb{N}$, $\mathsf{2}^n$ is the set of n-bit strings (we include the case of the empty string of zero bits).
c. The set of all binary strings: $\mathsf{2}^{<\omega} := \bigcup_{n=0}^{\infty} \mathsf{2}^n$. If σ and τ are binary strings we write $\sigma\tau$ for the string that results from concatenating σ and τ (in that order). We call a binary string *evenly split* if it contains the same number of 0's as it does 1's.
d. Cantor space: the set $\mathsf{2}^\omega$ of all infinite binary sequences.[5] For $S \in \mathsf{2}^\omega$, we write $S(k)$ for the kth bit of S and $S{\restriction}k$ for the string formed by concatenating the first k bits of S. We call a binary sequence *evenly split* if the limiting relative frequency of 0's in it is .5.
e. For any binary string τ, we denote by $[\![\tau]\!]$ the set of $S \in \mathsf{2}^\omega$ whose initial bits are given by τ. We call the $[\![\tau]\!]$ *basic subsets* of $\mathsf{2}^\omega$.[6] We use $\mathcal{B}$ to denote the σ-algebra of subsets of $\mathsf{2}^\omega$ generated by the family of basic sets: $\mathcal{B}$ is the smallest family of subsets of $\mathsf{2}^\omega$ that includes all the basic sets and which is closed under taking complements, countable unions, and countable intersections. We call the members of $\mathcal{B}$ the *Borel subsets* of $\mathsf{2}^\omega$.[7]
f. The space $\mathcal{P}$ of Borel probability measures on $\mathsf{2}^\omega$: each $\mu \in \mathcal{P}$ is a map from $\mathcal{B}$ to $[0, 1]$ such that: (a) $\mu(\mathsf{2}^\omega) = 1$; and (b) μ is countably additive (i.e., $\mu(\bigcup U_k) = \sum \mu(U_k)$, for any countable family U_1, U_2, ... of pairwise disjoint Borel sets). We call a map m from the family of basic sets to $[0, 1]$ a *probability function* if: (i) $m(\mathsf{2}^\omega) = 1$; and (ii) for any binary string σ,

[5] By *sequence* we always mean *one-way infinite sequence indexed by the positive natural numbers.*

[6] It is often convenient to take $\mathsf{2}^\omega$ to be equipped with the topology generated by the basic sets—this coincides with the product topology (induced by thinking of $\mathsf{2}^\omega$ as a product of a countable number of copies of $\mathsf{2}$, each equipped with the discrete topology). In this topology, each basic set is compact as well as open: $\mathsf{2}^\omega$ itself is compact (being the product of compact spaces); and each basic set is closed (if σ is a k-bit string, then the complement of $[\![\sigma]\!]$ is open, being the union of the basic sets determined by the k-bit strings other than σ).

[7] Note that since there are only countably many basic sets, each open set of the product topology (see footnote 6) is in $\mathcal{B}$—so $\mathcal{B}$ is the σ-algebra generated by this topology.

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press

$m(\llbracket\sigma\rrbracket) = m(\llbracket\sigma 0\rrbracket) + m(\llbracket\sigma 1\rrbracket)$.[8] Any Borel probability measure determines via restriction a probability function. And the restriction in fact determines the probability measure: probability measures that agree on all basic sets agree on all Borel sets.[9] Further, every probability function can be extended to a Borel probability measure.[10]

Example 1.1 (Delta-functions). *For any $S \in 2^\omega$, the* delta-function concentrated on S *is the Borel probability measure δ_S that for any binary string τ, assigns the basic set $\llbracket\tau\rrbracket$ probability one if S begins with τ, and otherwise assigns it probability zero.*

Example 1.2 (Bernoulli measures). *For any $r \in (0,1)$, the* Bernoulli measure with parameter r *is the Borel probability measure ν_r such that for any binary string τ, $\nu_r(\llbracket\tau\rrbracket) = r^k(1-r)^\ell$, where* k *is the number of* 0*'s in τ and ℓ is the number of* 1*'s in τ. So according to ν_r, bits are selected by independent tosses of a coin with bias* r *in favour of* 0*'s. We call $\nu_{.5}$ the* fair coin *measure.*

Example 1.3 (Hybrids). *At various points it will be helpful to have recourse to measures that initially behave like delta-functions before switching to the behaviour of the fair coin measure. For any binary string σ we define the measure $\nu^\sigma_{.5}$ as follows: $\nu^\sigma_{.5}(\llbracket\tau\rrbracket)$ is* 1 *if τ is an initial segment of σ, is $1/2^k$ if τ is a* k*-bit extension of σ, and is* 0 *otherwise.*

Since there is a natural bijection between binary strings and the basic sets that they determine, when we are focusing on the action of probability measures on basic sets we will often simplify notation by writing $\mu(\tau)$ for $\mu(\llbracket\tau\rrbracket)$ etc. Similarly, if b is a bit, and τ is a binary string, the expression $\mu(b \mid \tau)$ stands for $\mu(\llbracket\tau b\rrbracket \mid \llbracket\tau\rrbracket)$, the probability that μ gives for the next bit to be b, conditional on the sequence beginning with the bits of τ.

1.2. Computable numbers and measures. A binary sequence $S \in 2^\omega$ is *computable* if there is a Turing machine that on input of a natural number k gives as output the first k bits of S.

We say that a binary string $\sigma = b_1 b_2 \dots b_n$ is a *k-bit approximation* to a number x in the unit interval if

$$\left| x - \sum_{j=1}^{n} \frac{b_j}{2^j} \right| \leq \frac{1}{2^k}.$$

A real number x in the unit interval is *computable* if there is a Turing machine that on input of a natural number k gives as output a k-bit approximation to x.

So, intuitively, a real number (or a sequence) is computable if and only if it can be computably approximated in a controlled way. Similarly, a real-valued function

[8] An argument by induction shows that for any $k > 1$, any probability function satisfies a generalized form of condition (ii), with the sum ranging over all k-bit extensions of σ. From this it follows that any probability function is finitely additive as a function on basic sets. And in this case finite additivity implies countable (sub)additivity: recall from footnote 6 that in the product topology each basic set is both open and compact—so if a basic set arises as the union of a countable family of basic sets, then it also arises as the union of a finite subfamily.

[9] See, e.g., [43, lemma 1.42].

[10] This follows via the Carathéodory Extension Theorem (since the family of empty-or-basic subsets of 2^ω is a semi-ring and the extension of m to this family, via the stipulation $m(\varnothing) = 0$, is finitely additive and countably subadditive). See, e.g., [43, theorem 1.53].

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press

on a computably enumerable set is computable if there is a single machine that can approximate its value on any argument in a controlled way (this is stronger than just requiring that each of its values be computable). Specializing this to the case of probability measures—which, as we have seen, are determined by their restrictions to the basic subsets of 2^ω—it is natural to define a Borel probability measure μ on 2^ω to be *computable* if there is a Turing machine that on input of a binary string σ and a natural number k gives as output a k-bit approximation to $\mu([\![\sigma]\!])$.

Note that a delta-function measure δ_S is computable if and only if the sequence S that it is concentrated on is and that a Bernoulli measure ν_r is computable if and only if its parameter r is. Note, further, that since there are only countably many Turing machines, there can be only countably many computable sequences, computable real numbers, and computable probability measures.

1.3. Worlds and priors. We are going to use binary sequences to model certain simple possible worlds: time at these worlds is discrete, has a first instant, and is infinite towards the future; the state of one of these worlds at a time is encodable by a single bit. These worlds are too simple to be realistic, but allow us to study problems about credence and chance in a controlled setting (and of course they provide models of subsystems of more complex worlds).

We are going to work in a Bayesian framework in which agents satisfy three conditions.

PROBABILISM: Each agent begins life in a credal state encoded in a probability measure μ (the agent's *prior*), defined over some salient set of possible worlds.
EVIDENTIAL REGULARITY: Each agent's prior must assign non-zero weight to any possible evidence E that the agent might acquire.
CONDITIONALIZATION: The credal state of an agent with prior μ and total evidence E is encoded by $\mu(\cdot \mid E)$.

In our setting, the natural explication of *possible state of evidence* is: a finite initial segment of a world. So we will call a probability measure $\mu \in \mathcal{P}$ a *prior* if it satisfies $\mu(\tau) > 0$ for every binary string τ.

§2. Chance.

Our focus here is on reductionist theories of chance. The spirit behind these (broadly Humean) approaches can be summarized as follows:

> Philosophers are deeply divided on the status of modal truths in general and nomic truths in particular—truths related to physical possibility and necessity. Besides chance, prominent nomic phenomena include causation, counterfactuals, dispositions, and laws of nature. One approach to the nomic, going back to Hume and further to the medieval nominalists, holds that nomic facts about what *could* or *would* or *must* are always reducible to facts about what *is*. For the most part, Humeans do not deny the reality of nomic phenomena; they agree that there are laws of nature, chances, dispositions, etc. But they maintain that these things are derivative, determined by more fundamental, non-modal elements of reality. Thus Humeans might identify laws of nature with regularities in the history of physical events, and chances with relative frequencies. This ensures

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press

> that whenever two possible worlds agree in non-nomic aspects, they also agree with respect to laws and chance.[11]

> So what is a Humean reductionist theory of chance? The core idea is that: the chance facts are reducible to certain facts about *patterns* to be found in the whole panoply of events that happen in the history of our universe. What kinds of patterns? Broadly speaking, and unsurprisingly, they are stable and *stochastic-looking* or *random-looking* patterns.[12]

In our setting, a complete history of a world is a binary sequence—a good candidate for an object that can be described in non-modal terms but which might or might not exhibit the sorts of patterns that could ground talk about chance. A probability measure ν on 2^ω can be thought of in the following terms: $\nu(0)$ gives a probability for the initial state of the world to be 0 and $\nu(1)$ gives the probability for it to be 1; and given any binary string τ, $\nu(0 \mid \tau)$ and $\nu(1 \mid \tau)$ give the probability for the next bit to be 0 or 1, if τ gives the history so far. So we can picture ν as a book of instructions that tells Nature how to construct worlds: first, flip a coin of such and such bias to generate the first state; at any later time, flip a coin of *this* bias to generate the next state, if the history so far is *that*. More generally, for any (Borel) subset A of 2^ω, and any binary string τ, $\nu(A \mid \tau)$ gives the probability for A to obtain, given that τ encodes an initial segment of the history of the world.[13]

So in our setting, a reductionist theory of chance is an assignment of probability measures to (some) worlds. It is convenient to build a couple of weak technical conditions into the official definition.

Definition 2.1. *A* theory of chance *is a (possibly merely partially defined) map* $L : 2^\omega \to \mathcal{P}$ *such that for any* λ *in the image of* L *we have (i) that* $L^{-1}(\lambda)$ *is a Borel subset of* 2^ω *and (ii) that* $\lambda(L^{-1}(\lambda)) > 0$.

Interpretation: if L is a theory of chance and $L(S) = \lambda$ for some $S \in 2^\omega$ and $\lambda \in \mathcal{P}$, then we think of L as saying that λ encodes the facts about chance at S. We refer to any $\lambda \in \mathcal{P}$ in the image of L as a *law of chance* of L (or a *chance law* of L). If λ is a law of chance of L, then we refer to the proposition (= set of worlds) $\Lambda := L^{-1}(\lambda)$ as the *chance hypothesis* determined by λ. If, on the other hand, $L(S)$ is undefined, then L says that there are no facts about chance at S. We will refer to the set of worlds Λ_* at which L is undefined as the proposition that the world is *lawless* (it is the negation of the disjunction of all the chance hypotheses of the theory).

Clause (i) in Definition 2.1 requires that chance hypotheses be the sort of things that probabilities can be assigned to.[14] Clause (ii) rules out a certain sort of aberrant behaviour: a law of chance that says that there is zero chance that it is the law of chance.[15]

[11] See [60, p. 425].

[12] See [31, p. 2].

[13] So such a measure encodes the complete set of what Lewis [45, p. 97] calls “history to chance conditionals,” and hence gives what he calls “the complete theory of chance” for a world.

[14] This is a weak measurability requirement. We will consider a stronger one in Appendix B.4.

[15] This clause will be invoked below only in Section 5.1 and in Appendix A.2.

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press

In practice, we are most interested in theories of chance inspired by the heuristic considerations that underlie various classic philosophical programs for understanding chance reductively. Most prominent among these are: *frequentist approaches,* in which the chance of any event type is identified with the relative frequency with which that type of event occurs in some relevant reference class of events; and *Best System approaches* in which the facts about chance at a world are provided by the (in general stochastic) hypothesis that provides the best summary of facts at that world, according to ordinary scientific standards (very roughly speaking—the best hypothesis is the one that in some relevant sense ideally balances simplicity, logical strength, and fit with the data).[16] It will be helpful below to also have in mind a third (not especially popular) approach, *super-determinism,* on which the only laws of chance are deterministic (i.e., are encoded in delta-functions).

Consider the fair coin measure, $\nu_{.5}$, on 2^ω. What does it mean to think of it as the law of chance at a world $S \in 2^\omega$? Well, for starters, $\nu_{.5}$ tells you that the probability of 0 occurring at any time is .5—and that this probability is independent of what happens at other times. In order to be acceptable to frequentists, a theory of chance must assign the fair coin measure to all and only worlds that are evenly split between 0's and 1's. Advocates of the Best System approach may recognize some exceptions to this rule: perhaps (in the setting of finite worlds) it sometimes makes sense to take the fair coin measure to give the law of chance at a world at which the relative frequency of 0's is given by a number r that is close to .5, when r itself is sufficiently far from being simple.[17] And perhaps some evenly split sequences should be assigned a law of chance other than the fair coin measure—e.g., one might well want to assign a particularly simple $S \in 2^\omega$ the delta-function measure δ_S concentrated on S as its law of chance, whether or not S is evenly split.[18]

It will be helpful to have in mind a few theories of chance that can be thought of as very simple-minded implementations of super-determinism, frequentism, and the Best System approach.

Example 2.2 (Super-Determinism). *Each $S \in 2^\omega$ is assigned as its chance law δ_S, the delta-function measure concentrated on S.*

Example 2.3 (Computable Super-Determinism). *Each computable $S \in 2^\omega$ is assigned δ_S as its chance law. All other worlds are lawless.*

Example 2.4 (Basic Frequentism). *If 0's have relative frequency* r *in $S \in 2^\omega$, then the law of chance for S is the Bernoulli measure ν_r. If this relative frequency is not defined, S is lawless.*

Example 2.5 (Computable Frequentism). *If 0's have relative frequency* r *in $S \in 2^\omega$ and* r *is computable, then the law of chance for S is the Bernoulli measure ν_r. If the relative frequency of 0's in S is uncomputable or undefined, then S is lawless.*

There may seem to be little reason to opt for Computable Frequentism over Basic Frequentism: it doesn't seem to follow from the thought that probabilities are relative

[16] Schwarz [60, p. 424] again provides a helpful encapsulation: "chances are identified with probabilities in ideal physical theories whose aim is to provide a kind of summary statistic of actual outcomes; getting close to the frequencies is one virtue of probabilistic theories, but it trades off against other virtues such as comprehensiveness and simplicity."

