DOI: 10.1017/psa.2023.177

This is a manuscript accepted for publication in Philosophy of Science.

This version may be subject to change during the production process.

Pride and Probability

Francesca Zaffora Blando

Carnegie Mellon University, Pittsburgh, PA, USA
Email: fzaffora@andrew.cmu.edu

Abstract
Bayesian agents, argues Belot (2013), are orgulous: they believe in inductive success even when guaranteed to fail on a topologically typical collection of data streams. Here we shed light on how pervasive this phenomenon is. We identify several classes of inductive problems for which Bayesian convergence to the truth is topologically typical. However, we also show that, for all sufficiently complex classes, there are inductive problems for which convergence is topologically atypical. Lastly, we identify specific topologically typical collections of data streams, observing which guarantees convergence to the truth across all problems from certain natural classes of effective inductive problems.

1. Introduction

Convergence-to-the-truth theorems are a staple of Bayesian epistemology: their use in philosophy, especially in debates concerning the tenability of subjective Bayesianism, dates back to the work of Savage (1954) and Edwards et al. (1963). In a nutshell, these results establish that, in a wide array of learning scenarios, Bayesian agents expect their future credences to almost surely align with the truth as the evidence accumulates.

Rather than seeing convergence-to-the-truth results as an asset of the Bayesian framework, a number of authors take them to be the Achilles heel of Bayesianism.[1] Most recently, Belot (2013) argued that, because of these theorems, Bayesian reasoners are plagued by a pernicious type of epistemic immodesty. By the very nature of the Bayesian framework, Bayesian agents are barred from acknowledging that, for certain learning problems, failure, rather than success, is the *typical* outcome of inquiry—where, crucially, the notion of typicality that Belot's argument relies on is topological, rather than measure-theoretic (or probabilistic). There are learning problems for which a Bayesian agent's success set (the collection of data streams along which convergence to the truth occurs) is topologically atypical or "small"; yet, as a consequence of said convergence-to-the-truth results, the agent must nonetheless assign probability one to this set.[2]

[1] See (Glymour, 1980), (Earman, 1992), (Kelly, 1996), and (Belot, 2013).

[2] Kelly (1996) voices an analogous worry. His argument relies on cardinality, rather than topological considerations. In particular, he points out that there are learning situations where, even though the collection of data streams along which convergence to the truth occurs has probability one, the collection of data streams

It is well-known that measure-theoretic and topological typicality often come apart.[3] Belot's goal is to draw attention to the fact that this dichotomy occurs in contexts that he takes to be especially problematic due to their "obvious epistemological interest" (Belot, 2013, 499, footnote 43): for one, in situations where (i) the event witnessing the coming apart of these two types of typicality is a Bayesian agent's success set, and (ii) the agent's prior is one that, in Belot's view, otherwise displays a desirable type of open-mindedness. In such settings, the argument goes, this dichotomy is particularly alarming, because topological typicality is an objective notion—one that does not depend on any particular agent or their subjective degrees of belief—while the measure-theoretic notion of typicality, in this context, reflects a particular agent's opinion and is therefore subjective. These considerations lead Belot to conclude that Bayesian agents suffer from an irrational over-confidence in their ability to be inductively successful.

This objection has received considerable attention in the literature,[4] and many of the available responses recommend substantial modifications of the Bayesian framework in order to evade Belot's conclusion. For instance, Huttegger (2015) proposes to use *metric Boolean algebras*, which allow to avoid drawing distinctions between events that can only be made by infinitely many observations, Weatherson (2015) advocates passing to *imprecise Bayesianism*, while Elga (2016) and Nielsen and Stewart (2019) suggest dropping countable additivity in favour of finite additivity.

The goal of this article is not to further examine the merits or shortcomings of Belot's argument; rather, our aim is to shed light on how pervasive the phenomenon identified by Belot is by clarifying the conditions under which his objection does not apply—the conditions under which inductive success for a Bayesian agent is both probabilistically and topologically typical—and the conditions under which it does. To address this question, we will not depart from standard Bayesian lore: instead, comfortably situated within the Bayesian framework, we will consider a taxonomy of inductive problems, in the spirit of Kelly (1996), that will help us differentiate between the learning situations in which convergence to the truth is topologically typical and those in which it is not. We will focus on a canonical convergence-to-the-truth result—*Lévy's Upward Theorem* (Lévy, 1937)—and show that, by categorizing the random variables featuring in this result (the functions used to model the inductive problems faced by Bayesian agents) in terms of their descriptive complexity and computability-theoretic strength, we can gain a deeper and sharper understanding of when topological and probabilistic typicality agree or disagree in this setting.[5]

Continuity will play a crucial role in our investigation. We will show that, for several classes of random variables that are "sufficiently close" to being continuous and

along which convergence to the truth instead fails is uncountable. Kelly locates the culprit of Bayesian immodesty in the axiom of countable additivity.

[3] See (Oxtoby, 1980).

[4] See (Huttegger, 2015), (Weatherson, 2015), (Elga, 2016), (Belot, 2017), (Cisewski et al., 2018), (Pomatto and Sandroni, 2018), (Nielsen and Stewart, 2019), and (Gong et al., 2021).

[5] The computability-theoretic approach advocated in this paper is in line with (Huttegger et al., 2023) (see also (Zaffora Blando, 2020)), where Lévy's Upward Theorem is studied through the lens of computability theory and the theory of algorithmic randomness—a branch of computability theory on which we rely here, as well. The present work may also be seen as a Bayesian counterpart to work in formal learning theory. As mentioned above, see, in particular, (Kelly, 1996).

admit natural epistemic interpretations, convergence to the truth is indeed a topologically typical affair. We will also see, however, that, for all sufficiently complex classes of random variables, there are inductive problems for which convergence to the truth is instead topologically atypical. Even though topologically typical inductive success is guaranteed for several natural classes of inductive problems, "Bayesian orgulity"—as Belot calls it—is, in this sense at least, a pervasive phenomenon. Lastly, by bringing computability theory into the picture, we will identify several classes of effective inductive problems and specific topologically typical collections of data streams along which convergence to the truth is guaranteed to occur, no matter which inductive problem from those classes the agent is trying to solve. This will allow us to throw light on the kind of properties of data streams that are conducive to topologically typical inductive learning.

2. Lévy's Upward Theorem, Typicality, and Bayesian Immodesty

In keeping with much of the Bayesian epistemology literature on the topic,[6] our discussion of Bayesian convergence to the truth will focus on the setting of infinite binary sequences—i.e., the setting of *Cantor space*: the topological space whereby the set $\{0,1\}^{\mathbb{N}}$ of infinite binary sequences is endowed with the *topology of pointwise convergence*. This is the topology generated by the collection of *cylinders* $[\sigma]$, where $\sigma \in \{0,1\}^{<\mathbb{N}}$ is a finite binary string and $[\sigma] = \{\omega \in \{0,1\}^{\mathbb{N}} : \sigma \sqsubset \omega\}$ is the set of all sequences that begin with σ ("$\sigma \sqsubset \omega$" indicates that σ is a proper initial segment of ω). Every open set in Cantor space can be expressed as a countable union of cylinders and every clopen set as a finite union of cylinders. We will think of infinite binary sequences as data streams, sequences of experimental outcomes, environments, or possible worlds. From this viewpoint, cylinders intuitively encapsulate the information available to an agent after having made finitely many observations or having performed an imprecise measurement with a certain degree of precision.