[17] For this suggestion, see [46, sec. 4].

[18] On this point, see again [46, sec. 4].

frequencies that those frequencies must be given by computable numbers. But a restriction to computable chance laws is entirely natural on the Best System approach. If we are thinking of candidate chance laws for a world as something like putative scientific summaries of that world, with the actual chance law (if there is one) being the optimal such summary (as judged by ordinary scientific standards), then it would appear to be essential that candidate chance laws be finitarily specifiable. After all, what would it mean to compare the overall virtues of two infinite texts by the ordinary standards of scientific—or literary—criticism? To directly specify a binary sequence (or, equivalently, a real number in the unit interval) would be to specify the bits that make it up, one by one: an infinitary task. Turing launched modern theoretical computer science with the assertion that "The 'computable' numbers may be described briefly as the real numbers whose expressions as a decimal are calculable by finite means."[19] Turing's proposal is that the finitarily specifiable real numbers are the Turing-computable real numbers. A probability measure on 2^ω can be thought of as a certain sort of real-valued function on the countably infinite space of finite binary strings. Direct specification of such an object would be a doubly infinitary task, involving the specification of an infinite number of real numbers. But some such objects are finitarily specifiable: the Turing-computable measures (and the success of the Church–Turing thesis gives us reason to think that these are *all* of the finitarily specifiable such objects). So it is natural for those interested in Best System approaches to restrict their attention to theories in which all chance laws are computable.[20]

Example 2.6 (A Toy Best System Theory). *Any computable $S \in 2^\omega$ is assigned the delta-function measure concentrated on S. Any incomputable S in which 0's have computable relative frequency* r *is assigned the Bernoulli measure ν_r. All other S are lawless.*

Remark 2.7. *In a more plausible implementation of the best-system approach, it would be natural to further require that any sequence assigned a given computable probability measure as its law of chance must be algorithmically random relative to that measure.*[21] *It would also be natural to include further families of laws of chance: generalized Bernoulli measures (for which the bias of the coin used to determine the* n*th bit is allowed to depend on* n*); Markov chains (for which the bias of the coin used to determine the state at a time depends on the states at some fixed number of immediately preceding times); and so on.*

Definition 2.8 (Proper and Improper Theories). *Let L be a theory of chance. We call a law of chance λ of L with chance hypothesis Λ* proper *if $\lambda(\Lambda) = 1$, otherwise we call λ* improper. *We call* L proper *if all of its laws of chance are proper; otherwise we call* L improper.

Remark 2.9. *Each of the theories of chance mentioned above is proper. This follows from three facts. (a) Each delta-function δ_S assigns measure one to $\{S\}$. (b) Each Bernoulli measure ν_r assigns measure one to the set of sequences in which 0's have relative frequency* r *(the strong law of large numbers). (c) Each Bernoulli measure ν_r assigns measure zero to each countable set.*

[19] This is the opening sentence of [62].

[20] So here I am denying that numbers that can in a sense be defined but which cannot be computably approximated in a controlled way can play any role in science—maybe that is a mistake. For the most famous examples of such numbers, see, e.g., [4].

[21] See [11] for further discussion of this point.

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press

Example 2.10 (An Improper Theory of Chance). *As in Example 1.3, let $v^0_{.5}$ be the probability measure that is certain that the first bit will be* 0 *but thinks that all subsequent bits are sampled by tossing a fair coin. Consider the theory of chance that assigns the fair coin measure to all sequences that begin with* 1 *and assigns $v^0_{.5}$ all sequences that begin with* 0. *In this theory $v^0_{.5}$ is a proper chance law but the fair coin measure in an improper one.*

§3. Chance and credence. It has seemed to many that rationality requires there to be a certain sort of relation between chance and credence—that in certain situations, if you are rational, then your credence in an event must coincide with the chance of that event.

> Knowing only that the chance of drawing a red ball from an urn is 0.95, everyone agrees, in accordance with the law of likelihood, that a guess of 'red' about some trial is much better supported than one of 'not-red.' But nearly everyone will go further, and agree that 0.95 is a good measure of the degree to which 'red' is supported by the limited data.[22]

> [T]he chancemaking pattern in the arrangement of qualities must be something that would, if known, correspondingly constrain rational credence. Whatever makes it true that the chance of decay is 50% must also, if known, make it rational to believe to degree 50% that decay will occur.[23]

I will codify the constraint on priors suggested as follows.

Definition 3.1 (CCP). *Let L be a theory of chance. We say that a prior μ satisfies the* Chance–Credence Principle (CCP) *for L if: for any chance law λ of L with chance hypothesis Λ and for any Borel subset A of 2^ω we have*

$$\mu(A \mid \Lambda) = \lambda(A).$$

If there exists such a prior, we say that L is CCP-satisfiable; *otherwise we say that it is* CCP-unsatisfiable.

Remark 3.2 (Nomenclature). *The name* Miller's Principle *is often given to principles in this neighbourhood.*[24] *With more justice (and less snark) it might perhaps be called* Hacking's Principle, *in light of Hacking's somewhat earlier enunciation of a principle of this kind.*[25] *But, really, it seems hopeless to try to identify the originator of an idea like this: as Hacking notes in introducing his version of the principle, it "seems to be so universally accepted that it is hardly ever stated."*[26]

[22] See [27, p. 136].

[23] See [46, p. 478].

[24] Because Miller [49] argued that a principle along these lines leads to contradiction.

[25] See [27, esp. pp. 135 ff. and 193].

[26] See [27, p. 135]. Hacking himself offers a reading on which Bayes was at least implicitly committed to something like it—and also warns that although "many people readily grant [this principle], it is by no means certainly correct" (p. 193).

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press

CCP can be thought of as a weak form of the Principal Principle (see Appendix A). What does CCP require?

Definition 3.3. *Let μ be a prior and let λ be a law of chance of a theory of chance L. We say that μ* contains λ as a nugget of truth *if we can write $\mu = \nu + c \cdot \lambda$ with $c > 0$ and ν a (possibly trivial) measure that considers the chance hypothesis Λ of λ a null set.*[27]

Proposition 3.4. *Let L be a theory of chance and let μ be a prior on 2^ω. Then μ satisfies CCP for L if and only if L has a countable number of chance laws, each of which is proper and contained in μ as a nugget of truth.*

Proof. Suppose, first, that the chance laws of L are $\lambda_1, \lambda_2, \ldots$ (with chance hypotheses $\Lambda_1, \Lambda_2, \ldots$), that each λ_k is proper, and that $\mu = \nu + \sum c_k \cdot \lambda_k$ with each $c_k > 0$ and ν a measure that considers each Λ_k a null set. We show that μ satisfies CCP for L. Fix $j = 1, 2, 3, \ldots$ and a Borel set A. We have

$$\begin{aligned} \mu(A \mid \Lambda_j) &= \frac{\mu(A \,\&\, \Lambda_j)}{\mu(\Lambda_j)} \\ &= \frac{\nu(A \,\&\, \Lambda_j) + \sum c_k \cdot \lambda_k(A \,\&\, \Lambda_j)}{\nu(\Lambda_j) + \sum c_k \cdot \lambda_k(\Lambda_j)} \\ &= \frac{c_j \cdot \lambda_j(A \,\&\, \Lambda_j)}{c_j \cdot \lambda_j(\Lambda_j)} \\ &= \lambda_j(A), \end{aligned}$$

where the first equality holds by definition; the second follows from the expansion for μ given above; the third follows from the fact that ν and the λ_k other than λ_j assign measure zero to Λ_j; and the fourth follows via the propriety of λ_j.

Suppose, on the other hand, that μ satisfies CCP for L. Let λ be one of the chance laws of L and let Λ be the corresponding chance hypothesis.

(a) From the fact that CCP is satisfied, it follows in particular that the conditional probability $\mu(A \mid \Lambda)$ must be well-defined. So $\mu(\Lambda) > 0$. Since λ was arbitrary, the corresponding fact must hold for each chance law of L. Since the chance hypotheses corresponding to distinct chance laws must be disjoint, this means that L must have only countably many laws of chance.

(b) CCP tells us that a certain condition holds for all Borel sets A—so in particular, it holds when we set A to Λ. So L is proper:

$$\begin{aligned} \lambda(\Lambda) &= \mu(\Lambda \mid \Lambda) \\ &= 1. \end{aligned}$$

(c) Finally, let Λ^c be the complement of Λ in 2^ω. Then $\mu|_\Lambda$ and $\mu|_{\Lambda^c}$ are measures on 2^ω and $\mu = \mu|_\Lambda + \mu|_{\Lambda^c}$. Now, CCP says that for any measurable A,

[27] That is, μ admits an orthogonal decomposition in which λ appears with positive weight. Compared with the notion of a *grain of truth* of Kalai and Lehrer [37]: specialized to our setting, this requires that $\mu = \nu + c \cdot \lambda$ without requiring that $\nu(\Lambda) = 0$.

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press

$$\begin{aligned}\lambda(A) &= \mu(A \mid \Lambda)\\ &= \frac{\mu(A \,\&\, \Lambda)}{\mu(\Lambda)}\\ &= \frac{1}{\mu(\Lambda)}\mu|_{\Lambda}(A).\end{aligned}$$

So $\mu = \mu|_{\Lambda^c} + c \cdot \lambda$, where $c := \mu(\Lambda) > 0$: μ contains λ as a nugget of truth. □

Corollary 3.5. *Let L be a proper theory of chance with countably many chance laws $\lambda_1, \lambda_2, \ldots$ (with chance hypotheses $\Lambda_1, \Lambda_2, \ldots$). A prior μ satisfies CCP for L if and only if it can be written in the form*

$$\mu = \nu + \sum c_k \cdot \lambda_k,$$

with each $c_k > 0$ and ν a (possibly trivial) measure on 2^ω such that $\nu(\Lambda_k) = 0$ for each k.

Corollary 3.6. *A theory of chance is CCP-satisfiable if and only if it is proper and has only countably many laws of chance.*

Proof. The left to right implication follows immediately from the preceding proposition. For the other half, let L be a proper theory of chance with chance laws $\lambda_1, \lambda_2, \ldots$. It suffices to show the existence of a prior μ that contains each of the λ_k as a nugget of truth.

Let $\mu_0 = \sum \frac{1}{2^k}\lambda_k$. Let $M := \{\tau \in 2^{<\omega} \mid \mu_0([\![\tau]\!]) = 0\}$. If M is empty, then μ_0 is a prior and we can take $\mu = \mu_0$. Otherwise, enumerate M as $\tau_1, \tau_2, \ldots$ (M is countable, being a subset of $2^{<\omega}$, and is infinite, since any extension of a string in M is in M). Then, recalling the notation of Example 1.3, $\mu_1 := \sum \frac{1}{2^k}\nu_{.5}^{\tau_k}$ is a probability measure concentrated on the L-lawless sequences. So we can take $\mu = \frac{1}{2}\mu_0 + \frac{1}{2}\mu_1$. □

Of our examples of theories of chance from Section 2, Super-Determinism and Basic Frequentism are not CCP-satisfiable (they have too many chance laws), but Computable Super-Determinism, Computable Frequentism, and the Toy Best System account are.[28]

§4. Chance, credence, and computability.

As noted above, fans of the Principal Principle like to present it as Janus-faced, both giving us substantive information about the nature of chance and giving us a substantive constraint on rational Bayesian priors. From this perspective, something has gone wrong if we find no computable prior satisfies CCP for a certain reductionist theory of chance: the theory of chance should be discarded; we need to replace CCP with some weaker condition relating credence and chance; or we need to provide an account of the normative force of a putative requirement of rationality that requires agents to perform supra-computable tasks.

[28] If we consider the analogs of these theories in the context of worlds modelled by finite binary strings, we find that obvious versions of Frequentism and of the Best System approach are unsatisfiable with respect to the finite-world version of CCP (because they are improper), while the obvious finite-world version of Super-Determinism is satisfiable for it.

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press

Definition 4.1 (CCCP). *Let L be a theory of chance and let μ be a prior. We say that μ satisfies the* Computable Chance–Credence Principle (CCCP) *for L if μ is computable and satisfies CCP for L.*

Definition 4.2. *We call a theory of chance* CCCP-satisfiable *if there is a prior satisfying CCCP for it; otherwise we call it* CCCP-unsatisfiable.

We are going to see that there are nontrivial obstructions to CCCP-satisfiability. Indeed, *none* of our paradigm examples of theories of chance are CCCP-satisfiable.

4.1. Learning and delta-functions. Our first objective will be to show that Computable Super-Determinism and the Toy Best System Theory are not CCCP-satisfiable. We will approach this result indirectly via one of the classic computer science models of inductive learning (this will give us purchase in Section 4.3 on the question of what sort of theories *are* CCCP-satisfiable).[29]

Definition 4.3. *An* extrapolating machine *is a computable map $m : 2^{<\omega} \to \{0, 1\}$.*

Definition 4.4. *Let* m *be an extrapolating machine and let $S \in 2^{\omega}$. We say that* m NV-learns S *if there is an N such that $m(S\restriction n) = S(n+1)$ for all $n > N$.*

Informal picture: a learning agent is being fed an infinite binary data stream S bit-by-bit and each time a bit is revealed, the agent makes a guess as to the identity of the next bit. An extrapolating machine m is a computable strategy that an agent might use for arriving at such guesses: when given a binary string τ (a finite data set), m outputs a bit $m(\tau)$. Extrapolating machine m successfully NV-learns S just in case from some point onwards in processing S, all of m's guesses are correct (here NV=*next value*). Note that any binary sequence NV-learned by an extrapolating machine must be computable.

Proposition 4.5 (Putnam [54]a,b). *No extrapolating machine NV-learns each computable sequence.*

Proof. Let m_0 be an extrapolating machine and define the binary sequence S as follows: the first bit of S is $1 - m_0(\varnothing)$ (the opposite of what m_0 predicts on empty input); and in general, $S(k+1)$ is the opposite of what m_0 predicts when shown the first k bits of S. S is computable (since m_0 is). And m_0 does not NV-learn S (since it never predicts a bit correctly). □

Given a computable prior μ we can define an extrapolating machine m_μ as follows. Let κ be a computable number in the open unit interval that is not in the set $X_\mu := \{\mu(b \mid \sigma) \mid b \in 2, \sigma \in 2^{<\omega}\}$ (i.e., μ never uses κ as the conditional probability that the next bit will be 0 or that it will be 1).[30] Define $m_\mu : 2^{<\omega} \to 2$ as follows: on input of

[29] This model of learning has its roots [54, 55] and was codified and developed in [6] and in [16]. For classic surveys see [2] and [50, chap. VII].

[30] It is always possible to find such a κ: let $\tau_1, \tau_2, \ldots$ be a computable enumeration of the binary strings; then $\mu(0|\tau_1), \mu(1|\tau_1), \mu(0|\tau_2), \mu(1|\tau_2), \ldots$ is an enumeration of some computable real numbers; since we have μ, we can calculate any of these to any desired degree of accuracy; so the numbers listed are uniformly computable and so must not include all computable numbers in the unit interval (see Remark 4.14). A variant of this argument shows that we can choose a κ arbitrarily close to .5: otherwise the computable numbers in some open subinterval around .5 would be uniformly computable—and by omitting some initial bits,

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press

binary string σ, m_μ outputs 0 if $\mu(0 \mid \sigma) > \kappa$ and outputs 1 if $\mu(0 \mid \sigma) < \kappa$. Since κ and μ are computable, so is m_μ.[31]

PROPOSITION 4.6. *Let μ be a prior on 2^ω and let S be a binary sequence. If $\mu = \nu + c \cdot \delta_S$, where $c > 0$ and ν is a measure, then m_μ NV-learns S.*

Proof. We can assume without loss of generality that $\nu(\{S\}) = 0$. We then have

$$\begin{aligned}\mu(S(n+1) \mid S\restriction n) &= \frac{\mu(S\restriction(n+1))}{\mu(S\restriction n)}\\ &= \frac{\nu(S\restriction(n+1)) + c \cdot \delta_S(S\restriction(n+1))}{\nu(S\restriction n) + c \cdot \delta_S(S\restriction n)}\\ &= \frac{\nu(S\restriction(n+1)) + c}{\nu(S\restriction n) + c},\end{aligned}$$

since $\delta_S(S\restriction n) = 1$ for all n. Since $\nu(\{S\}) = 0$, by making n sufficiently large we can make $\nu(S\restriction n)$—and therefore also $\nu(S\restriction(n+1))$—as small as we like.[32] So by making n sufficiently large, we can make $\mu(S(n+1) \mid S\restriction n)$ as close to one as we like. So for sufficiently large n, m_μ always predicts the next bit of S correctly (i.e., m_μ NV-learns S—no matter what choice of κ we made in defining m_μ). □

PROPOSITION 4.7. *Computable Super-Determinism and the Toy Best System Theory are CCCP-unsatisfiable.*

Proof. Any prior μ on 2^ω that satisfied CCCP for one of these theories would in particular satisfy CCP for that theory—and so, by Proposition 3.4, would contain each computable delta-function measure as a nugget of truth. But then by Proposition 4.6, the extrapolating machine m_μ induced by μ would NV-learn each computable sequence. But Proposition 4.5 shows that to be impossible. □

So Computable Super-Determinism and the Toy Best System Theory, although CCP-satisfiable, are CCCP-unsatisfiable: they admit priors that satisfy the weak form of the Principal Principle that we have been considering; but all such priors are uncomputable.

REMARK 4.8. *Adleman and Blum* [1] *show that if A is an oracle, then there exists a Turing machine with access to A that NV-learns every computable sequence if and only if A is high (i.e., A enables the computation of a function that grows more quickly than any computable function). So the problem of NV-learning all computable sequences is unsolvable, but strictly easier than the Halting Problem.*

4.2. Learning and Bernoulli measures. The above discussion shows that Computable Super-Determinists must rein in their ambitions, maintaining that only a well-behaved proper subset of computable delta-function measures can correspond to chance laws, if they want their theory to support the existence of a prior satisfying

we can transform a listing of the computable numbers in such a subinterval into a listing of all computable numbers in the unit interval. Note that it is not required for present purposes that the identification of a suitable cutoff κ for a given extrapolating machine need itself be a computable task.

[31] We choose $\kappa \notin X_\mu$ because determining that two computable numbers are equal is not in general as computable task. On this point see, e.g., [53, theorem 3.3].

[32] This follows from the general fact that measures exhibit continuity from above—see e.g., [43, theorem 1.36].

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press

CCCP. We will see next that Computable Frequentists must likewise fall back to a position on which only some of the computable Bernoulli measures arise as laws of chance, if they want their theory to support the existence of a prior satisfying CCCP. And, of course, fans of the Toy Best System will be driven to make both moves at once.[33]

We begin by recalling some basic notions from the theory of algorithmic randomness deriving from [48]. A family $\mathcal{U} = \{U_k\}_{k\in\mathbb{N}}$ of subsets of 2^ω is called a *uniformly effective family of open sets* if there is a Turing machine that on input $k \in \mathbb{N}$ outputs a sequence $\tau^k_1, \tau^k_2, \ldots$ of binary strings with the feature that

$$U_k = \bigcup_{j=1}^{\infty} [\![\tau^k_j]\!].$$

That is: each U_k can be effectively approximated as a union of basic subsets of 2^ω; and this approximation is uniform in k, in the sense that a single machine can handle each of the U_k.[34] Let ν be a computable probability measure on 2^ω. Then an *effective* ν*-null set* is a subset T of 2^ω that can be written in the form $T = \bigcap U_k$, for some uniformly effective family of open sets, $\{U_k\}$, satisfying $\nu(U_k) \leq 2^{-k}$.[35] We say that a binary sequence is ν*-Martin-Löf random* if it does not belong to any effective ν-null set and we denote by ML_ν the set of such sequences. Note that each computable $\nu \in \mathcal{P}$ assigns measure one to its ML_ν.[36]

Proposition 4.9. *Let* $R = \{r_k\}_{k\in I}$ *be a set of computable binary sequences and for each* $k \in I$ *let* ν_k *be the Bernoulli measure whose parameter admits* r_k *as a binary expansion. If* μ *is a prior that can be written in the form* $\mu = \nu + \sum c_k \cdot \nu_k$ (*with* ν *a possibly trivial measure and each* $c_k > 0$), *then there is an extrapolating machine that NV-learns each* r_k.

Proof. (1) We claim that for any $m \in I$ and any binary sequence S, if $S \in \mathsf{ML}_{\nu_m}$, then $S \in \mathsf{ML}_\mu$. For suppose that $S \notin \mathsf{ML}_\mu$. Then there exists a uniformly effective family of open sets, $U_1, U_2, \ldots$, with $\mu(U_k) \leq 2^{-k}$ (for each $k \in \mathbb{N}$) and $S \in \bigcap_{k=1}^{\infty} U_k$. Choosing j large enough to ensure that $1/c_m \leq 2^j$, we have

$$\begin{aligned} \nu_m(U_k) &\leq \frac{1}{c_m}\mu(U_k) \\ &\leq 2^{-k+j} \end{aligned}$$

for each $k \in \mathbb{N}$. So if we take $V_k := U_{k+j}$, then the V_k form a uniformly effective family of open sets and satisfy $\nu_m(V_k) \leq 2^{-k}$, with $S \in \bigcap V_k = \bigcap U_j$. So $S \notin \mathsf{ML}_{\nu_m}$.

[33] Much of the argument of this section was suggested by Christopher Porter (in private communication).

[34] And if we equip 2^ω with the product topology (as in footnote 6), then each U_k is an open subset of 2^ω.

[35] Commentary: such T are ν-null sets that can be effectively specified in a certain sense. To belong to a given effective ν-null set is to be special in a certain effectively specifiable way. The intuition behind the present approach is that a sequence is ν-random if it avoids each such null set. You can think of the ν-Martin-Löf random sequences as those that exhibit no effectively specifiable behaviour that would be arbitrarily surprising to agents expecting to see data sampled from ν.

[36] Since there can only be countably many effective ν-null sets, their union must be a ν-null set.

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press

(2) Next, we invoke a result due to Porter: for any computable measure λ on 2^ω, there is a computable measure ζ on 2^ω such that for each $r \in [0, 1]$, if $\mathsf{ML}_\lambda \cap \mathsf{ML}_{\nu_r} \neq \varnothing$, then $r \in \mathsf{ML}_\zeta$.[37] Applied to our case, this implies that there exists a computable probability measure ζ such that $R \subseteq \mathsf{ML}_\zeta$.

(3) It follows that ζ must assign positive probability to each $\{r_k\}$: otherwise, since ζ and r_k are both computable, we could choose initial segments $\tau_1, \tau_2, \ldots$ of r_k so that the family of basic sets $[\![\tau_1]\!], [\![\tau_2]\!], \ldots$ would be uniformly effectively open and satisfy $\zeta([\![\tau_m]\!]) < 2^{-m}$, with $r_k \in \bigcap [\![\tau_j]\!]$, contradicting $r_k \in \mathsf{ML}_\zeta$. It then follows from Proposition 4.6 that the extrapolating machine m_ζ determined by ζ must NV-learn each r_k. □

Corollary 4.10. *Computable Frequentism and the Toy Best System are CCCP-unsatisfiable.*

Remark 4.11. *Vitányi and Chater* [63] *develop a model of learning in which an agent viewing a binary data stream attempts to guess a code number for a Turing machine that calculates a probability measure on* 2^ω *relative to which the data stream is Martin-Löf random. Barmpalias* et al. [5, *theorem* 1.6] *show that the set of computable Bernoulli measures is not learnable in this sense—indeed, they further show that this set is learnable relative to an oracle if and only if that oracle is high. It is natural to conjecture that an analogous result holds in our setting: there exists a prior satisfying CCP for computable Frequentism that is computable relative to a given oracle if and only if that oracle is high.*

4.3. Satisfying CCCP. So in order for a prior to satisfy CCCP for a given theory of chance, the set sequences on which the theory's delta-function laws of chance are concentrated and of parameters of its Bernoulli measure laws of chance must both be NV-learnable. Here we consider the contours of NV-learnability. We denote by $NV(m)$ that set of sequences NV-learned by extrapolating machine m.

Definition 4.12. *A set* $\mathcal{F}$ *of binary sequences is* dense *if for each binary string* τ*, there is a sequence* S *in* $\mathcal{F}$ *with* τ *as an initial segment.*[38]

Definition 4.13. *A set* $\mathcal{F}$ *of binary sequences is* uniformly computable *if there is an enumeration* $S_1, S_2, S_3, \ldots$ *of the elements of* $\mathcal{F}$ *such that the map that sends* $(k, \ell) \in \mathbb{N}^2$ *to* $S_k(\ell)$ *is computable.*

Remark 4.14. *It is immediate that each member of a uniformly computable set of sequences is computable. A straightforward diagonalization argument shows that the set of computable sequences is not itself uniformly computable. And since any finite subset of* 2^ω *is uniformly computable and the union of any two uniformly computable sets of sequences in uniformly computable, it follows that any uniformly computable family of binary sequences must exclude infinitely many computable binary sequences. Moral: to say that a set of sequences is uniformly computable is to say that it is computationally tractable in a certain sense.*

[37] See [52, theorem 3.7].

[38] This is just the usual topological notion, if we impose the product topology on 2^ω (see footnote 6).

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press

Proposition 4.15. *Let $\mathcal{F}$ be a set of binary sequences. The following are equivalent*:

i) *$\mathcal{F}$ is uniformly computable and dense.*
ii) *There is an extrapolating machine* m *such that $\mathcal{F} = NV(m)$.*

Proof. Let $\mathcal{F} \subset 2^{\omega}$ be uniformly computable and dense. Fix an enumeration S_1, S_2, ... of the members of $\mathcal{F}$ such that the map $(k, \ell) \mapsto S_k(\ell)$ is computable. Define an extrapolating machine m as follows: on input of an n-bit binary string τ, m finds the least k such that $S_k{\restriction}n = \tau$ and gives output $m(\tau) = S_k(n+1)$ (such a k always exists since $\mathcal{F}$ is dense). For any $S_k \in \mathcal{F}$, we can find n large enough so that there is no $j < k$ with $S_j \neq S_k$ and $S_j{\restriction}n = S_k{\restriction}n$. It follows that m NV-learns each sequence in $\mathcal{F}$. Suppose, on the other hand, that m NV-learns S and let τ be an initial segment of S that contains all of the bits that m predicts incorrectly. On input τ, m uses some S_k that has τ as an initial segment to predict that next bit. This prediction is correct, so m also uses S_k to predict the next bit—and all subsequent bits. So $S = S_k$. So each sequence NV-learned by m is in $\mathcal{F}$. So $\mathcal{F} = NV(m)$.

Suppose that there is an extrapolating machine m such that $\mathcal{F} = NV(m)$. Then $\mathcal{F}$ must be dense, since it must include, for each binary string τ, the sequence S_m^{τ} that has τ as an initial segment and in which all subsequent bits are chosen to vindicate m's predictions. Further, if $\tau_1, \tau_2, \tau_3, \ldots$ is a computable enumeration of the binary strings, then $S_m^{\tau_1}, S_m^{\tau_2}, S_m^{\tau_3}, \ldots$ is an enumeration of the sequences NV-learned by m and the map $(k, \ell) \mapsto S_m^{\tau_k}(\ell)$ is computable, since m and our enumeration of the strings are both computable. So $\mathcal{F}$ is uniformly computable. □

Corollary 4.16. *Let $\mathcal{F}$ be a set of binary sequences. There is an extrapolating machine that NV-learns each sequence in $\mathcal{F}$ if and only if $\mathcal{F}$ is a subset of a uniformly computable set of binary sequences.*

Proof. Note that the union of a uniformly computable set of sequences with the set of periodic sequences (i.e., those sequences that can be written in the form $\sigma\sigma\sigma\ldots$ for some binary string σ) is both uniformly computable and dense. □

Remark 4.17. *This corollary gives a version of one of the two standard characterizations of NV-learnability.*[39] *The other is*: *A set $\mathcal{F}$ of computable binary sequences is a subset of some $NV(m)$ if and only if it is a subclass of an abstract complexity class.*[40] *One gloss that has been given to this latter characterization*: *"It says, in essence, that the* [*computably*] *extrapolable sequences are the ones that can be computed rapidly."*[41] *Perhaps a more accurate gloss would be*: *for any computably NV-learnable set S of sequences, there is an at least somewhat natural notion of* rapidly computable *relative to which every sequence in S is rapidly computable.*

[39] It is usually given in the form: a set of sequences is a subset of some $NV(m)$ if and only if it is a subset of a computably enumerable family of computable binary sequences. See the references in footnote 29.

[40] See the references in footnote 29. Roughly, one defines an abstract complexity class by choosing a way of measuring the complexity of a computation (satisfying certain weak conditions) and choosing a computable upper bound on this complexity, and then restricting attention to sequences satisfying this bound. For a brief introduction to abstract complexity classes, see [32, sec. 12.7]. For a thorough treatment, see [50, chap. VII].