Measure-theoretic vs. Topological Typicality. One prominent way to think about typicality—the one that Bayesian convergence-to-the-truth theorems capitalize on—is measure-theoretic. Recall that the *Borel σ-algebra* $\mathfrak{B}$ on $\{0,1\}^{\mathbb{N}}$ is the smallest σ-algebra containing all open sets in Cantor space. The elements of $\mathfrak{B}$ are called *Borel sets*. A probability measure μ on $\mathfrak{B}$ assigns to each Borel subset of $\{0,1\}^{\mathbb{N}}$ a value in $[0,1]$ in such a way that, for any countable collection $\{\mathscr{A}_n\}_{n\in\mathbb{N}}$ of pairwise disjoint Borel sets, $\mu(\bigcup_{n\in\mathbb{N}} \mathscr{A}_n) = \sum_{n\in\mathbb{N}} \mu(\mathscr{A}_n)$. Every probability measure on $\mathfrak{B}$ can be identified with a function μ that maps cylinders to real numbers in $[0,1]$ and satisfies the following two conditions: (i) $\mu([\varepsilon]) = 1$ (where ε denotes the empty string) and (ii) $\mu([\sigma]) = \mu([\sigma 1]) + \mu([\sigma 0])$ for all $\sigma \in \{0,1\}^{<\mathbb{N}}$ (where $\sigma 1$ is the string consisting of σ followed by 1 and $\sigma 0$ the string consisting of σ followed by 0). By Carathéodory's Extension Theorem,[7] any such function can in fact be uniquely extended to a full probability measure on $\mathfrak{B}$.

A probability measure that we will often make use of is the *uniform measure* λ: the probability measure that results from tossing a fair coin infinitely many times, given by $\lambda([\sigma]) = 2^{-|\sigma|}$ for all $\sigma \in \{0,1\}^{<\mathbb{N}}$ (where $|\sigma|$ denotes the length of σ).

[6] See (Earman, 1992), (Belot, 2013), and (Huttegger, 2015, 2017).
[7] See (Williams, 1991, Theorem 1.7, 20).

https://doi.org/10.1017/psa.2023.177 Published online by Cambridge University Press

While probability measures admit various interpretations, here we will take them to represent the subjective priors of Bayesian reasoners. So, a probability measure μ on $\mathfrak{B}$ will always stand for a specific agent's initial degrees of belief, or credences, about the events in $\mathfrak{B}$, and it will be understood as capturing that agent's background knowledge and inductive assumptions at the beginning of the learning process.

Given a probability measure μ, a set is measure-theoretically typical relative to μ if it has μ-measure one and measure-theoretically atypical relative to μ if it has μ-measure zero. A μ-measure-one set corresponds to an event of which the agent with prior μ is essentially certain, while a μ-measure-zero set corresponds to an event that the agent considers negligible. Measure-theoretic typicality is therefore inextricably tied to the underlying prior: distinct probability measures may disagree wildly as to which sets count as measure-theoretically typical.[8]

Topological typicality (the type of typicality that Belot's criticism is grounded on) is instead defined as follows. Recall that a set is *nowhere dense* if its closure has empty interior—intuitively, if it corresponds to a hypothesis that, no matter what evidence has been observed so far, can always be refuted by further evidence. Equivalently, a set $\mathscr{S} \subseteq \{0, 1\}^{\mathbb{N}}$ is nowhere dense if, for every open set $\mathscr{U} \subseteq \{0, 1\}^{\mathbb{N}}$, $\mathscr{S} \cap \mathscr{U}$ is not *dense* in the subspace topology on $\mathscr{U}$—where a set is dense if it has a non-empty intersection with every non-empty open set (intuitively, a dense set corresponds to a hypothesis that cannot be refuted by any finite amount of evidence). A subset of a topological space is topologically atypical if it is *meagre*: i.e., if it is expressible as a countable union of nowhere dense sets.[9] On the other hand, a set is topologically typical if it is *co-meagre*: if it is the complement of a meagre set.

Measure-theoretic and topological (a)typicality have several features in common. For instance, the class of measure-zero sets and the class of meagre sets are both σ-*ideals*: measure-zero and meagre sets are both closed under subsets and countable unions. However, while they both aim at capturing notions of "largeness" and "smallness", these concepts often diverge. For a well-trodden example, consider the collection of sequences that satisfy the Strong Law of Large Numbers relative to the uniform measure λ (the set

[8] For a simple example, take the uniform measure λ and the collection of sequences that satisfy the Strong Law of Large Numbers relative to λ: namely, the set of sequences along which the relative frequency of 1 converges to $\frac{1}{2}$ in the limit. This is a set with λ-measure one and is therefore measure-theoretically typical relative to λ. Its complement—the set of sequences that fail to satisfy the Strong Law of Large Numbers relative to λ—has λ-measure zero and is therefore measure-theoretically atypical relative to λ. Yet, if one takes a Bernoulli measure other than λ, the situation changes drastically. Consider, for instance, the probability measure β given by $\beta([\sigma]) = \frac{1}{3}^{\#1(\sigma)} \frac{2}{3}^{\#0(\sigma)}$ for all $\sigma \in \{0, 1\}^{<\mathbb{N}}$, where $\#1(\sigma)$ denotes the number of 1's occurring in σ and $\#0(\sigma)$ the number of 0's occurring in σ. According to β, the set of sequences that satisfy the Strong Law of Large Numbers relative to λ not only fails to be measure-theoretically typical, it is measure-theoretically atypical, since β assigns probability one to the set of sequences along which the relative frequency of 1 converges to $\frac{1}{3}$ in the limit.

[9] For instance, every singleton set $\{\omega\}$ is nowhere dense, so every countable set is meagre. But nowhere dense sets—and, a fortiori, meagre sets—can also be uncountable. Consider the set $\mathscr{S} = \{\omega \in \{0, 1\}^{\mathbb{N}} : (\forall n)\ \omega(2n+1) = \omega(2n)\}$: i.e., the set of all sequences whose odd entries agree with the preceding even entry (where the enumeration starts at 0). This is an uncountable set and, yet, it is nowhere dense. To see this, let $[\sigma]$ be an arbitrary cylinder. Let τ be a string of even length that extends σ such that τ's last entry is 1, while its penultimate entry is 0. Then, $\mathscr{S} \cap [\tau] = \emptyset$, which suffices to conclude that $\mathscr{S}$ is nowhere dense.

https://doi.org/10.1017/psa.2023.177 Published online by Cambridge University Press

of sequences with limiting relative frequency $\frac{1}{2}$ for 1): this is a set with λ-measure one and, yet, it is also meagre.[10]

Lévy's Upward Theorem. Bayesian convergence to the truth is epitomized by *Lévy's Upward Theorem* (1937), which establishes that, given some quantity that a Bayesian agent is trying to measure, the probability of observing a data stream that will lead the agent's successive estimates to asymptotically align with the truth is one. In other words, a Bayesian reasoner conducting repeated experiments to gauge some quantity expects that almost every sequence of observations will bring about inductive success.