[41] See [16, p. 127]. For natural senses in which any NV-learnable set of computable sequences forms a "small" subset of the set of computable sequences, see [22]. For further discussion, see [10].

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press

All of this may seem congenial to fans of the Best System approach. True, they may have to recognize fewer chance laws than they might have originally hoped—but to the extent that we can think of there being some sort of complexity cut-off that rules out certain delta-function measures and Bernoulli measures as respectable chance laws, that may seem to uphold the spirit of the Best System approach. (*But here it is important to keep in mind the arbitrariness entailed by the considerations canvassed in Remark* 4.14.)

Let us end this discussion on a positive note.

Definition 4.18. *We say that a set* $\mathcal{M}$ *of probability measures on* 2^ω *is* uniformly computable *if there is an enumeration* $\nu_1, \nu_2, \ldots$ *of the elements of* $\mathcal{M}$ *and a Turing machine that on input of a binary string* σ *and non-zero natural numbers* j *and* k, *give as output a* k-*bit approximation* (*as in Section* 1.2) *to* $\nu_j(\sigma)$.

Proposition 4.19. *If L is a proper theory of chance whose set of chance laws is uniformly computable and includes a prior, then L is CCCP-satisfiable.*

Proof. Let $\lambda_1, \lambda_2, \ldots$ be an enumeration of the chance laws of L in virtue of which these laws are uniformly computable. Consider the measure

$$\mu := \sum \frac{1}{2^k}\lambda_k.$$

It is a prior (since by assumption one of the λ_k's is). It is immediate that μ satisfies CCP for L. And it is straightforward to show that μ is computable.[42] □

§5. Will we never learn?

A reaction that people sometimes have to the Principal Principle: *Who cares how my prior behaves conditional on an event E like that of the coin being fair? If I am a true Humean, I should regard this E as something that could never be part of my evidence, since it is a holistic fact about the history of the entire universe.*[43]

Lewis has an answer to this challenge.[44]

> To the believer in chance, chance is a proper subject to have beliefs about. Propositions about chance will enjoy various degrees of belief, and other propositions will be believed to various degrees conditionally upon them.[45]

[42] See, e.g., [10, proposition 4.7].

[43] For this point of view, see [57, p. S104]. For critical discussion, see [17, sec. 3.3].

[44] See [45, pp. 84–86 and 106–109].

[45] See [45, p. 84]. Lewis [45, p. 86] offers the following example: suppose that you have credence .27 that the chance of Heads is .5, credence .22 that the chance of Heads is .35, and credence .51 that the chance of Heads is .8—what credence should you have in Heads? The obvious answer is

$$(.5 \times .27) + (.35 \times .22) + (.8 \times .51) = .62.$$

This makes sense as an application of the law of total probability (writing H for the event of Heads, A for the chance being .5, B for it being .35, and C for it being .8),

$$p(H) = p(H \mid A) \times p(A) + p(H \mid B) \times p(B) + p(H \mid C) \times p(C),$$

so long as the probability function p, which represents your credences, satisfies $p(H \mid A) = .5$, $p(H \mid B) = .35$, and $p(H \mid C) = .8$. That is, the obvious calculation works here so long as you credences satisfy a basic form of the Principal Principle.

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press

> To the subjectivist who believes in objective chance, particular or general propositions about chance are nothing special. We believe them to varying degrees. As new evidence arrives, our credence in them should wax and wane in accordance with Bayesian confirmation theory. It is reasonable to believe such a proposition, like any other, to the degree given by a reasonable initial credence function conditionalized on one's present total evidence.[46]

We are interested in the Principal Principle, not because we want to be prepared just in case we should learn the truth of some law of chance, but because chance laws can be confirmed and disconfirmed: we have varying degrees of credence in various chance laws and the Principal Principle is a constraint on the rational connection between those theoretical beliefs and concrete expectations about future events.[47]

Learnability is in fact an important (if largely subterranean) Lewisian theme in the paper in which the Principal Principle made its debut.[48] He has recourse to non-standard-valued priors because he thinks that rational priors must be regular (i.e., assign non-zero probability to each non-trivial proposition):

> it is required as a condition of reasonableness: one who started out with an irregular credence function (and who then learned from experience by conditionalizing) would stubbornly refuse to believe some propositions no matter what the evidence in their favor.[49]

In a post-script he suggests that "if we start with a reasonable initial credence function and do enough feasible investigation, we may expect our credences to converge to the chances."[50]

Now, within the standard framework that we have been working in, only Bayesian agents who assign non-zero prior probability to a chance hypothesis can become more confident of that chance hypothesis through experience. And within our standard framework, no Bayesian agent can assign non-zero prior probability to each chance hypothesis of a theory of chance that has uncountably many laws of chance. It is natural to think that something has gone wrong here. Surely Basic Frequentism, despite having uncountably many laws of chance, should provide an extraordinarily hospitable environment for a Bayesian agent attempting to learn the chance facts!

One option would be too simply excuse priors from having to defer to chance laws when they assign the corresponding chance hypothesis probability zero.[51]

Definition 5.1 (CCP*). *Let L be a theory of chance and let μ be a prior. We say that μ satisfies* CCP* *for L if: for any Borel subset A of 2^ω and for any chance law λ such that μ assigns positive probability to $\Lambda := L^{-1}(\lambda)$, we have $\mu(A \mid \Lambda) = \lambda(A)$.*

Relative to Basic Frequentism, the fair coin measure satisfies CCP* (but not CCP of course). Someone who begins life with this prior is certain from birth that the chance

[46] See [45, p. 106].

[47] On this point, see [17] as well as [45, pp. 106–109].

[48] On the other hand, in this strand of his work, Lewis shows no interest in considerations of computability.

[49] See [45, p. 88—see also p. 96].

[50] See [45, p. 121].

[51] Compare with, e.g., the version of Miller's Principle offered in [51, sec. 1].

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press

of a 0 at any particular future time is $1/2$ and remains certain of this no matter how strongly the data speak against this hypothesis and in favour of others.[52] To accept CCP* as an explication of the idea that a rational agent's credence function should defer to chance is to give up entirely on the idea that part of what deference to chance involves is being open to learning chance hypotheses.[53]

I think that we should be interested in finding variants of CCP that require priors to be open to the learning of chance hypotheses.[54] To this end, let us zoom out. The thought behind the Best System approach is that a chance law of a world is a sort of optimal scientific summary of the pattern of events at that world (relative to the ordinary standards of scientific practice). The thought behind the Principal Principle is that (in ordinary circumstances at least) to the extent that you are confident that the chance law of your world is given by λ, your credences should be close to the λ-chances. Suppose that a given prior μ satisfies at least the spirit of the Principal Principle and embodies the canons of scientific rationality (preferring simpler theories to more complex ones, etc.).[55] Then it seems that if μ sees larger and larger portions of a world with chance law λ_0, it should become more and more confident that the law of chance is λ_0-like (since μ is a good scientist, it will eventually home in on the best scientific description of its world) and hence, upon being conditionalized on larger and larger data sets, the probabilities that μ assigns events will approach those that λ_0 assigns them (since μ satisfies the spirit behind the Principal Principle).

Let L be a proper theory of chance and let λ be a chance law of L with corresponding chance hypothesis $\Lambda = L^{-1}(\lambda)$. Let S be typical in Λ. Our line of thought above tells us that as $n \to \infty$, we expect that $\mu(\cdot \mid S{\restriction}n)$ comes closer and closer to $\lambda(\cdot \mid S{\restriction}n)$ in some sense. So let us consider generalizations of CCP of the form: there exists a prior μ such that for every chance law λ, for λ-almost all sequences S in the corresponding chance hypothesis, the $\mu(\cdot, \mid S{\restriction}n)$ converge to the $\lambda(\cdot \mid S{\restriction}n)$ as $n \to \infty$. The question is: What notion of convergence should we demand here?

5.1. Merging. Recall that the *total variation distance* between probability measures λ and ν on 2^ω is $\sup_{A\in\mathcal{B}} |\nu(A) - \lambda(A)|$.[56] The following is a specialization to our context of the notion of merging introduced in [14] and subsequently widely discussed by statisticians and game theorists.

[52] Some seem to feel the temptation to say: Yes, but if that if the fair coin *is* your prior that is exactly what you should do! That is of course correct according to subjective Bayesianism. But the Principal Principle is not consistent with subjective Bayesianism: it was put forward as a part of an attempt to articulate constraints on rational priors beyond mere coherence. There is not much point to such constraints unless one is willing to say: some agents who are coherent and update by conditionalization are nonetheless irrational.

[53] The same sort of case can be made against a reformulation of CCP in terms of regular conditional probabilities—see Appendix B.4. Dmitri Gallow pointed out to me that a version of this problem arises for New Principle of Hall [28] and Lewis [46] (discussed in Appendix A): in the setting of finite worlds, the fair coin measure satisfies the New Principle for the theory of chance that assigns each world the Bernoulli measure whose parameter gives the relative frequency of 0's at that world.

[54] Further remarks about alternative directions can be found in Appendix B.

[55] I do not actually believe that it is possible to design priors that always prefer simpler hypotheses—see [9]. But here I bracket my various heretical views about the shortcomings of the Bayesian account of rationality.

[56] For background on this notion of distance and its relation to other ways of topologizing spaces of probability measures, see, e.g., [26].

Definition 5.2 (Merging). *Let λ and μ be probability measures on 2^ω. We say that μ* merges *with λ if there is a set of sequences of λ-measure one such that for each S in that set*:

$$\lim_{n\to\infty} \sup_{A\in\mathcal{B}} |\mu(A \mid S{\restriction}n) - \lambda(A \mid S{\restriction}n)| = 0.$$

Definition 5.3 (M-CCP). *Let L be a theory of chance. We say that a prior μ on 2^ω satisfies the* Merging Chance–Credence Principle (M-CCP) *if μ merges with each chance law of L.*

If a prior μ merges with a proper law of chance λ of a theory of chance, then we can take the set of sequences that witnesses merging to be a subset of the chance hypothesis of λ.

Definition 5.4. *We call a theory of chance* L M-CCP-satisfiable *if there is a prior that satisfies M-CCP for L. If there is a computable prior with this feature, we call* L M-CCCP-satisfiable.

A pair of classic results clarify what it takes for a prior to merge with a probability measure ν. Recall that if ν_1 and ν_2 are measures on a measurable space, then we say that ν_1 is *absolutely continuous* with respect to ν_2 if for each measurable set A, $\nu_1(A) > 0$ implies $\nu_2(A) > 0$. In this case we write $\nu_1 \ll \nu_2$.

Remark 5.5. *Note that if we have some measures satisfying $\mu = \nu_1 + \nu_2$, then $\nu_1, \nu_2 \ll \mu$. So, in particular, if a prior μ contains a proper law of chance λ as a nugget of truth, then $\lambda \ll \mu$.*

Proposition 5.6 (Blackwell and Dubins [14]). *Let λ be a probability measure on 2^ω and let μ be a prior on 2^ω. If $\lambda \ll \mu$, then μ merges with λ.*

Proposition 5.7 (Lehrer and Smorodinsky [44]). *Let λ be a probability measure on 2^ω and let μ be a prior on 2^ω. If μ merges with λ, then $\lambda \ll \mu$.*

Each prior merges with uncountably many probability measures.[57] But only countably many of these can be chance laws of any given reductionist theory of chance.

Proposition 5.8. *Let L be a theory of chance. If L is M-CCP satisfiable, then L has only countably many chance laws.*

Proof. Suppose that μ satisfies the M-CCP for L. Let λ be a law of chance of L with chance hypothesis $\Lambda = L^{-1}(\lambda)$. Since L is a theory of chance, we have $\lambda(\Lambda) > 0$. Since μ merges with λ, Proposition 5.7 then implies that $\mu(\Lambda) > 0$. So μ must assign positive probability to each chance hypothesis of L—so L can have only countably many chance laws. □

Corollary 5.9. *Super-Determinism and Basic Frequentism are M-CCP-unsatisfiable (and hence also M-CCCP-unsatisfiable).*

As an immediate consequence of Remark 5.5 and Propositions 3.4 and 5.6 we have:

Proposition 5.10. *If a prior satisfies CCP for a theory of chance, then it also satisfies M-CCP for that theory.*

[57] See, e.g., [10, proposition 4.4].

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press

Corollary 5.11. *A theory is M-CCP satisfiable if it is CCP satisfiable. In particular, Computable Super-Determinism, Computable Frequentism, and the Toy Best System Theory are M-CCP-satisfiable.*

Remark 5.12 (Mixtures and Merging). *Let $\{\mu_k\}_{k \in I}$ be a countable family of probability measures at least one of which is a prior and for each $k \in I$, let $q_k > 0$ with $\sum_{k \in I} q_k = 1$. Then $\mu := \sum q_k \cdot \mu_k$ is a prior. And μ merges with any measure merged with by one of the μ_k: if ν is merged with by μ_k, then $\nu \ll \mu_k$ (by Proposition 5.7), so $\nu \ll \mu$ (given how μ was defined), and so μ merges with ν (by Proposition 5.6).*

In the case of Computable Super-Determinism, a prior satisfies CCP if and only if it satisfies M-CCP. But, for more typical theories, we expect that there will be priors that satisfy M-CCP without also satisfying CCP.

Example 5.13. *A theory L and a prior μ may satisfy M-CCP without satisfying CCP (even if L is proper and μ is computable): let L be the theory that assigns the fair coin measure $\nu_{.5}$ to every sequence and let $\mu^\dagger$ be the prior that thinks there is a .9 chance that the first bit is a 0 and thinks that each subsequent bit is chosen by flipping a fair coin. We have merging (because $\nu_{.5}$ and $\mu^\dagger$ agree completely when conditionalized on any non-trivial initial segment) but CCP is violated.*

Example 5.14. *Improper theories of chance are never CCP-satisfiable. But they can be M-CCCP satisfiable. Consider the theory of chance that assigns the measure $\mu^\dagger$ of the preceding example to each sequence that begins with 0 as its law of chance (and which considers sequences that begin with 1 to be lawless). This theory is CCP-unsatisfiable (because it is improper—$\mu^\dagger$ assigns its chance hypothesis probability .9). But the fair coin measure (considered as a prior) satisfies M-CCP for this theory.*

We have a partial converse to Corollary 5.11.

Proposition 5.15. *Any proper theory of chance that is M-CCP-satisfiable is also CCP-satisfiable.*

Proof. Let L be a proper theory of chance that is also M-CCP satisfiable. We know from Proposition 5.8 that L has a countable family of laws of chance: $\lambda_1, \lambda_2, \ldots$. Let μ be a prior that merges with each λ_k. A result of Ryabko tells us that if $\mathcal{K}$ is a set of probability measures on 2^ω and there exists a probability measure μ that merges with each measure in $\mathcal{K}$, then there exists a measure μ^* that merges with each measure in $\mathcal{K}$ and which can be written as a sum of measures, each of which is a positive multiple of a measure in $\mathcal{K}$.[58] Applied to our case: there must be a μ^* that is a weighted sum of the λ_k and which merges with each λ_k. By Proposition 5.7, each $\lambda_k \ll \mu^*$—so each λ_k appears with positive weight in μ^*. So μ^* (or the result of adding something to it to make sure that we have a prior as in the proof of Corollary 3.6) satisfies CCP for L (by Proposition 3.4). □

Of course, it does not follow from this that M-CCCP-satisfiability implies CCCP-satisfiability for a proper theory of chance L: it could be that the only computable priors that merge with every chance law of L are not themselves weighted sums of the chance laws of L. So, in particular, the above results leave open the questions whether Computable Super-Determinism, Computable Frequentism, and the Toy Best System

[58] See [58, theorem 4].

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press

Theory are M-CCCP-satisfiable. We will see in Proposition 5.25 that Computable Super-Determinism is M-CCCP-unsatisfiable. We can use a variant of the proof of Proposition 4.9 to show that there are no computable priors that satisfy M-CCP for Computable Frequentism or for the Toy Best System.

PROPOSITION 5.16. *Let $R = \{r_k\}_{k \in J}$ be a set of computable binary sequences and for each $k \in J$, let ν_k be the Bernoulli measure whose parameter admits the binary expansion r_k. If μ is a computable prior that merges with each ν_k, then there is an extrapolating machine that NV-learns each r_k.*

Proof. It suffices to show that for each $r_k \in R$, $\mathsf{ML}_\mu \cap \mathsf{ML}_{\nu_k} \neq \varnothing$, and then to argue as in steps (2) and (3) of the proof of Proposition 4.9.

For each $k \in J$, let N_k be the set of sequences in which 0's have limiting relative frequency r_k. Then we can write $\mu = \nu + \sum_{k \in J} c_k \mu_k$, where ν is a (possibly trivial measure) that considers each N_k a null set, each $c_k > 0$, and each μ_k is a probability measure that assigns N_k probability one (so $c_k \cdot \mu_k$ is the restriction of μ to N_k). Since μ merges with each ν_k, we know via Proposition 5.7 that $\nu_k \ll \mu$, from which it follows that $\nu_k \ll \mu_k$.

The μ_k may or may not be computable. But the notion of Martin-Löf randomness can be generalized to apply to arbitrary probability measures: we call a sequence S *blindly Martin-Löf random* relative to the probability measure μ if there is no uniformly effective family of open sets $U_1, U_2, \dots$ such that $S \in \bigcap U_k$ and $\mu(U_k) \leq 2^{-k}$ for each k.[59] We write BML_μ to denote the set of sequences blindly Martin-Löf random relative to the probability measure μ (of course $\mathsf{BML}_\mu = \mathsf{ML}_\mu$ when μ is computable). Note that we have $\mu(\mathsf{BML}_\mu) = 1$ for any probability measure μ.[60]

We claim that for any $k \in J$, we have $\mathsf{BML}_{\mu_k} \cap \mathsf{ML}_{\nu_k} \neq \varnothing$. For suppose otherwise. Then of course $\mu_k(\mathsf{ML}_{\nu_k}) = 0$ (since $\mu_k(\mathsf{BML}_{\mu_k}) = 1$). But since $\nu_k(\mathsf{ML}_{\nu_k}) = 1$, this contradicts the fact that $\nu_k \ll \mu_k$.

And we can now argue, as in the first step of the proof of Proposition 4.9, that since μ can be decomposed as a sum of measures in which each μ_k appears with non-zero weight, every sequence in BML_{μ_k} must also be in ML_μ. So we have, as desired, that for each $r_k \in R$, $\mathsf{ML}_\mu \cap \mathsf{ML}_{\nu_k} \neq \varnothing$. □

COROLLARY 5.17. *There are no computable priors that satisfy M-CCP for Computable Frequentism or for the Toy Best System.*

5.2. *Weak merging.* M-CCP is weaker than CCP and M-CCCP is weaker than CCCP. But for proper theories of chance M-CCP satisfiability is equivalent to CCP-satisfiability and it would be surprising if there were a significant gap between M-CCCP-satisfiability and CCCP-satisfiability for such theories. So it makes sense to consider a further alternative to CCP. Here is another criterion of learning that has played some role in game theory.[61]

[59] For this notion, see [42]. For some purposes it is natural to work with another generalization of the notion of Martin-Löf randomness to the setting of arbitrary measures, on which one requires μ-random sequences to avoid all μ-null sets that are effectively definable relative to an oracle encoding μ. See [56].

[60] The reasoning of footnote 36 carries over unchanged.

[61] Weak merging was introduced in [38]. The presentation here follows that of Lehrer and Smorodinsky [44]. For further discussion, see [10] (where weak merging is called *next-chance learning*).

Definition 5.18 (Weak merging). *Let λ and μ be probability measures on 2^ω. We say that μ* weakly merges *with λ if there is a set of sequences of λ-measure one such that for each S in that set*:

$$\lim_{n\to\infty} |\mu(0 \mid S\restriction n) - \lambda(0 \mid S\restriction n)| = 0.$$

Definition 5.19 (WM-CCP). *Let L be a theory of chance. We say that a prior μ on 2^ω satisfies the* Weak Merging Chance–Credence Principle (WM-CCP) *for L if μ weakly merges with each chance law of L.*

If a prior μ weakly merges with a proper law of chance λ of a theory of chance, then we can take the set of sequences that witnesses weak merging to be a subset of the chance hypothesis of λ.

Definition 5.20. *We call a theory of chance* L WM-CCP-satisfiable *if there is a prior that satisfies WM-CCP for L. If there is a computable prior with this feature, we say that L is* WM-CCCP-satisfiable.

The following is immediate.

Proposition 5.21. *If one probability measure merges with another, then the first also weakly merges with the second. If a theory of chance is CCP- or M-CCP-satisfied (CCCP- or M-CCCP-satisfied) by a given prior, then it is also WM-CCP-satisfied (WM-CCCP-satisfied) by it. If a theory is CCP- or M-CCP-satisfiable (CCCP- or M-CCCP-satisfiable) then it is also WM-CCCP-satisfiable (WM-CCCP-satisfiable).*

Remark 5.22 (Stronger than it looks). *In moving from merging to weak merging we lower our sights—rather than requiring that, in the limit of large data sets sampled from λ, for every event, the judgement of μ of the probability of that event must conform to the judgement of λ (in a fashion uniform across events) we require this only for a special class of events—the identity of the next bit to be revealed.*

But to require weak merging is to require more than may be immediately apparent. It turns out that μ weakly merges with ν if and only if: as the size of the data set goes to infinity, for each k, *μ must asymptotically learn to defer to ν* in a way uniform across all events that look k time-steps beyond the data *about the chance of all such events.*[62] *So merging is stronger than weak merging precisely in paying attention to events involving an infinite number of times (a type events about whose epistemological salience a Humean might well be suspicious).*

Example 5.23. *Consider the Bayes–Laplace prior.*[63] *This is the computable probability measure $\bar{\nu}$ determined by the rule that if τ is an* m*-bit string containing ℓ 0's, then*

$$\bar{\nu}([\![\tau]\!]) := \frac{\ell!\,(m-\ell)!}{(1+m)!}.$$

Famously, there is a sense in which $\bar{\nu}$ is a Lebesgue-uniform mixture of the Bernoulli measures—and from this it follows that $\bar{\nu}$ weakly merges with each Bernoulli measure.[64]

[62] See [44, definition 9 and remark 5].

[63] For more on this prior, see, e.g., [21, pp. 119–121], [65, 740 f. and 750], and [34, examples 7.2.4, 7.2.27, and 7.4.14].

[64] See, e.g., [23, 24].

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press

It then follows that $\bar{\nu}$ satisfies WM-CCP (indeed, WM-CCCP) for Basic Frequentism and for Computable Frequentism.

This example shows that theories with uncountably many laws of chance can be WM-CCP-satisfiable (and even WM-CCCP-satisfiable), even though they must be CCCP-, CCP-, M-CCCP, and M-CCP-unsatisfiable. But there are limits.

Proposition 5.24. *Let L be a theory of chance and let μ be a prior that satisfies WM-CCP for L. Then the laws of chance of L include only countably many delta-functions. In particular, Super-Determinism is WM-CCP-unsatisfiable.*

Proof. We generalize the notion of an extrapolating machine: an *extrapolator* is a (not necessarily computable) function from binary strings to bits. We extend the notion of NV-learning to extrapolators: extrapolator m *NV-learns* sequence S if when shown S bit by bit, m is eventually always correct in its predictions about the next bit. The set of sequences NV-learned by an extrapolator m is always countable (each such sequence can be generated by extending some binary string by asking m what it expects to see next). Any prior μ determines an extrapolator m^*_μ via the rule: $m^*_\mu(\sigma) = 0$ if $\mu(0 \mid \sigma) \geq .5$, otherwise $m^*_\mu(\sigma) = 1$.

Suppose that prior μ satisfies WM-CCP for theory of chance L. Suppose that for some $S \in 2^\omega$, δ_S is a law of chance of L. We can make $\mu(0 \mid S\restriction n)$ as close as we like to $\delta_S(0 \mid S\restriction n)$ by choosing n sufficiently large. It follows that m^*_μ NV-learns S. Since m^*_μ cannot NV-learn an uncountable set of sequences, there can only be countably many delta-functions among the laws of chance of L. □

Computable Super-Determinism and the Toy Best System Theory are WM-CCP-satisfiable (being CCP-satisfiable). But they are not WM-CCCP-satisfiable.

Proposition 5.25. *No computable prior weakly merges with each delta-function concentrated on a computable sequence.*

Proof. Let μ be a computable prior with associated extrapolating machine m_μ (as in Section 4.1) and let S be a binary sequence. Arguing as in Proposition 5.24 we see that if μ weakly merges with δ_S, then m_μ NV-learns S. But Proposition 4.5 implies that no extrapolating machine NV-learns each computable sequence. □

Remark 5.26 (Mixtures and weak merging). *In Remark 5.12, we saw that when we build a prior by taking a non-trivial mixture of some measures, that prior merges with every probability measure merged with by any component of the mixture. The picture is different with weak merging: Ryabko and Hutter give an example in which μ is a prior, ν is a probability measure, and $\frac{1}{2}(\mu + \nu)$ fails to weakly merge with a delta-function measure that is weakly merged with by μ.*[65] *So it is not obvious that you can build a prior suited to a Best System theory of chance by taking a mixture of a prior (such as the Bayes–Laplace prior) that weakly merges with a desirable range of Bernoulli measures with a measure that weakly merges with a desirable range of delta-function measures.*

[65] See [59, proposition 10]. In this connection, see also [44, corollary 6].

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press

§6. Scorecard.

Satisfiable?	Super-Determinism	Computable Super-Determinism	Basic Frequentism	Computable Frequentism	Toy Best System Theory
CCP	No	Yes	No	Yes	Yes
M-CCP	No	Yes	No	Yes	Yes
WM-CCP	No	Yes	Yes	Yes	Yes

Computably Satisfiable?	Super-Determinism	Computable Super-Determinism	Basic Frequentism	Computable Frequentism	Toy Best System Theory
CCCP	No	No	No	No	No
M-CCCP	No	No	No	No	No
WM-CCCP	No	No	Yes	Yes	No

Super-Determinism: Because this theory has uncountably many chance laws, it is CCP-unsatisfiable (Proposition 3.4) and M-CCP-unsatisfiable (Corollary 5.9). A learning-theoretic argument shows that it is also WM-CCP-unsatisfiable (Proposition 5.24). It is, then, of course also CCCP-unsatisfiable, M-CCCP-unsatisfiable, and WM-CCCP-unsatisfiable.

Computable Super-Determinism: For this theory, any prior that can be written as a sum in which each delta-function measure concentrated on a computable sequence appears with non-zero weight will satisfy CCP (Proposition 3.4); it will therefore also satisfy M-CCP (Corollary 5.11) and WM-CCP (Proposition 5.21). However, it turns out that no such prior is computable—and, indeed, that the theory is CCCP-unsatisfiable (Proposition 4.7). It is WM-CCCP-satisfiable (Proposition 5.25) and therefore (via Proposition 5.21) also M-CCCP-unsatisfiable.

Basic Frequentism: Because this theory has uncountably many chance laws, it is CCP-unsatisfiable (Proposition 3.4) and M-CCP-unsatisfiable (Corollary 5.9). It is of course therefore also CCCP-unsatisfiable and M-CCCP-unsatisfiable. But it is WM-CCP-satisfiable and even WM-CCCP-satisfiable (Example 5.23).

Computable Frequentism: For this theory, any prior that can be written as a sum in which each Bernoulli measure ν_r with a computable parameter r appears with non-zero weight will satisfy CCP (Proposition 3.4). Priors of this form will also satisfy M-CCP (Proposition 5.10) and WM-CCP (Proposition 5.21). But since no prior that can be written as a sum in which each computable Bernoulli measure appears with non-zero weight can be computable (Corollary 4.10), Computable Frequentism is not CCCP-satisfiable (by Proposition 3.4). Nor is it computably M-CCP satisfiable (see Corollary 5.17). On the other hand: Computable Frequentism *is* WM-CCCP-satisfiable (Example 5.23).

Toy Best System Theory: The Toy Best System is not CCCP-satisfiable (see Proposition 4.7 or Corollary 4.10). It is also not WM-CCCP-satisfiable (Proposition 5.25)—from which it follows that is not M-CCCP-satisfiable (Proposition 5.21—or see Corollary 5.17). On the other hand, this theory is CCP-satisfiable (Proposition 3.4). So it is also M-CCP-satisfiable (Corollary 5.11) and WM-CCP-satisfiable (Proposition 5.21).

§7. Conclusion. The main goal of this paper has been to show that it is more difficult one might think to find a consistent package consisting of a reductionist theory of chance and a principle encoding the rational relation between credence and chance—especially if one expects there to be rationally permitted computable priors.[66]

In Section 3 we saw there exists a prior probability measure satisfying CCP relative to a given proper theory of chance if and only if that theory has only countably many chance laws. This is, of course, a direct consequence of the choice to use ordinary (i.e., countably additive) probability measures to model rational credal states. Some reductionists about chance may be happy to ignore theories with uncountably many chance laws—e.g., advocates of best-system approaches who take fully seriously the idea that each chance law should be something like a humanly expressible theory of the world. But many reductionists will, I suspect, regard Basic Frequentism, the laws of chance of which are the Bernoulli measures, as a theory that should be compatible with our account of the relation between chance and credence. In the first instance, such reductionists face a choice between accepting that credal states can be represented by objects more general than probability measures and replacing CCP with something weak enough to accommodate the existence of rationally permitted priors adapted to theories of chance with uncountably many chance laws.