Let $f : \{0,1\}^{\mathbb{N}} \to \mathbb{R}$ be a random variable relative to some probability measure μ on $\mathfrak{B}$. Think of the values of f as quantities that the agent with prior μ wishes to estimate: for instance, f could record the value of some unknown physical parameter which may vary between possible worlds. The unconditional expectation of f with respect to μ (the average value of f weighted by μ, given by $\int_{\{0,1\}^{\mathbb{N}}} f\, d\mu$) is abbreviated as $\mathbb{E}_{\mu}[f]$. If $\mathbb{E}_{\mu}[|f|] < \infty$, then f is *integrable*. For each $n \in \mathbb{N}$, let $\mathfrak{F}_n$ be the sub-σ-algebra of $\mathfrak{B}$ generated by the cylinders $[\sigma]$ centred on strings $\sigma \in \{0,1\}^{<\mathbb{N}}$ of length n. This collection of algebras has an especially natural epistemic interpretation: each $\mathfrak{F}_n$ intuitively captures the possible information that the agent may obtain at the n-th stage of the learning process—any string of outcomes that could result from n experiments. The conditional expectation $\mathbb{E}_{\mu}[f \mid \mathfrak{F}_n] : \{0,1\}^{\mathbb{N}} \to \mathbb{R}$ of f given $\mathfrak{F}_n$ is itself a random variable that, on input $\omega \in \{0,1\}^{\mathbb{N}}$, returns the best estimate of f's value, from the perspective of μ, conditional on the first n digits $\omega \upharpoonright n$ of ω. More suggestively, when ω is the true state of the world, $\mathbb{E}_{\mu}[f \mid \mathfrak{F}_n](\omega)$ can be seen as encoding the agent's beliefs regarding the true value of f (namely, $f(\omega)$) after having observed the outcomes $\omega \upharpoonright n$ of the first n experiments. We use throughout the following version of the conditional expectation (since it is unique only up to μ-measure zero)—though, as will become clear, this choice is immaterial for our results. For all $\omega \in \{0,1\}^{\mathbb{N}}$,

$$\mathbb{E}_{\mu}[f \mid \mathfrak{F}_n](\omega) = \begin{cases} \dfrac{1}{\mu([\omega \upharpoonright n])} \displaystyle\int_{[\omega \upharpoonright n]} f\, d\mu & \text{if } \mu([\omega \upharpoonright n]) > 0, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

Lévy's Upward Theorem is the following result:[11]

Theorem 2.1 (Lévy's Upward Theorem, Lévy (1937)). Let $f : \{0,1\}^{\mathbb{N}} \to \mathbb{R}$ an integrable random variable. Then, for μ-almost every $\omega \in \{0,1\}^{\mathbb{N}}$, $\lim_{n\to\infty} \mathbb{E}_{\mu}[f \mid \mathfrak{F}_n](\omega) = f(\omega)$.[12]

Call the set of sequences $\{\omega \in \{0,1\}^{\mathbb{N}} : \lim_{n\to\infty} \mathbb{E}_{\mu}[f \mid \mathfrak{F}_n](\omega) = f(\omega)\}$ that satisfy Lévy's Upward Theorem the *success set* of agent μ with respect to f and its complement the *failure set*. Lévy's Upward Theorem says that, from the agent's viewpoint,

[10] See (Oxtoby, 1980, 85).

[11] See (Williams, 1991, §14.2).

[12] Technically, Lévy's Upward Theorem gives us more than this. A sequence of functions $\{g_n\}_{n\in\mathbb{N}}$ converges to an integrable function g in the L^1-norm if $\lim_{n\to\infty} \int_{\{0,1\}^{\mathbb{N}}} |g_n - g|\, d\mu = 0$. Besides establishing almost-sure pointwise convergence, Lévy's Upward Theorem also establishes that $\{\mathbb{E}_{\mu}[f \mid \mathfrak{F}_n]\}_{n\in\mathbb{N}}$ converges to f in the L^1-norm.

their failure set is negligible: the agent assigns probability one to the hypothesis that they will eventually converge to the truth about the value of f (i.e., $\mu(\{\omega \in \{0,1\}^{\mathbb{N}} : \lim_{n\to\infty} \mathbb{E}_\mu[f \mid \mathfrak{F}_n](\omega) = f(\omega)\}) = 1$).

The philosophical literature on Lévy's Upward Theorem generally focuses on the special case where the integrable random variable being estimated is the indicator function $\mathbb{1}_{\mathscr{A}}$ of some Borel set $\mathscr{A}$ (as we shall see, this is the setting within which Belot frames his objection). This restriction corresponds to the case where the inductive problem faced by the agent is a binary decision problem: does the true world—the observed data stream—belong to $\mathscr{A}$? Or, put differently, does the true world possess the property corresponding to $\mathscr{A}$? In this setting, the quantity that the agent is trying to estimate is the truth value of $\mathscr{A}$ and learning proceeds by standard Bayesian conditioning. Whenever $\mu([\omega \restriction n]) > 0$, we in fact have that

$$\mathbb{E}_\mu[\mathbb{1}_{\mathscr{A}} \mid \mathfrak{F}_n](\omega) = \frac{1}{\mu([\omega \restriction n])} \int_{[\omega\restriction n]} \mathbb{1}_{\mathscr{A}} \, d\mu = \frac{\mu(\mathscr{A} \cap [\omega \restriction n])}{\mu([\omega \restriction n])} = \mu(\mathscr{A} \mid [\omega \restriction n]).$$

So, since the *support* $\mathrm{supp}(\mu) = \{\omega \in \{0,1\}^{\mathbb{N}} : (\forall n)\, \mu([\omega \restriction n]) > 0\}$ of μ has μ-probability one, Lévy's Upward Theorem entails that $\lim_{n\to\infty} \mu(\mathscr{A} \mid [\omega \restriction n]) = \mathbb{1}_{\mathscr{A}}(\omega)$ for μ-almost every $\omega \in \{0,1\}^{\mathbb{N}}$. An agent with prior μ expects their beliefs, given by the above sequence of posterior probabilities, to converge almost surely to the truth about whether $\mathscr{A}$ is the case with increasing information.

As noted above, the almost-sure convergence to the truth achieved via Lévy's Upward Theorem is always relative to the agent's prior. Before performing any experiments or measurements, the agent is essentially certain that, with increasing information, their beliefs will eventually converge to the truth. Yet, there is no objective or external guarantee that this will indeed be the case. Thus, Lévy's Upward Theorem does not establish the universal reliability of Bayesian learning methods from an objective, third-person standpoint. Its epistemic significance stems from the fact that it establishes that a certain kind of scepticism about induction is impossible: if an agent is independently committed to probabilistic coherence, then, by Lévy's Upward Theorem, that agent cannot be a sceptic about the possibility of learning from experience. The agent's independent commitment to the Bayesian framework implies that, *by their own light*, their recourse to inductive reasoning is justified. As observed by Skyrms (1984), from the perspective of a Bayesian agent, it is "inappropriate for you to ask the standard question, "Why should I believe that the real situation is not in that set of measure zero?" The measure in question *is* your degree of belief. You *do* believe that the real situation is not in that set, with degree of belief one" (Skyrms, 1984, 62).

Belot's Argument. Precisely because of its barring a certain type of scepticism about induction, Lévy's Upward Theorem has however been accused of being a drawback of the Bayesian approach, rather than serving in its favour. In particular, Belot (2013) argues that, because of Lévy's Upward Theorem (and other convergence-to-the-truth results), Bayesian reasoners are forced to be epistemically immodest in an especially pernicious way. Belot's worry is that

> Bayesian convergence-to-the-truth theorems tell us that Bayesian agents are forbidden to think that there is any chance that they will be fooled in the long run, even when they know that their credence function is defined on a

https://doi.org/10.1017/psa.2023.177 Published online by Cambridge University Press

> space that includes many [data streams] that would frustrate their desire to reach the truth. (Belot, 2013, 500)

Bayesian reasoners are, in a sense, incapable of entertaining the possibility of inductive failure, and this is so even when "their desire to reach the truth" is thwarted along many data streams—in fact, on a topologically typical collection of data streams.