In Section 4 we saw that even if priors exist that satisfy CCP relative to a given theory of chance, there may be no computable priors with this feature. In particular this happens when the laws of chance of a theory include each delta-function measure concentrated on a computable sequence—the root of the problem being that the computable sequences, though enumerable, are not computably enumerable. A related problem arises for theories whose chance laws include each Bernoulli measure with a computable parameter. I do not think we should rest content with an account of chance and credence that tells us that rationality requires us to adopt a non-computable credence function—any more than we would be satisfied with an account of rationality that required rational agents to be able to solve the Halting Problem. Some reductionists may be happy with the moral that certain *prima facie* attractive theories of chance are unsuited to computable agents—such reductionists will then need to be careful to avoid using computationally intractable sub-families of the delta-function measures and Bernoulli measures as laws of chance. Others may again be interested in investigating whether the problem can be avoided by generalizing the Bayesian framework or by endorsing a constraint of theories of chance on rationally permitted priors weaker than CCP.

[66] Let me reiterate that non-reductionist accounts of chance are neglected here not because I take them to be immune to the sort of problems encountered above, but because the present framework would need to be extended in order to accommodate them.

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press

There are a number of standard strategies that Bayesians have explored in other contexts in which the presence of an uncountable number of alternatives and/or awkward null sets causes trouble, including infinitesimal-valued credence functions, merely finitely additive probabilities, regular conditional probabilities, and primitive conditional probabilities. These are considered briefly in Appendix B, where it is argued that none of them offers much help in our current predicament.

Section 5 explored two ways of weakening CCP with an eye to satisfying those desiderata: M-CCP and WM-CCP. The latter turns out to be the more fruitful for our purposes. It says, roughly, that the rational priors are those with the feature that, almost certainly, in the limit of large data sets the posterior probabilities will converge to the true chances for events that depend on only finitely many times. WM-CCP has the attractive feature that it is satisfied by the computable Laplace–Bayes prior, both relative to Basic Frequentism (with its uncountable family of chance laws) and relative to Computable Frequentism (with its enumerable but computationally intractable family of chance laws). So it is a substantive constraint on priors that is consistent with the intuition that frequentism in its several forms is one of the more forgiving of reductionist approaches—and with the thesis that rationality should be consistent with computability. But like CCP, WM-CCP cannot be satisfied by any computable prior relative to a theory of chance that includes among its laws of chance all delta-function measures concentrated on computable sequences.

Other responses to the problems encountered in Sections 3 and 4 are of course possible. One of the most natural would be to reject the assumption that we have implicitly carried over from the literature on the Principal Principle: that our account of the relation between credence and chance should be the basis of a dichotomy between priors that are rationally permitted and those that are not rationally permitted. Perhaps we should instead think of considerations about chance as inducing a partial ordering *more rational than* on possible priors. For instance, given a theory of chance L and a prior μ we could look for the largest sub-theory of L (= restriction of L to a subset of its domain of definition) such that μ satisfies one of our principles (CCP, M-CCP, or WM-CCP) for that sub-theory, considering μ_1 to be no more rational than μ_2 if the sub-theory L_1 of L associated with μ_1 is itself a sub-theory of the sub-theory L_2 or L associated with μ_2. If there are priors that satisfy the given principle for a given theory of chance then they will be fully rational in the sense that no prior is more rational then them—but otherwise, there will just more and less rational priors without any being fully rational.

Applied to CCP or to M-CCP, this strategy is not particularly enticing: it tells us that the fair-coin measure is a more rational prior for Basic Frequentism than the Laplace–Bayes indifference prior—even though agents with the fair coin measure as their prior will remain certain the chance that the next bit will be 0 is .5 no matter what data they see, while the posterior probabilities that agents with the Laplace–Bayes prior assign to that event will converge (almost surely) to the true chance (if one there be) in the limit of large data sets. But applied to WM-CCP, this strategy suggests a way forward for fans of the best-system-type approaches in grappling with the problems surrounding delta-function measures concentrated on computable sequences (although some challenges would remain).

This is far from a complete account of possible responses to problems exposed above. But I hope, at least, that one moral is clear: interesting work remains to be done for anyone who is attracted to Lewis's accounts of chance and credence and

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press

who is also inclined to take the computability of priors to be consistent with their rationality.

§Appendix A. Lewisiana. This appendix examines the relation between the approach adopted above and the main stream of literature on the Principal Principle. The two chief goals are to clarify the relation between the Chance–Credence Principle and the Principal Principle and to examine the proper role of improper theories of chance.

A.1. CCP and the principal principle. Lewis thought of his Principal Principle (PP) as a generalization of the sort of principle associated with Miller and Hacking.[67] Roughly speaking: PP says not only that a rational prior should agree with the chance facts when conditionalized on the chancemaking facts, but also there are many further propositions that are screened off by the chance facts.[68] CCP is intended to be a special case of (a reasonable explication of) Lewis's PP in which we ignore such further screened-off propositions (along with attendant complications about admissibility). For the curious, I offer further comments on the relation between CCP and PP.

1. As Lewis formulates it, PP has no relativity to a theory of chance. But such relativity is unavoidable: PP is intended to compatible with a range of reductionist theories of chance (Lewis was not committed to the Best-System account when he first formulated PP); and priors that satisfy the spirit of PP with respect to one such theory may be incompatible with the spirit of PP with respect to other such theory (consider, e.g., the question whether it is rationally required to put non-zero prior credence on the sequence that eternally alternates between 0 and 1).
2. As Lewis formulates PP, it involves talking about the chances at a time, which in the present framework would involve conditionalizing a law of chance on a binary string. Of course, one such string is the empty string—and specializing a Lewis-style formulation to that case gives a version in which time no longer plays a role. And: no content is lost in via this specialization.[69]
3. Omitting all the bits about the screened-off event E and the time t, Lewis's PP reads:

   > Let C be any reasonable initial credence function. Let x be any real number in the unit interval. Let X be the proposition that the chance of A's holding equals x. Then $C(A \mid X) = x$.[70]

   For Lewis, a proposition is a set of worlds. So in our context, the X must correspond to a subset of 2^{ω}. Which one? Recall that above we quoted Lewis as saying that the idea behind PP is that a rational prior should have the feature that it says that the chance of an outcome is 50% when conditionalized on the

[67] See [7, Letters 659 and 660].

[68] It is not clear that Lewis achieves what he set out to do: if a prior satisfies CCP then it is (essentially) a weighted sum of the laws of chance—so the only freedom available in choosing a prior reduces to choosing these weights, and screening-off conditions are not going to help to fix these. In this connection, see the discussion of Pettigrew [51, sec. 1].

[69] On this point, see [51, fns. 3 and 4].

[70] Adapted from [45, p. 87]. See also [46, sec. 5].

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press

proposition that makes it true that the chance of that outcome is 50%. So in the schema above, X should be the proposition that the chancemaking pattern obtains in virtue of which the chance of A is x. In the case where λ is the only law of chance that assigns A chance x, this is the pattern that makes λ be the law of chance—that is, the chance hypothesis Λ corresponding to λ. More generally, there may be more than one chance law that assigns A chance x—and then X will be the disjunction of the corresponding chance hypotheses. But if a prior satisfies a Lewis-style principle that deals in such disjunctive events, then it also satisfies one in the style of CCP (since you can always replace A by the conjunction of A with the proposition that a certain chance law holds, which lands you in the simple case considered above). And, at least in the case where the disjunction of chance laws under consideration involves only finitely many chance laws, the converse also holds.[71]

A.2. What about the New Principle? Each of the five paradigmatic theories of chance that we have been concerned with is proper. If we were to work with worlds whose complete histories corresponded to finite binary strings instead of to infinite binary sequences, it would have been difficult to find interesting proper theories of chance.[72] For instance, since the fair coin measure assigns positive probability to each finite string, it can be a proper law of chance of a theory for finite worlds only if it is the sole law of chance.

However, there are improper theories of chance even in the infinite setting. Example 2.10 provided a toy example: a theory of chance in which all and only sequences beginning with 1 get the fair coin measure $\nu_{.5}$ as their (improper) law of chance.

It is not obvious that it makes sense to assert the proposition corresponding to an improper law of chance. In Example 2.10, for instance, the relevant proposition implies both that the first bit will be 1 and that for each bit, there is only a 50–50 chance that it will be 1. Now, corresponding to any theory of chance L there is a proper theory of chance $\bar{L}$ whose laws of chance are given by $\bar{\lambda} := \lambda(\cdot \mid L^{-1}(\lambda))$ (so that $L = \bar{L}$ if and only if L is proper).[73] To my mind, it is more natural (because less Moore-paradoxical) to interpret someone apparently advancing an improper chance L as having chosen an clumsy way to assert $\bar{L}$.[74]

The observation that the Principal Principle is inconsistent with improper theories of chance launched the large and messy literature on undermining with its many attempts to correct the Principal Principle. If you are the sort of person who likes improper theories of chance, then it is worth noting the following. Specialized to our setting, the New Principle of Hall [28] and Lewis [46] takes the following form.

Definition A.1. *Let L be a theory of chance and let μ be a prior. We say that μ satisfies the* New Principle *for L if: for any chance law λ of L with chance hypothesis $\Lambda = L^{-1}(\lambda)$ and for any Borel subset A of 2^ω, we have $\mu(A \mid \Lambda) = \lambda(A \mid \Lambda)$.*

[71] To see this, use [61, lemma C.10].

[72] The literature on the Principal Principle largely focuses on this case. There is something odd about that: the New Principle (a modification of the original Principal Principle crafted to handle improper theories of chance) does not do its job in this setting. See footnote 53.

[73] For the transformation $L \mapsto \bar{L}$, see [3].

[74] On this point, see [8]. For critical discussion, see [30].

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press

Let L be an improper theory of chance with countably many chance laws, $\lambda_1, \lambda_2, \ldots$. Let $\bar{L}$ be the corresponding proper theory with laws $\bar{\lambda}_1, \bar{\lambda}_2, \ldots$. Let A be any measurable set. Note that for each k we have $\Lambda_k = \bar{\Lambda}_k$ and $\bar{\lambda}_k(A) = \lambda_k(A \mid \Lambda_k)$. It follows that any prior μ satisfies CCP for the proper theory $\bar{L}$ if and only if it satisfies the New Principle for the improper theory L. So if you think that improper theories make sense, you can understand the New Principle as saying: a prior μ is rationally permitted relative to a (not necessarily proper) theory of chance L if and only if μ satisfies CCP for the corresponding proper theory $\bar{L}$.[75]

Of course, given an improper theory of chance L we can also ask whether there are any computable priors that satisfy the New Principle for it. In the exceptional case (as in Example 2.10) where all of the chance hypotheses of L are basic subsets of 2^ω, there will be a computable prior satisfying the New Principle for L if and only if there is a computable prior satisfying CCP for $\bar{L}$. But in general there need be no such direct relation. The following toy example of a variant of Computable Super-Determinism shows, however, that working with improper theories of chance and the New Principle does not on its own allow one to evade the problems encountered in Section 4.

EXAMPLE A.2. *Fix some distinction between those* 100*-bit sequences that are random-looking* (= *anyone would reasonably suspect them of being generated by a fair coin*) *and others. Consider the theory of chance that only assigns laws of chance to computable sequences, assigning any sequence whose first* 100 *bits are random-looking and which is afterwards all* 0*'s the law of chance* $\nu^\dagger_{.5}$ *that says that the first* 100 *bits are sampled from the fair coin measure and the rest are all* 0*'s; and assigning any other computable sequence* S *the law of chance* δ_S. *Note that* $\nu^\dagger_{.5}$ *considers each sequence that is all* 0*'s after the first hundred bits to be equally likely. So it is just the equal-weighted mixture of the delta-function measures concentrated on those sequences* (*and it is an improper law of chance since, for instance, the all* 0*'s sequence is not included in its chance hypothesis*). *It follows that if* S *is computable then* $\lambda = L(S)$ *can be written in the form* $\lambda = \nu + c \cdot \delta_S$ *for some* $c > 0$ *and some* (*possibly trivial*) *measure* ν. *It then follows that* $\bar{\lambda}(\cdot) := \lambda(\cdot \mid L^{-1}(\lambda))$ *can also be written as a sum in which* δ_S *appears with non-zero weight. From this it follows that if* μ *satisfies the New Principle for* L, *then* μ *can be written as a sum in which each computable delta-function appears with non-zero weight. So* μ *cannot be computable.*

REMARK A.3 (An alternative to the New Principle). *A Chance–Credence constraint suggested by Roberts and by Ismael has been widely discussed as an alternative to the New Principle.*[76] *Specialized to our setting, the Roberts–Ismael constraint says that, relative to a theory of chance* L, *a rational prior* μ *should satisfy* $\mu(A) = \sum \mu(\Lambda_k)\lambda_k(A)$, *where* $\lambda_1, \lambda_2, \ldots$ *are the chance laws of* L, *each* $\Lambda_k = L^{-1}(\lambda_k)$, *and we do* not *assume that the* λ_k *are proper. Applied to the theory of Example* 2.10, *this proposal is implausibly restrictive, requiring that a rational prior must put probability one on one of the theory's two laws of chance.*

§Appendix B. I am not a zero.

CCP embodies a straightforward way of understanding the idea that credence should defer to chance: for a given theory of chance,

[75] One could of course also consider conditions that stand to M-CCP and to WM-CCP as the New Principle stands to CCP.

[76] See [35, 57] (their proposals differ in general, but coincide in our setting). For critical discussion, see [17, sec. 3.3f].

it requires of priors that when conditionalized on the set of worlds corresponding to a law of chance of that theory they give the same probability to each event that the law of chance itself does. The discussion in the main body of this paper took for granted that the conditional probability of A given B is the ratio of two real numbers, the probability of $A\&B$ and the probability of B, and so is undefined when the probability of B is zero. This led to Proposition 3.4—and from there, along the path that led us to consider CCCP, M-CCP, and the rest. Many readers will want off the boat as soon as CCP appears—either because they think it is too intolerant of null sets (see the discussion of CCP* in Section 5) or because they prefer some more sophisticated approach to handling conditional probabilities.

Here I will briefly give my reasons for cleaving to the simple-minded approach followed above. In brief: allowing primitive conditional probabilities to model rational credal states would in any case push us in the direction pursued in Section 5; allowing merely finitely additive probabilities (or non-standard-valued probabilities) to model rational credal states would not help us to evade the delta-function version of the computational obstruction encountered in Section 4; and reformulating CCP in terms of regular conditional probabilities requires us to abandon the connection between learning and deference to chance discussed in Sections 1 and 5.

B.1. What about primitive conditional probabilities? The idea that conditional probability should be taken as primitive rather than unconditional probability has its attractions. Why not just skip all the aggravation and work in a formalism in which the basic probabilistic notion takes two arguments, rather than one?

There are several of ways of working this idea out.[77] Note that under the most plausible ones, we can always conditionalize on logical truths and (when defined) the result of conditionalizing on a proposition is a probability measure. So we always have a surrogate within this framework for our ordinary one-place prior probabilities. So we can think of primitive conditional probabilities as plans of the following form: I will start life with prior μ_0 (an ordinary probability measure); if I face evidence assigned positive probability by μ_0, then I update by conditionalization; if I face evidence corresponding to a null set of μ_0, then I deploy a backup plan (essentially, a fall-back prior that happens to assign the evidence in question probability one).

Fix some fancy gizmos to your liking that encode primitive conditional probabilities. Cook up a version CCP‡ of CCP adapted to such fancy gizmos. Fix a well-behaved theory of chance L with uncountably many laws of chance that is CCP‡-satisfied by a fancy gizmo γ relative to this theory of chance. Let us call an ordinary prior on μ on 2^ω L-*admissible* if it arises by conditionalizing a fancy gizmo satisfying CCP‡ for L on a logical truth. It will count against this approach if there are no L-admissible priors: this would tell us that in order to obey the principle that credence should defer to L-chance, we must assign prior probability zero to some possible finite data sets (and so be willing to bet our lives against observing such data etc.). It would also be bad if the fair coin measure were L-admissible: when it comes to forecasting probabilities of future events, someone with this prior never learns from experience—so it is hard to see how they could be deferring to L-chance.

So, presumably, there must be some interesting difference between priors that are L-admissible and those that are not. And since we are working in a context in which

[77] For some options, consult [20, 29].

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press

possible data sets are finite binary strings and in which all priors assign positive probability to each such string, any agent whose initial credal state is given by a fancy gizmo γ_0 that determines an admissible prior μ_0 is always going to behave exactly like someone whose initial credal state is given by μ_0. Now, we know that μ_0 doesn't satisfy CCP (no prior does). This strongly suggests that it will be more fruitful to search directly for a generalization of CCP that focuses on the response of priors to finite data sets (as M-CCP and WM-CCP do) rather than proceeding indirectly via the apparatus of primitive conditional probabilities.

B.2. What about merely finitely additive priors? Note, first, that merely finitely additive priors defined on σ-algebras are non-constructive objects: the existence of such objects cannot be proven without recourse to some choice-like axiom.[78] So such objects can, presumably, be set aside given our interest in computable priors.

At the same time: any finitely additive probability measure on 2^ω defined only on the algebra of sets $\mathcal{A}$ generated by the basic sets is in fact countably additive (see footnote 8).[79] So in order to be of interest for present purposes, a finitely additive prior would need to be defined on an algebra intermediate between the algebra $\mathcal{A}$ generated by the basic sets and the σ-algebra $\mathcal{B}$ generated by them. Our notion of a computable probability measure on $(2^\omega, \mathcal{B})$ exploits two facts: there is a natural sense in which by computably enumerating the binary strings, we effectively list the basic sets; and the behaviour of a probability measure on the basic sets determines its behaviour on all Borel sets. Presumably, in order for finitely additive probability measures defined on an algebra intermediate between $\mathcal{A}$ and $\mathcal{B}$ to be candidates for computability, we would want to demand something similar: that there be a sense in which we can effectively list a set of generators for this algebra and that any finitely additive probability measure on this algebra is determined by its restriction to these generators.

EXAMPLE B.1. *Let $\mathcal{C}_0$ algebra generated by the basic sets together with $\{0^\omega\}$ (the singleton set containing the all 0's sequence).*[80] *In order to specify a finitely additive probability measure M on $(2^\omega, \mathcal{C})$ it suffices to specify: (i) the probability that M assigns to each basic set; (ii) the probability that M assigns to $\{0^\omega\}$ (freely chosen, subject to the constraint that it does not exceed $R := \inf\{M([\![0^k]\!]) \mid k \in \mathbb{N}\}$).*[81] *So long as these data*

[78] On the one hand, any merely finitely additive probability measure on a measurable space determines a merely finitely additive probability measure on the natural numbers, equipped with the power-set σ-algebra, that considers every finite set a null set—see, e.g., the proof of Theorem 13.5 in [64]. And it follows from results due to Solovay and to Shelah that the existence of such a measure on $\mathbb{N}$ cannot be proven in ZF—for discussion and references see [47]. On the other hand, the existence of merely finitely additive probability measures on $(2^\omega, \mathcal{B})$ can be proven using the axiom of choice (or somewhat weaker assumptions)—see [12, sec. 2.1] and [47].

[79] $\mathcal{A}$ is the family of finitely determined subsets of 2^ω—i.e., the empty set and finite unions of basic sets. See, e.g., [12, theorem 1.1.9(3)].

[80] $\mathcal{C}_0$ consists of the empty set plus finite disjoint unions of sets of the following kinds: the singleton $\{0^\omega\}$, basic sets, and the result of deleting $\{0^\omega\}$ from basic sets. See, e.g., [12, theorem 1.1.9(3)].

[81] Further, every finitely additive probability measure on $(2^\omega, \mathcal{C}_0)$ arises in this way. See, e.g., [12, propositions 3.2.7 and 3.3.1 and theorem 3.3.3].

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press

are computable, we can reasonably consider the finitely additive measure $M : \mathcal{C}_0 \to [0,1]$ to be computable.[82]

More generally, it is clear how to make sense of the notion of a computable but merely finitely additive probability measure on an algebra $\mathcal{C}_k$ whose generators include the basic sets together with k singleton sets of computable sequences.

Let $\mathcal{C}$ be the algebra generated by the basic sets together with the singleton sets of all of the computable sequences.[83] A $\mathcal{C}$*-prior* is a finitely additive probability measure on $(2^\omega, \mathcal{C})$ that assigns positive probability to each basic set. We say that a $\mathcal{C}$-prior M satisfies $\mathcal{C}$-CCP for Computable Frequentism if for each computable sequence S and each $B \in \mathcal{C}$ we have $M(B \mid \{S\}) = \delta_S(B)$. There are two reasons that this will not allow us to evade the obstruction to CCCP-satisfiability that we encountered in Section 4.

1. In order for us to think of a finitely additive measure M on $\mathcal{C}$ as being computable, we need to think of there being an effective listing of its generators, the basic sets and the singletons of computable sequences. This would appear to require that the computable sequences be uniformly computable—which they are not (see Remark 4.14).
2. Let M be a $\mathcal{C}$-prior on $(2^\omega, \mathcal{C})$ and let μ be the unique prior on $(2^\omega, \mathcal{B})$ that whose restriction to basic sets coincides with that of M. Note that if $M(\{S\}) > 0$ for some sequence S, then $\mu(\{S\}) > 0$ as well.[84] So if M satisfies $\mathcal{C}$-CCP for Computable Frequentism, then $\mu(\{S\}) > 0$ for each computable S—which means that μ cannot be computable. So M cannot be computable even in the weak sense that its restriction to the basic sets is computable in the usual sense.

B.3. What about priors taking non-standard values? Many philosophers follow Lewis [45] in using as representors of rational credal states generalized probability measures taking values in extensions of the unit interval that include infinitesimals. The real part of such an object is a finitely additive probability measure. So the considerations adduced above against the helpfulness of merely finitely additive priors carry over the case of non-standard-valued priors.

B.4. What about regular conditional probabilities? In the following, 2^ω always carries the product topology (generated by the basic subsets), any subspace of 2^ω always carries the corresponding subspace topology, the unit interval always carries its standard topology (generated by its open subintervals), and $\mathcal{P}$ always carries the weak topology.[85] Unless otherwise noted, when considered as measurable spaces, we take these spaces to be equipped with the Borel σ-algebra generated by their open sets.[86]

[82] Let μ be the unique (countably additive) probability measure on $(2^\omega, \mathcal{B})$ that agrees with M when restricted to basic sets. Then $R = \mu(\{0^\omega\})$ and M is the restriction to $\mathcal{C}_0$ of μ if and only if $M(\{0^\omega\}) = R$.

[83] A finitely additive probability measure on $\mathcal{C}$ is determined by its behaviour on sets of the following kinds: basic sets, finite unions of singleton sets of computable sequences, and basic sets with such finite unions of singletons deleted. And $\mathcal{C}$ consists of finite disjoint unions of sets of these kinds (plus the empty set). See, e.g., [12, theorems 1.1.9(3) and 3.5.1(ii)].

[84] See [12, propositions 3.2.7 and 3.3.1].

[85] Recall that the weak topology on $\mathcal{P}$ can be characterized as follows: for any $p, q \in \mathbb{Q}$ with $p < q$ and any basic set $[\![\tau]\!]$ of 2^ω we let $S_{p,q,\tau} := \{\nu \in \mathcal{P} \mid p < \nu([\![\tau]\!]) < q\}$; the open sets of the weak topology are the unions of finite intersections the $S_{p,q,\tau}$. For details, see [40, sec. 17E].

[86] For case of $\mathcal{P}$, again see [40, sec. 17E].

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press

For any theory of chance L, we denote by Λ_+ the set of sequences to which L assigns chance laws and by Λ_* the set of lawless sequences. We have all along required that each chance hypothesis of a theory of chance be Borel. The following is a natural strengthening of that requirement.

Definition B.2. *A theory of chance L is* measurable *if it is measurable as a map from* Λ_+ *to* $\mathcal{P}$.[87]

Our five paradigmatic reductionist theories of chance are measurable. This is more or less immediate for Super-Determinism, Computable Super-Determinism, Computable Frequentism, and the Toy Best System.[88] And the measurability of the Basic Frequentism is implicit in standard proofs of the de Finetti Representation Theorem.[89]

Remark B.3. 2^ω *and* $\mathcal{P}$ *can both be viewed as computable metric spaces in a natural way.*[90] *So it makes sense to ask whether an acceptable theory of chance should be computable as well as measurable. There is a powerful intuitive reason for resisting the idea that they should be: even computable frequentism is not computable as a map from* 2^ω *to* $\mathcal{P}$.[91] *However, computable frequentism satisfies the weaker condition of layerwise computability.*[92] *So it would perhaps be reasonable to require an acceptable theory of chance to be layerwise computable.*[93]

Definition B.4. *Let L be a measurable theory of chance and let* $\mathcal{L}$ *be the set of chance hypotheses of L. The* kernel *of L is the map* $p_L(\cdot \parallel \cdot) : \mathcal{B} \times \mathcal{L} \to [0,1]$ *defined by*

$$p_L(B \parallel \Lambda) = \lambda(B)$$

(*where* λ *is the chance law corresponding to* Λ).[94]

[87] When considering a theory of chance L, it is sometimes helpful to consider the σ-algebra $\mathcal{M}$ on 2^ω generated by Λ_* together with the chance hypotheses of L. Note that if L is measurable, then $\mathcal{M}$ is a sub-σ-algebra of $\mathcal{B}$.

[88] For the case of Super-Determinism, see, e.g., [39, lemma 3.1]. For the others, note that for any theory of chance L and any Borel subset D of $\mathcal{P}$, $L^{-1}(D)$ will be a union of chance hypotheses of L. When L has only countably many chance laws, each such union is Borel (being a countable union of Borel sets).

[89] See, e.g., [43, sec. 12.3]—thanks to Chris Mierzewski for this labour-saving observation. For a direct proof of the measurability of Basic Frequentism note that by [39, lemma 3.1.iii] it suffices to show that for any fixed string τ and any closed interval J with rational endpoints, the set of sequences S such that $\nu_{r(S)}([\![\tau]\!]) \in J$ is Borel (where $r(S)$ is the limiting relative frequency of 0's in S, if defined). This will just be the set of sequences in which the limiting relative frequency of 0's lies in some closed interval, which is Borel—see, e.g., [40, p. 70].

[90] See, e.g., [25, Appendix B].

[91] The problem is that because for each r, the set of sequences in which 0's have relative frequency r forms a dense subset of 2^ω, there is no way to force the relative frequency of 0's in $S\upharpoonright n$ to be close to the limiting relative frequency of 0's in S by choosing n sufficiently large. On this point see, e.g., [13, sec. IV].

[92] See, e.g., [13, sec. IV]. For further discussion, see [33, sec. 3].

[93] Thanks to Chris Mierzewski and to Francesca Zaffora Blando for the discussion of these points.

[94] Let $\mathcal{M}$ be as in footnote 87 and let S^* be some L-lawless sequence. Define $\kappa : \mathcal{B} \times 2^\omega \to [0,1]$ by $\kappa(B,S) := p_L(B \parallel L^{-1}(L(S)))$ (unless S is lawless, in which case set $\kappa(B,S) = \delta_{S^*}(B)$). Then, for any sequence S, $\kappa(\cdot, S) \in \mathcal{P}$ (by the definition of p_L) and for any $B \in \mathcal{B}$, $\kappa(B, \cdot)$

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press

Definition B.5. *Let L be a proper and measurable theory of chance with kernel p_L. Let μ be a prior that considers the set of lawless sequences of L to be a null set. Then p_L is a* proper regular conditional probability *for μ if for each $B \in \mathcal{B}$ we have*

$$\mu(B) = \int p(B \parallel \Lambda(S))\ d\mu(S)$$

(*where we write $\Lambda(S)$ for $L^{-1}(L(S))$*).[95]

Example B.6. *Suppose that L is a proper theory of chance with only countably many chance laws λ_1, λ_2 ... (with chance hypotheses Λ_1, Λ_2, ...). Then L is measurable (see footnote 88). If μ is a prior that assigns positive probability to each Λ_k and probability zero to the lawless sequences of L, then p_L is a proper regular conditional probability for μ if and only if, for each $B \in \mathcal{B}$,*

$$\mu(B) = \sum p_L(B \parallel \Lambda_k)\mu(\Lambda_k),$$

which happens if and only if $p_L(B \parallel \Lambda_k) = \mu(B \bigcap \Lambda_k)/\mu(\Lambda_k)$.

The core idea behind the Principal Principle, codified in the Chance–Credence Principle, is that, conditional on knowing which law of chance is actual, your credences in events should coincide with their actual chances. In the setting of a proper theory of chance with only countably many chance laws this says:

> Adopt any prior you like, so long as the probability measure that results from conditionalizing it on a chance hypothesis is the corresponding chance law.

The preceding example shows that one way to generalize this advice to the setting of uncountably many chance laws is to advise:

> Adopt any prior you like, so long as the kernel for the theory of chance under consideration is a proper regular conditional probability for it.