To make his point,[13] Belot considers a specific class of priors, which he calls *open-minded*. The notion of an open-minded prior is, by definition, always relative to a particular Borel set—more suggestively, to a particular hypothesis under consideration: given some $\mathscr{S} \in \mathfrak{B}$, μ is open-minded with respect to $\mathscr{S}$ if, no matter what string $\sigma \in \{0,1\}^{<\mathbb{N}}$ has been observed so far, there are always two possible distinct extensions $\tau, \rho \in \{0,1\}^{<\mathbb{N}}$ of σ (i.e., $\sigma \sqsubset \tau$ and $\sigma \sqsubset \rho$) such that $\mu(\mathscr{S} \mid [\tau]) \geq \frac{1}{2}$ and $\mu(\mathscr{S} \mid [\rho]) < \frac{1}{2}$. If μ is open-minded with respect to $\mathscr{S}$, then no finite number of observations will ever suffice for μ to settle on whether the data stream being observed belongs to $\mathscr{S}$.

Now, suppose the hypothesis under consideration corresponds to a countable dense Borel set $\mathscr{D}$ (for instance, $\mathscr{D}$ could be the set of sequences that are eventually 0). Given a Bayesian agent with prior μ, what do the success set and the failure set of μ with respect to the binary estimation problem encoded by $\mathbb{1}_{\mathscr{D}}$ respectively look like? The answer to this question of course depends on the particular prior adopted by the agent. Since $\mathscr{D}$ and its complement are both dense, any finite sequence of observations is compatible with the true data stream being in $\mathscr{D}$, but also with it not being in $\mathscr{D}$. Hence, according to Belot, in this case it is reasonable to adopt a prior μ that is open-minded with respect to $\mathscr{D}$. Yet, Belot shows, if μ is open-minded with respect to $\mathscr{D}$, then its failure set—the set of sequences $\omega \in \{0,1\}^{\mathbb{N}}$ along which the conditional probabilities $\mu(\mathscr{D} \mid [\omega \upharpoonright n])$ fail to converge to $\mathbb{1}_{\mathscr{D}}(\omega)$ in the limit—is co-meagre, despite being a μ-measure-zero set by Lévy's Upward Theorem. Equivalently, the success set of μ relative to $\mathbb{1}_{\mathscr{D}}$ is meagre (and, so, topologically negligible), despite having probability one according to the agent. Probabilistic and topological typicality are thus orthogonal notions in this setting.

In light of these (and other analogous) observations, Belot concludes that the Bayesian approach is irremediably flawed: the account of rationality it yields "renders a certain sort of arrogance rationally mandatory, requiring agents to be certain that they will be successful at certain tasks, even in cases where the task is so contrived as to make failure the typical outcome" (Belot, 2013, 484).

3. Meagre and Co-meagre Success

There are several moving parts in Belot's argument that one may call into question to avoid his conclusion. For instance, one may doubt the reasonableness of Belot's notion of an open-minded prior or challenge the very significance of topological considerations for Bayesian epistemology.[14] Huttegger (2015), for example, notes the following:

[13] Here we focus on the argument from (Belot, 2013, §4).

[14] See, for instance, (Huttegger, 2015) and (Cisewski et al., 2018).

https://doi.org/10.1017/psa.2023.177 Published online by Cambridge University Press

> [T]he mathematical structure of measure theory is very different from the mathematical structure of topology. [...] Taking all of this together suggests that topological and probabilistic concepts are fairly independent of each other, and that results about the topology of a space do not prescribe specific probability distributions for that space. From a Bayesian perspective, this makes a lot of sense. Topology is a mathematical theory of concepts like closeness and limit point, whereas probability is a mathematical theory of rational degrees of belief. The two theories have very different domains, and so there is no reason to suppose that there are any general principles connecting the two in the way required by Belot's argument. (Huttegger, 2015, 92)

The question of whether topological typicality is a relevant concept when evaluating the rationality of probabilistic reasoners is a divisive one. It is however worth pointing out that the use of topological typicality in the setting of Bayesian convergence-to-the-truth results has some notable precedents that bear upon the long-standing debate between Bayesians and frequentists in the foundations of statistics—see, in particular, Freedman (1963, 1965), who employs notions of topological typicality in studying the consistency of Bayes' estimates. Regardless of whether Belot's specific argument is ultimately successful, we take the following questions to be of independent interest: are there any ordinary learning situations where Bayesian agents (or, at the very least, certain types of Bayesian agents) are guaranteed to be inductively successful on a typical set of data streams both in the probabilistic and the topological sense? Is it possible to provide an informative classification of the learning scenarios where the two notions of typicality are in agreement (with respect to the success sets of Bayesian agents) and the learning scenarios where they instead come apart, so as to be able to understand how pervasive the phenomenon identified by Belot is? These are the questions that will keep us occupied in the remainder of this article.

Recall that a probability measure on $\mathfrak{B}$ has *full support* if it assigns positive probability to all cylinders. The uniform measure, for example, is a probability measure with full support, and so are all other Bernoulli measures with bias strictly less than 1, as well as their mixtures. All of the learning situations identified in this article for which convergence to the truth occurs on a co-meagre set will feature priors with full support. Priors with full support have a natural epistemic interpretation: they, too, correspond to a form of open-mindedness—in particular, they intuitively capture the credences of Bayesian reasoners who are open-minded with respect to the evidence, in that they do not a priori exclude any finite sequence of observations.

Of course, the type of open-mindedness encoded by having full support is compatible with various forms of closed-mindedness. Take, for instance, the set of data streams that are eventually 0, which, as remarked earlier, is both countable and dense. According to the uniform measure λ, this is a measure-zero set. Thus, though open-minded with respect to all finite sequences of observations, λ is closed-minded with respect to the possibility of observing only finitely many 1's. So, λ fails to be open-minded in Belot's sense with respect to the hypothesis encoded by this set. We take this to be a feature, rather than a bug, as no prior can be open-minded with respect to every event in $\mathfrak{B}$. In particular, Belot's notion of open-mindedness is just as susceptible to the charge of entailing various forms of closed-mindedness as having full support is.

https://doi.org/10.1017/psa.2023.177 Published online by Cambridge University Press

Lévy's Upward Theorem holds very generally for any integrable random variable. A crucial component of our analysis relies on classifying random variables in terms of their descriptive complexity and computational strength. This will allow us to identify well-behaved families of integrable random variables for which Lévy's Upward Theorem can be shown to hold on a co-meagre set. To this end, we need to introduce a few more definitions.

The Borel Hierarchy and the Arithmetical Hierarchy. The events in $\mathfrak{B}$ can be classified in terms of their *rank*, or *descriptive complexity*, within the *Borel hierarchy*:[15]

Definition 3.1 (Borel hierarchy). The *Borel hierarchy* of subsets of Cantor space consists of the following three types of classes: $\mathbf{\Sigma}^0_\alpha$, $\mathbf{\Pi}^0_\alpha$, and $\mathbf{\Delta}^0_\alpha$, with α a countable ordinal greater than 0. Given a positive natural number n,[16] a set $\mathscr{S} \in \mathfrak{B}$ is in

- $\mathbf{\Sigma}^0_1$ if and only if it is open;
- $\mathbf{\Pi}^0_n$ if and only if its complement $\overline{\mathscr{S}}$ is in $\mathbf{\Sigma}^0_n$;
- $\mathbf{\Sigma}^0_n$ ($n > 1$) if and only if there is a sequence $\{\mathscr{S}_i\}_{i\in\mathbb{N}}$ of $\mathbf{\Pi}^0_{n-1}$ sets such that $\mathscr{S} = \bigcup_{i\in\mathbb{N}} \mathscr{S}_i$;
- $\mathbf{\Delta}^0_n$ if and only if $\mathscr{S}$ is both in $\mathbf{\Sigma}^0_n$ and $\mathbf{\Pi}^0_n$.

For instance, the $\mathbf{\Pi}^0_1$ sets are the closed sets, the $\mathbf{\Delta}^0_1$ sets are the clopen sets, the $\mathbf{\Sigma}^0_2$ sets are countable unions of closed sets, and the $\mathbf{\Pi}^0_2$ sets are countable intersections of open sets.

The Borel hierarchy has an effective counterpart called the *arithmetical hierarchy*, which allows to classify certain Borel sets in terms of their *arithmetical complexity*:[17]

Definition 3.2 (Arithmetical hierarchy). The *arithmetical hierarchy* of subsets of Cantor space consists of the following three types of classes: Σ^0_n, Π^0_n, and Δ^0_n, with n a positive natural number. A set $\mathscr{S} \in \mathfrak{B}$ is in

- Σ^0_1 if and only if it is *effectively open* (i.e., if there is a computably enumerable set $S \subseteq \{0,1\}^{<\mathbb{N}}$ such that $\mathscr{S} = [S] = \bigcup_{\sigma\in S}[\sigma]$);
- Π^0_n if and only if its complement $\overline{\mathscr{S}}$ is in Σ^0_n;
- Σ^0_n ($n > 1$) if and only if there is a computable sequence $\{\mathscr{S}_i\}_{i\in\mathbb{N}}$ of Π^0_{n-1} sets[18] such that $\mathscr{S} = \bigcup_{i\in\mathbb{N}} \mathscr{S}_i$;
- Δ^0_n if and only if $\mathscr{S}$ is in both Σ^0_n and Π^0_n.

For instance, the Π^0_1 sets are the effectively closed sets, the Δ^0_1 sets are the (effectively) clopen sets, the Σ^0_2 sets are effective countable unions of effectively closed sets, and the

[15] See, for instance, (Kechris, 1995, §11B).

[16] Here we only focus on finite ordinals—and, so, on Borel sets of *finite rank*.

[17] See, e.g., (Soare, 2016, Chapter 4) or (Downey and Hirschfeldt, 2010, §2.19).

[18] This means that there is a computable function $g : \mathbb{N} \to \mathbb{N}$ such that, for each $i \in \mathbb{N}$, $\mathscr{S}_i = \overline{\mathscr{E}^{n-1}_{g(i)}}$, where $\mathscr{E}^{n-1}_0, \mathscr{E}^{n-1}_1, \mathscr{E}^{n-1}_2, \ldots$ is a fixed effective enumeration of all the Σ^0_{n-1} subsets of $\{0,1\}^{\mathbb{N}}$ (see, for instance, (Downey and Hirschfeldt, 2010, 75-76)).

https://doi.org/10.1017/psa.2023.177 Published online by Cambridge University Press

Π^0_2 sets are effective countable intersections of effectively open sets. A set with a classification within the arithmetical hierarchy is said to be *arithmetical* (or *arithmetically definable*).

The levels of the arithmetical hierarchy can also be characterized in terms of the complexity of the formulas in the language of first-order arithmetic that define the sets belonging to those levels (hence the name "arithmetical hierarchy"). A set $\mathscr{S} \in \mathfrak{B}$ is in Σ^0_n if and only if it is definable by a Σ^0_n formula: namely, if $\mathscr{S} = \{\omega \in \{0,1\}^{\mathbb{N}} : (\exists k_1)(\forall k_2)...(Qk_n)\, R(\omega \upharpoonright k_1, \omega \upharpoonright k_2, ..., \omega \upharpoonright k_n)\}$ for some computable relation R, with $Q = \exists$ if n is odd and $Q = \forall$ if n is even. On the other hand, a set $\mathscr{S} \in \mathfrak{B}$ is in Π^0_n if and only if it is definable by a Π^0_n formula: namely, if there is a computable relation R such that $\mathscr{S} = \{\omega \in \{0,1\}^{\mathbb{N}} : (\forall k_1)(\exists k_2)...(Qk_n)\, R(\omega \upharpoonright k_1, \omega \upharpoonright k_2, ..., \omega \upharpoonright k_n)\}$, with $Q = \forall$ if n is odd and $Q = \exists$ if n is even.

Algorithmic Randomness and Effective Genericity. As notions of "largeness" go, having measure one and being co-meagre are rather coarse-grained. There are many sets that, while measure-theoretically or topologically (a)typical, seem to intuitively differ in "size". For instance, the concept of *Hausdorff dimension*, which is a generalization of the uniform measure, was introduced to formalize the intuition that certain subsets of a metric space differ in size, even though, from the viewpoint of the uniform measure, they all have measure zero. In what follows, we will consider some more refined notions of typicality whose definitions rely on the machinery of computability theory. These notions of effective typicality allow to make more fine-grained distinctions between intuitively "large" sets. Here, we will use them to provide a more detailed analysis of the collections of data streams along which convergence to the truth holds for several classes of inductive problems.

Let us start with *algorithmic randomness*: a branch of computability theory that offers an account of effective measure-theoretic typicality.[19] According to algorithmic randomness, given a probability measure μ fixed in the background, a sequence is random relative to μ if it is a representative outcome of μ. One naïve idea is that an outcome is representative of μ if it satisfies every property that, according to μ, "most" sequences possess: i.e., if it belongs to all μ-measure-one sets. Since being representative in this sense is not possible in general,[20] algorithmic randomness instead identifies representativeness—and, so, randomness—with membership in certain countable collections of μ-measure-one sets: more precisely, μ-measure-one sets of a certain arithmetical complexity that correspond to natural statistical laws (such as the set of sequences that satisfy the Strong Law of Large Numbers relative to μ). In a nutshell, given such a countable collection of arithmetically definable properties that hold with μ-measure one (of arithmetically definable statistical laws), a sequence is algorithmically μ-random relative to that collection if and only if it possesses all of the corresponding properties (if and only if it satisfies all of the corresponding statistical laws).

[19] See (Nies, 2009) or (Downey and Hirschfeldt, 2010).

[20] For every probability measure μ that assigns probability zero to every singleton set (i.e., for every *atomless* probability measure), every sequence ω belongs to at least one μ-measure-zero set: its own singleton set $\{\omega\}$. Hence, defining randomness in terms of the satisfaction of all μ-measure one properties can lead to a vacuous notion.

The field of algorithmic randomness is teeming with notions of differing logical strength, each determined by the particular family of measure-one arithmetically definable properties a sequence must satisfy to count as random. In what follows, we will focus on two such notions.

Arguably, the simplest algorithmic randomness notion is *Kurtz randomness* (Kurtz, 1981):

Definition 3.3 (Kurtz randomness). Let μ a probability measure.[21] A sequence $\omega \in \{0,1\}^{\mathbb{N}}$ is *μ-Kurtz random* if and only if ω belongs to all Σ^0_1 sets of μ-measure one.

In other words, to qualify as μ-Kurtz random, a sequence has to possess all the properties that correspond to μ-measure-theoretically typical effectively open subsets of Cantor space (such as the property of having at least one prime-numbered 0 entry when μ is a non-trivial Bernoulli measure). Since there are only countably many Σ^0_1 sets, there are only countably many of them that have μ-measure-one. Hence, for every μ, the collection of μ-Kurtz random sequences is itself a μ-measure-one set.