Think of this kernel as a plan: if you find out which chance hypothesis obtains, then you will adopt credences given by the corresponding chance law; requiring that the kernel be a proper regular conditional probability is a way of constraining priors so that this plan can be thought of arising naturally out the priors rather than being wholly arbitrary.

is measurable as a function from $(2^\omega, \mathcal{M})$ to the unit interval (by the measurability of L, via Lemma 3.1 of [39]). So p_L determines a probability kernel from $(2^\omega, \mathcal{M})$ to $(2^\omega, \mathcal{B})$. If Λ_* is empty, then this probability kernel is proper, in the sense that for each $M \in \mathcal{M}$ we have $\kappa(M, S) = 1$ if $S \in M$. Otherwise, it is proper modulo Λ_* (in the sense that the given condition holds for any M disjoint from Λ_*).

[95] For a helpful philosophical introduction to regular conditional probabilities, see [20]. Blackwell and Dubins [15, sec. 1] observe that if $(X, \mathcal{C}, \nu)$ is a probability space and $\mathcal{D}$ is a sub-σ-algebra of $\mathcal{C}$, then a proper probability kernel $\kappa : \mathcal{C} \times X \to [0, 1]$ from $(X, \mathcal{D})$ to $(X, \mathcal{C})$ is a regular conditional probability for ν (in the standard sense) if and only if $\mu(\cdot) = \int \kappa(\cdot, S)\ d\nu(S)$. Here we specialize to the case $(X, \mathcal{C}, \nu) = (2^\omega, \mathcal{B}, \mu)$ and $\mathcal{D} = \mathcal{M}$ (with $\mathcal{M}$ as in footnote 87). Since Λ_* is by assumption a μ-null set, the distinction of footnote 94 between being proper and being proper modulo Λ_* is immaterial.

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press

Definition B.7 (CCP**). *Let L be a proper and measurable theory of chance and let μ be a prior that considers the set of L-lawless sequences to be null set. We say that μ satisfies* CCP** *for L if the kernel p_L of L is a regular conditional probability for μ.*

Good news: the Laplace–Bayes prior of Example 5.23 satisfies CCP** for Basic Frequentism.[96] Bad news:

Proposition B.8. *Let L be a proper and measurable theory of chance. Then any law of chance of L that is also a prior satisfies CCP** for L.*

Proof. Let λ be a prior that is also law of chance of L (with chance hypothesis Λ). Let B be a Borel subset of 2^ω. Then we have

$$\begin{aligned}\int p_L(B \parallel \Lambda(s))\, d\lambda(s) &= \int_\Lambda p_L(B \parallel \Lambda(s))\, d\lambda(s)\\ &= \int_\Lambda p_L(B \parallel \Lambda)\, d\lambda(s)\\ &= \int_\Lambda \lambda(B)\, d\lambda(s)\\ &= \lambda(B)\int_\Lambda d\lambda(s)\\ &= \lambda(B),\end{aligned}$$

where the first equality follows because L is proper (so that $\lambda(\Lambda) = 1$), the second because $\Lambda(s) = \Lambda$ for each $S \in \Lambda$, the third by the definition of the kernel p_L, the fourth because $\lambda(B)$ doesn't depend on the variable of integration, and the fifth again via propriety. □

Corollary B.9. *The fair coin measure satisfies CCP** for Basic Frequentism.*

Example B.10 (Super-Determinism). *Also worrying*: every *prior satisfies CCP** for Super-Determinism (i.e., the map $L : S \mapsto \delta_S$). For let μ be a prior and let B be Borel subset of* 2^ω*. For any Borel set B and any binary sequence S, we have*

$$p_L(B \parallel \Lambda(S)) = I_B,$$

where I_B is the characteristic function of B. So

$$\begin{aligned}\int p_L(B \parallel S)\, d\mu(S) &= \int I_B(S)\, d\mu(S)\\ &= \mu(B).\end{aligned}$$

(*Super-Determinism is of course a pathological theory—but it would be more considerably more reassuring to be told that* no *prior is rational for someone who hopes to defer to this theory of chance.*)

Acknowledgments. Previous versions of the paper were presented at NYU, CMU, and PSA 2022. For helpful correspondence and suggestions, thanks very much to Anonymous, Dave Baker, Cian Dorr, John Earman, Dmitri Gallow, Kevin Kelly, Alex

[96] On this point see, e.g., [43, sec. 8.3].

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press

Meehan, Richard Pettigrew, Laura Ruetsche, Tom Sterkenburg, Francesca Zaffora Blando, Snow Zhang, and, especially, Chris Mierzewski and Chris Porter.

REFERENCES

[1] Adleman, L., & Blum, M. (1991). Inductive inference and unsolvability. *Journal of Symbolic Logic*, **56**, 891–900.

[2] Angluin, D., & Smith, C. (1983). Inductive inference: Theory and methods. *ACM Computing Surveys*, **15**, 237–269.

[3] Arntzenius, F., & Hall, N. (2003). On what we know about chance. *British Journal for the Philosophy of Science*, **54**, 171–179.

[4] Barmpalias, G. (2020). Aspects of Chaitin's omega. In Franklin, J. and Porter, C., editors. *Algorithmic Randomness: Progress and Prospects*. Cambridge: Cambridge University Press, pp. 175–205.

[5] Barmpalias, G., Fang, N., & Stephan, F. (2018). Equivalences between learning of data and probability distributions and their applications. *Information and Computation*, **262**, 132–140.

[6] Bārzdiņš, J. (1972) Prognostication of automata and functions. In Freiman, C., editor. *Information Processing '71*, Vol. 1. Amsterdam: North-Holland, pp. 81–84.

[7] Beebee, H., & A. Fisher, editors. (2021). *Philosophical Letters of David Lewis*, Vol. 2. Oxford: Oxford University Press.

[8] Belot, G. (2016). Undermined. *Australasian Journal of Philosophy*, **94**, 781–791.

[9] ———. (2017). Curve-fitting for Bayesians? *British Journal for the Philosophy of Science*, **68**, 689–702.

[10] ———. (2020). Absolutely no free lunches! *Theoretical Computer Science*, **845**, 159–180.

[11] ———. (2023). That does not compute: David Lewis on credence and chance. *Philosophy of Science*, forthcoming.

[12] Bhaskara Rao, K. P. S., & Bhaskara Rao, M. (1983). *Theory of Charges: A Study of Finitely Additive Measures*. Cambridge, MA: Academic Press.

[13] Bienvenu, L., & Monin, B. (2012). Von Neumann's biased coin revisited. In Dershowitz, N., editor. *Proceedings of the 2012 27th Annual ACM/IEEE Symposium on Logic in Computer Science*. Los Alamitos, CA: Association for Computing Machines.

[14] Blackwell, D., & Dubins, L. (1962). Merging of opinion with increasing information. *Annals of Mathematical Statistics*, **33**, 882–886.

[15] ———. (1975). On existence and non-existence of proper, regular, conditional distributions. *Annals of Probability*, **3**(1975), 741–752.

[16] Blum, L., & Blum, M. (1975). Toward a mathematical theory of inductive inference. *Information and Control*, **28**, 125–155.

[17] Briggs, R. (2009). The anatomy of the big bad bug. *Noûs*, **43**, 428–449.

[18] Downey, R., & Hirschfeldt, D. (2010). *Algorithmic Randomness and Complexity*. New York: Springer.

[19] Earman, J. (1992). *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. Cambridge: MIT Press.

[20] Easwaran, K. (2019). Conditional probabilities. In Pettigrew, R. and Weisberg, J., editors. *The Open Handbook of Formal Epistemology*. London: PhilPapers Foundation, pp. 131–198.

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press

[21] Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, Vol. I (third edition). Hoboken: Wiley.

[22] Fortnow, L., Freivalds, R., Gasarch, W., Kummer, M., Kurtz, S., Smith, C., & Stephan, F. (1998). On the relative sizes of learnable sets. *Theoretical Computer Science*, **197**, 139–156.

[23] Freedman, D. (1963). Asymptotic behavior of Bayes' estimates in the discrete case. *Annals of Mathematical Statistics*, **34**, 1386–1403.

[24] ———. (1965). Bernard Friedman's urn. *Annals of Mathematical Statistics*, **36**, 956–970.

[25] Gács, P. (2021). Lecture notes on descriptional complexity and randomness. Preprint, arXiv:2105.04704v1.

[26] Gibbs, A., & Su, F. (2002). On choosing and bounding probability metrics. *International Statistical Review*, **70**, 419–435.

[27] Hacking, I. (1965). *Logic of Statistical Inference*. Cambridge: Cambridge University Press.

[28] Hall, N. (1994). Correcting the guide to objective chance. *Mind*, **103**, 505–517.

[29] Halpern, J. (2010). Lexicographic probability, conditional probability, and nonstandard probability. *Games and Economic Behavior*, **68**, 155–179.

[30] Hoefer, C. (2018). "Undermined" Undermined. Unpublished. pittphilsci: 14886.

[31] ———. (2019). *Chance in the World: A Humean Guide to Objective Chance*. Oxford: Oxford University Press.

[32] Hopcroft, J., & Ullman, J. (1979). *Introduction to Automata Theory, Languages, and Computation*. Boston: Addison-Wesley.

[33] Hoyrup, M. (2020). Algorithmic randomness and layerwise computability. In Franklin, J. and Porter, C., editors. *Algorithmic Randomness: Progress and Prospects*. Cambridge: Cambridge University Press, pp. 115–133.

[34] Hoyrup, M., & Rute, J. (2021). Computable measure theory and algorithmic randomness. In Brattka, V. and Hertling, P., editors. *Handbook of Computability and Complexity in Analysis*. Cham: Springer, pp. 227–270.

[35] Ismael, J. (2008). Raid! Dissolving the big, bad bug. *Noûs*, **42**, 292–307.

[36] Jain, S., Osherson, D., Royer, J., & Sharma, A. (1999). *Systems that Learn* (second edition). Cambridge: MIT Press.

[37] Kalai, E., & Lehrer, E. (1993). Rational learning leads to Nash equilibrium. *Econometrica*, **61**, 1019–1045.

[38] ———. (1994). Weak and strong merging of opinions. *Journal of Mathematical Economics*, **23**, 73–86.

[39] Kallenberg, O. (2021). *Foundations of Modern Probability* (third edition). Cham: Springer.

[40] Kechris, A. (1995). *Classical Descriptive Set Theory*. New York: Springer.

[41] Kelly, K. (1996). *The Logic of Reliable Inquiry*. Oxford: Oxford University Press.

[42] Kjos-Hanssen, B. (2010). The probability distribution as a computational resource for randomness testing. *Journal of Logic and Analysis*, **10**, 1–13.

[43] Klenke, A. (2020). *Probability Theory: A Comprehensive Course* (third edition). Cham: Springer.

[44] Lehrer, E., & Smorodinsky, R. (1996). Merging and learning. In Ferguson, T., Shapley, L., and MacQueen, J., editors. *Statistics, Probability, and Game Theory:*

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press

Papers in Honor of David Blackwell. Hayward, CA: Institute of Mathematical Statistics, pp. 147–168.

[45] Lewis, D. (1980). A subjectivist's guide to objective chance. In Jeffrey, R., editor. *Studies in Inductive Logic and Probability* II. Berkeley: University of California Press, pp. 263–294; reprinted with additional post-scripts in Lewis, D. (1986). *Philosophical Papers* II. Oxford: Oxford University Press, pp. 83–132.

[46] ———. (1994). Humean supervenience debugged. *Mind*, **103**, 473–490.

[47] Luxemburg, W., & Väth, M. (2001). The existence of non-trivial bounded functionals implies the Hahn–Banach extension theorem. *Zeitschrift für Analysis und ihre Anwendungen*, **20**, 267–279.

[48] Martin-Löf, P. (1966). The definition of random sequences. *Information and Control*, **9**, 602–619.

[49] Miller, D. (1966). A paradox of information. *British Journal for the Philosophy of Science*, **17**, 59–61.

[50] Odifreddi, P. (1999). *Classical Recursion Theory II*. Amsterdam: North-Holland.

[51] Pettigrew, R. (2012). Accuracy, chance, and the principal principle. *Philosophical Review*, **121**, 241–275.

[52] Porter, C. (2019). Effective aspects of Bernoulli randomness. *Journal of Logic and Computation*, **29**, 933–964.

[53] Porter, M., Day, A., & Downey, R. (2017). Notes on computable analysis. *Theory of Computing Systems*, **60**, 53–111.

[54] Putnam, H. (1963a). 'Degree of confirmation' and inductive logic. In Schilpp, P., editor. *The Philosophy of Rudolf Carnap*. Chicago: Open Court, pp. 761–783.

[55] ———. (1963b). *Probability and Confirmation*. Washington: United States Information Agency.

[56] Reimann, J. (2009). Randomness—Beyond Lebesgue measure. In Cooper, B., Geuvers, H., Pillay, A., and Väänänen, J., editors. *Logic Colloquium '06*. Cambridge: Cambridge University Press, pp. 247–279.

[57] Roberts, J. (2001). Undermining undermined: Why Humean supervenience never needed to be debugged (even if it's a necessary truth). *Philosophy of Science*, **68**, S98–S108.

[58] Ryabko, D. (2010). On finding predictors for arbitrary families of processes. *Journal of Machine Learning Research*, **11**, 581–602.

[59] Ryabko, D., & Hutter, M. (2007). On sequence prediction for arbitrary measures. In Goldsmith, A., Medard, M., Shokrollahi, A., and Zamir, R., editors. *Proceedings of the 2007 IEEE International Symposium on Information Theory*. Piscataway, NJ: The Institute of Electrical and Electronics Engineers, pp. 2346–2350.

[60] Schwarz, W. (2014). Best system approaches to chance. In Hájek, A. and Hitchcock, C., editors. *The Oxford Handbook of Probability and Philosophy*. Oxford: Oxford University Press, pp. 423–439.

[61] Titelbaum, M. (2013). *Quitting Certainties: A Bayesian Framework for Modeling Degrees of Belief*. Oxford: Oxford University Press.

[62] Turing, A. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, **42**, 230–265.

[63] Vitányi, P., & Chater, N. (2017). Identification of probabilities. *Journal of Mathematical Psychology*, **76**, 13–24.

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press

[64] Wagon, S. (1985). *The Banach–Tarski Paradox*. Cambridge: Cambridge University Press.

[65] Zabell, S. (2009). Philosophy of inductive logic: The Bayesian perspective. In Haaparanta, L., editor. *The Development of Modern Logic*. Oxford: Oxford University Press, pp. 724–774.

DEPARTMENT OF PHILOSOPHY
UNIVERSITY OF MICHIGAN
435 SOUTH STATE STREET
2215 ANGELL HALL
ANN ARBOR, MI 48104, USA
E-mail: belot@umich.edu

https://doi.org/10.1017/S1755020323000151 Published online by Cambridge University Press