Another fundamental algorithmic randomness notion is *Martin-Löf randomness* (Martin-Löf, 1966), which can be easily seen to entail Kurtz randomness:

Definition 3.4 (Martin-Löf randomness). Let μ a probability measure. A *μ-Martin-Löf test* is a computable sequence $\{\mathscr{U}_n\}_{n\in\mathbb{N}}$ of Σ^0_1 sets with $\mu(\mathscr{U}_n) \leq 2^{-n}$ for all $n \in \mathbb{N}$. A sequence $\omega \in \{0,1\}^{\mathbb{N}}$ is *μ-Martin-Löf random* if and only if, for all μ-Martin-Löf tests $\{\mathscr{U}_n\}_{n\in\mathbb{N}}$, $\omega \notin \bigcap_{n\in\mathbb{N}} \mathscr{U}_n$.

The requirement that, for a μ-Martin-Löf test $\{\mathscr{U}_n\}_{n\in\mathbb{N}}$, $\mu(\mathscr{U}_n) \leq 2^{-n}$ for all $n \in \mathbb{N}$ ensures that $\bigcap_{n\in\mathbb{N}} \mathscr{U}_n$ is a set of *effective μ-measure zero*: it is a μ-measure-zero set whose measure can be approximated at a computable rate (2^{-n}) using the measures of the components $\mathscr{U}_n$ of the test. And since the intersection of a computable sequence of Σ^0_1 sets is a Π^0_2 set, a sequence is μ-Martin-Löf random if and only if it does not possess any Π^0_2 properties of effective μ-measure zero—equivalently, if and only if it possesses all Σ^0_2 properties of effective μ-measure one. Once again, seeing that there are only countably many Σ^0_2 properties of (effective) μ-measure one, the collection of μ-Martin-Löf random sequences is itself a μ-measure-one set.

Since μ-Martin-Löf randomness entails μ-Kurtz randomness while the reverse implication does not hold in general, μ-Martin-Löf randomness yields a more fine-grained notion of measure-theoretic typicality than μ-Kurtz randomness does; in turn, μ-Kurtz randomness provides a more fine-grained notion of measure-theoretic typicality than simply having μ-measure one.

The second family of effective typicality notions that will be relevant for our discussion falls under the umbrella of *effective genericity*: a theory of effective topological

[21] Algorithmic randomness is often defined with respect to *computable* probability measures (see footnote 31). Here we will not impose such a restriction and focus on *blind* randomness (see, for instance, (Kjos-Hanssen, 2010)): namely, on notions where the underlying probability measure μ, which may be uncomputable, is not used as an oracle when specifying the class of μ-measure one arithmetically definable properties that a sequence has to satisfy to be random.

typicality.[22] Just as algorithmic randomness is defined in terms of membership in every measure-one set from some pre-specified countable collection of sets, effective genericity essentially amounts to membership in every co-meagre set from some pre-specified countable collection of sets. While there are many notions of effective genericity in the literature, here we will only consider the *n-genericity* hierarchy: a linearly ordered family of canonical genericity notions. We will begin by defining 1-*genericity* (the first level of the hierarchy) and discuss the rest of the hierarchy in the last section of the paper.

Given $\omega \in \{0,1\}^{\mathbb{N}}$ and $S \subseteq \{0,1\}^{<\mathbb{N}}$, ω is said to *meet* S if $\omega \in [S] = \bigcup_{\sigma \in S}[\sigma]$. A set $S \subseteq \{0,1\}^{<\mathbb{N}}$ is *dense along* $\omega \in \{0,1\}^{\mathbb{N}}$ if ω is in the closure of $[S]$: in other words, if, for every $n \in \mathbb{N}$, there is some $\sigma \in \{0,1\}^{<\mathbb{N}}$ with $\omega \upharpoonright n \sqsubseteq \sigma$ such that $[\sigma] \subseteq [S]$.[23] Then, 1-genericity is defined as follows:

Definition 3.5 (1-Genericity). A sequence $\omega \in \{0,1\}^{\mathbb{N}}$ is 1-*generic* if and only if ω meets every computably enumerable set $S \subseteq \{0,1\}^{<\mathbb{N}}$ that is dense along ω.

Equivalently, a sequence is 1-generic if and only if it is not on the boundary of any Σ^0_1 set. Intuitively, a 1-generic sequence ω is such that, for any Σ^0_1 hypothesis $\mathscr{S}$, if no imprecise measurement of ω can rule $\mathscr{S}$ out, then ω is in $\mathscr{S}$ (and, so, satisifies the hypothesis).

It is not difficult to see that every 1-generic sequence belongs to every dense Σ^0_1 set and that every such set is co-meagre. Moreover, the following well-known facts will be important for our discussion.[24]

Proposition 3.6. The set of 1-generic sequences is co-meagre.

Proposition 3.7. Let μ a probability measure with full support. If $\omega \in \{0,1\}^{\mathbb{N}}$ is 1-generic, then ω is μ-Kurtz random.[25]

When μ is a probability measure with full support, the set of μ-Kurtz random sequences is thus itself a co-meagre set. So, both 1-genericity and μ-Kurtz randomness provide more fine-grained notions of topological typicality than simply being co-meagre.

[22] See (Downey and Hirschfeldt, 2010, §2.24).

[23] For example, the set $A = \{0^n 1 \in \{0,1\}^{<\mathbb{N}} : n \in \mathbb{N} \text{ with } n \geq 1\}$ (where $0^n 1$ is the string consisting of n consecutive 0's followed by a 1) is dense along the constant 0 sequence 000000..., even though this sequence does not meet A.

[24] See (Kurtz, 1981), (Nies, 2009), or (Downey and Hirschfeldt, 2010).

[25] When μ does not have full support, Proposition 3.7 may fail to hold. Consider the probability measure δ concentrated on the constant 1 sequence 111111... (that is, δ is the measure given by $\delta([\varepsilon]) = 1$ and, for all $\sigma \neq \varepsilon$, $\delta([\sigma]) = 1$ if σ consists of $|\sigma|$ consecutive 1's and $\delta([\sigma]) = 0$ otherwise). Clearly, δ does not have full support (for one, $\delta([0]) = 0$). Now, the constant 1 sequence is not 1-generic—for instance, it fails to belong to the set $\{\omega \in \{0,1\}^{\mathbb{N}} : (\exists n)\ \omega(n) = 0\}$ of sequences with at least one 0 entry, which is both dense and Σ^0_1. However, the constant 1 sequence is the only δ-Kurtz random sequence. Hence, the set of 1-generic sequences and the set of δ-Kurtz random sequences are disjoint.

Continuous Functions. With these taxonomic tools at our disposal, we are ready to consider some specific families of inductive problems for which Bayesian convergence to the truth is not susceptible to Belot's charge of epistemic immodesty.

One way to achieve co-meagre success is of course to be inductively successful no matter what data stream is observed: that is, for convergence to the truth to occur everywhere, and not just almost everywhere according to the agent's prior. So, an immediate question is whether there are any ordinary learning situations where convergence to the truth can be achieved everywhere. Below, we highlight one such class of learning situations that will guide the rest of our discussion.

One simple type of inductive problem consists in estimating a continuous quantity. Quantities of this kind are naturally modelled in terms of continuous random variables. First, recall that the standard topology on $\mathbb{R}$ is the topology generated by the open intervals: namely, by sets of the form $(a,b) = \{r \in \mathbb{R} : a, b \in \mathbb{R} \text{ and } a < r < b\}$. Earlier, we introduced the Borel hierarchy of subsets of $\{0,1\}^{\mathbb{N}}$ (Definition 3.1). The very same taxonomy in terms of $\boldsymbol{\Sigma}^0_n$, $\boldsymbol{\Pi}^0_n$, and $\boldsymbol{\Delta}^0_n$ sets also applies to the Borel subsets of $\mathbb{R}$—where the Borel subsets of $\mathbb{R}$ are the elements of the Borel σ-algebra on $\mathbb{R}$: the smallest σ-algebra containing all open intervals. Continuous functions from $\{0,1\}^{\mathbb{N}}$ to $\mathbb{R}$ are defined as follows:

Definition 3.8 (Continuous function). A function $f : \{0,1\}^{\mathbb{N}} \to \mathbb{R}$ is *continuous* if and only if, for every open ($\boldsymbol{\Sigma}^0_1$) subset $\mathscr{U}$ of $\mathbb{R}$, $f^{-1}(\mathscr{U}) = \{\omega \in \{0,1\}^{\mathbb{N}} : f(\omega) \in \mathscr{U}\}$ is an open ($\boldsymbol{\Sigma}^0_1$) subset of $\{0,1\}^{\mathbb{N}}$.

Suppose an experiment is being conducted which involves measuring some real-valued physical parameter, such as the temperature at a given location or the concentration of some substance in a fluid. Such a learning situation may be modelled via the function $f : \{0,1\}^{\mathbb{N}} \to \mathbb{R}$ that maps each sequence in $\{0,1\}^{\mathbb{N}}$ to the real number in $[0,1]$ of which that sequence is the binary expansion. Let the true parameter be given by $f(\omega)$. Then, at each finite stage n of the learning process, the observed data $\omega \upharpoonright n$ provides an approximation of $f(\omega)$. The map f is a continuous function. For another simple example of a continuous function, let $\mathscr{U}$ be a clopen subset of $\{0,1\}^{\mathbb{N}}$ and take its indicator function $\mathbb{1}_{\mathscr{U}}$. Intuitively, $\mathbb{1}_{\mathscr{U}}$ represents a binary decision problem that can be settled with a finite amount of data (such as the question of whether the first n patients from a given sample all recovered after being treated for a certain disease).

The observation below is entirely straightforward, but it is a useful starting point. First, note that, since Cantor space is compact,[26] every continuous function on it is bounded both below and above;[27] hence, every continuous random variable on Cantor space is integrable. When the quantity to be estimated is a continuous random variable, it is easy to see that Lévy's Upward Theorem holds for every sequence in the support of the agent's prior. So, when the agent's prior has full support, Lévy's Upward Theorem holds everywhere.

[26] This means that each of its open covers has a finite subcover.

[27] For each positive $n \in \mathbb{N}$, let $\mathscr{U}_n = \{\omega \in \{0,1\}^{\mathbb{N}} : n > f(\omega) > -n\}$. Then, each $\mathscr{U}_n$ is open, $\mathscr{U}_n \subseteq \mathscr{U}_{n+1}$ for all $n \geq 1$, and $\bigcup_{n\geq 1} \mathscr{U}_n = \{0,1\}^{\mathbb{N}}$. By the compactness of Cantor space, there is some $n_0 \geq 1$ such that $\mathscr{U}_{n_0} = \{0,1\}^{\mathbb{N}}$. This establishes that f is bounded (below and above).

https://doi.org/10.1017/psa.2023.177 Published online by Cambridge University Press

Proposition 3.9. Let μ a probability measure and $f : \{0,1\}^{\mathbb{N}} \to \mathbb{R}$ a continuous random variable. Then, for all $\omega \in \text{supp}(\mu)$, $\lim_{n\to\infty} \mathbb{E}_\mu[f \mid \mathfrak{F}_n](\omega) = f(\omega)$. When μ has full support, $\text{supp}(\mu) = \{0,1\}^{\mathbb{N}}$ and, so, $\lim_{n\to\infty} \mathbb{E}_\mu[f \mid \mathfrak{F}_n](\omega) = f(\omega)$ for all $\omega \in \{0,1\}^{\mathbb{N}}$.

Proof. Let $\omega \in \text{supp}(\mu)$. Then, for all $n \in \mathbb{N}$, $\mathbb{E}_\mu[f \mid \mathfrak{F}_n](\omega) = \int_{[\omega\restriction n]} f\,d\mu / \mu([\omega\restriction n])$. Let $\varepsilon > 0$. By continuity, there is $m \in \mathbb{N}$ such that, for all $\alpha \in [\omega \restriction m]$, $|f(\omega) - f(\alpha)| < \varepsilon$. Hence, for all $n \geq m$,

$$\left| \frac{\int_{[\omega\restriction n]} f\,d\mu}{\mu([\omega\restriction n])} - f(\omega) \right| \leq \frac{\int_{[\omega\restriction n]} |f - f(\omega)|\,d\mu}{\mu([\omega\restriction n])} < \frac{\int_{[\omega\restriction n]} \varepsilon\,d\mu}{\mu([\omega\restriction n])} = \varepsilon,$$

where the first inequality holds because $f(\omega)$ is a constant. This establishes the claim. □

When the agent's prior μ does not have full support, convergence to the truth is not guaranteed to happen on a co-meagre set. To see this, let μ be the probability measure that results from first flipping a coin that lands heads with probability one and then flipping a fair coin forever after.[28] Take the indicator function $\mathbb{1}_{[0]}$ of the cylinder $[0]$, which, as noted above, is continuous. Lévy's Upward Theorem fails everywhere on $[0]$, so the success set of μ with respect to $\mathbb{1}_{[0]}$ is not co-meagre, as the cylinder $[1]$ is not co-meagre.

Baire Class n Functions. Continuity, while natural, is a strong condition. What we will consider next is a family of functions that, relying on the classifications afforded by the Borel hierarchy, provides a broad generalization of the class of continuous functions:

Definition 3.10 (Baire class n function.). Let $n \in \mathbb{N}$. A function $f : \{0,1\}^{\mathbb{N}} \to \mathbb{R}$ is of *Baire class n* if and only if, for every $\mathbf{\Sigma}^0_1$ subset $\mathscr{U}$ of $\mathbb{R}$, $f^{-1}(\mathscr{U})$ is a $\mathbf{\Sigma}^0_{n+1}$ subset of $\{0,1\}^{\mathbb{N}}$.

Clearly, the collection of Baire class 0 functions coincides with the collection of continuous functions. Moreover, for each $n \in \mathbb{N}$, every Baire class n function is also a Baire class $(n+1)$ function (while the converse does not hold).

For each $n \geq 1$, the indicator functions of the $\mathbf{\Delta}^0_n$, $\mathbf{\Sigma}^0_n$, $\mathbf{\Pi}^0_n$, and $\mathbf{\Delta}^0_{n+1}$ subsets of Cantor space are straightforward examples of Baire class n functions. These functions have natural epistemic interpretations. For instance, the indicator functions of $\mathbf{\Sigma}^0_1$ sets intuitively capture binary decision problems membership in which can be verified with a finite amount of data, while the indicator functions of $\mathbf{\Pi}^0_1$ sets intuitively capture binary decision problems membership in which can be refuted with a finite amount of data. Similarly, the indicator functions of $\mathbf{\Sigma}^0_2$ sets correspond to binary decision problems membership in which can be verified in the limit, the indicator functions of $\mathbf{\Pi}^0_2$ sets correspond to binary decision problems membership in which can be refuted in the

[28]More precisely, μ is the probability measure given by $\mu([\varepsilon]) = 1$ and, for all strings $\sigma \neq \varepsilon$, $\mu([\sigma]) = 0$ if the first entry of σ is a 0 and $\mu([\sigma]) = 2^{-|\sigma|+1}$ otherwise.

limit, and the indicator functions of $\mathbf{\Delta}^0_2$ sets correspond to binary decision problems membership in which can be decided in the limit.[29]

Another family of functions that are of Baire class 1 (in addition to the indicator functions of $\mathbf{\Delta}^0_1$, $\mathbf{\Sigma}^0_1$, $\mathbf{\Pi}^0_1$, and $\mathbf{\Delta}^0_2$ sets) is the class of *semicontinuous* functions, which includes the *lower semicontinuous* and the *upper semicontinuous* functions:

Definition 3.11 (Lower and upper semicontinuous function). A function $f : \{0, 1\}^{\mathbb{N}} \to \mathbb{R}$ is *lower semicontinuous* if and only if all sets of the form $f^{-1}((a, +\infty)) = \{\omega \in \{0, 1\}^{\mathbb{N}} : f(\omega) > a\}$ are $\mathbf{\Sigma}^0_1$. A function $f : \{0, 1\}^{\mathbb{N}} \to \mathbb{R}$ is *upper semicontinuous* if and only if all sets of the form $f^{-1}((-\infty, b)) = \{\omega \in \{0, 1\}^{\mathbb{N}} : f(\omega) < b\}$ are $\mathbf{\Sigma}^0_1$.

Semicontinuity is a weaker form of continutiy, and it is not difficult to see that a function is continuous if and only if it is both lower semicontinuous and upper semicontinuous. The collection of lower semicontinuous functions includes the indicator functions of $\mathbf{\Sigma}^0_1$ sets, while the collection of upper semicontinuous functions includes the indicator functions of $\mathbf{\Pi}^0_1$ sets. For another simple example of a lower semicontinuous function (and, so, of a Baire class 1 function), let f be given by $f(\omega) = 1$ if ω's prime-numbered entries feature a 1 infinitely often and $f(\omega) = 1 - 2^{-n}$ if ω has exactly $n \in \mathbb{N}$ prime-numbered entries featuring a 1 (basically, for any ω, f is a normalized function that counts the number of 1's in ω that occur at prime-numbered positions). For one last example, take a bounded function $c : \{0, 1\}^{<\mathbb{N}} \to \mathbb{R}$ recording the daily values of some stock market share. Then, the function f given by $f(\omega) = \sup_{n,m\in\mathbb{N}} |c(\omega \upharpoonright n) - c(\omega \upharpoonright m)|$, which tracks the greatest spread between this share's values over its history, is lower semicontinuous. Similarly, the function g given by $g(\omega) = \inf_{n,m\in\mathbb{N}} |c(\omega \upharpoonright n) - c(\omega \upharpoonright m)|$, which tracks the lowest spread between the share's values over its history, is upper semicontinuous.

The following is a classical result due to Baire[30] that will help us shed light on Lévy's Upward Theorem in the context of Baire class 1 functions:

Theorem 3.12 (Baire). Let $f : \{0, 1\}^{\mathbb{N}} \to \mathbb{R}$ a function of Baire class 1. The points of discontinuity of f form a meagre $\mathbf{\Sigma}^0_2$ set—equivalently, the points of continuity of f form a co-meagre $\mathbf{\Pi}^0_2$ set.

With Theorem 3.12 at hand, the following can be easily seen to hold:

Corollary 3.13. Let μ a probability measure with full support and $f : \{0, 1\}^{\mathbb{N}} \to \mathbb{R}$ a Baire class 1 integrable random variable. Then, the collection of all $\omega \in \{0, 1\}^{\mathbb{N}}$ with $\lim_{n\to\infty} \mathbb{E}_\mu[f \mid \mathfrak{F}_n](\omega) = f(\omega)$ is co-meagre.

Proof. Let ω be a point of continuity of f. Since μ has full support, $\mathbb{E}_\mu[f \mid \mathfrak{F}_n](\omega) = \int_{[\omega\upharpoonright n]} f\, d\mu / \mu([\omega\upharpoonright n])$ for all $n \in \mathbb{N}$. By the very same argument used in the proof of Proposition 3.9, $\lim_{n\to\infty} \mathbb{E}_\mu[f \mid \mathfrak{F}_n](\omega) = f(\omega)$. So, the set of points of continuity of

[29]For a detailed discussion of how to provide a learning-theoretic interpretation of the levels of the Borel hierarchy (and of the arithmetical hierarchy), see (Kelly, 1996).

[30]See (Kechris, 1995, Theorem 24.14) or (Oxtoby, 1980, Theorem 7.3).

https://doi.org/10.1017/psa.2023.177 Published online by Cambridge University Press

f is a subset of the set of sequences along which Lévy's Upward Theorem holds. By Theorem 3.12, the former set is co-meagre. Hence, so is the set of sequences along which Lévy's Upward Theorem holds. □

When the underlying prior μ has full support, the success set of μ relative to a Baire class 1 integrable random variable is a co-meagre set. We thus have another class of inductive problems relative to which convergence to the truth is not only probabilistically typical, but also topologically typical.

There are priors that do not have full support for which Corollary 3.13 does not hold. By virtue of being continuous, the indicator function $\mathbb{1}_{[0]}$ of $[0]$ is also of Baire class 1, and we have already mentioned an example of a probability measure that does not have full support whose success set with respect to $\mathbb{1}_{[0]}$ fails to be co-meagre.

For $n \geq 2$, it is not in general true that the points of discontinuity of a Baire class n function form a meagre set. Consider once again the set of all sequences that are eventually 0: i.e., the set $\mathscr{Z} = \{\omega \in \{0,1\}^{\mathbb{N}} : (\exists n)(\forall m > n)\ \omega(m) = 0\}$. The indicator function $\mathbb{1}_{\mathscr{Z}}$ of this set is of Baire class 2, yet $\mathbb{1}_{\mathscr{Z}}$ is discontinuous everywhere. Hence, the points of continuity of $\mathbb{1}_{\mathscr{Z}}$ not only fail to form a co-meagre set: they form a meagre set (since this set is empty).

Of course, this remark does not preclude the possibility that an analogue of Corollary 3.13 may hold for Baire class n integrable random variables in general. The following proposition establishes that Corollary 3.13 does not generalize:

Proposition 3.14. Let μ a computable probability measure[31] with full support. For each $n \geq 2$, there is a Baire class n integrable random variable $f : \{0,1\}^{\mathbb{N}} \to \mathbb{R}$ such that the collection of all $\omega \in \{0,1\}^{\mathbb{N}}$ with $\lim_{n\to\infty} \mathbb{E}_{\mu}[f \mid \mathfrak{F}_n](\omega) = f(\omega)$ is meagre.

Proof. Let μ-MLR denote the set of μ-Martin-Löf random sequences (Definition 3.4). The computability of μ ensures the existence of a *universal* μ-Martin-Löf test: a μ-Martin-Löf test $\{\mathscr{V}_n\}_{n\in\mathbb{N}}$ such that, to determine whether a sequence is μ-Martin-Löf random or not, it suffices to check whether that sequence belongs to $\bigcap_{n\in\mathbb{N}} \mathscr{V}_n$ (if it does, the sequence is not μ-Martin-Löf random, if it does not, it is).[32] Then, $\mu\text{-MLR} = \overline{\bigcap_{n\in\mathbb{N}} \mathscr{V}_n}$. Since $\bigcap_{n\in\mathbb{N}} \mathscr{V}_n$ is a Π^0_2 set, μ-MLR is a Σ^0_2 set. A fortiori, μ-MLR is in $\boldsymbol{\Sigma}^0_2$. The set $\bigcap_{n\in\mathbb{N}} \mathscr{V}_n$ is also dense. For, suppose not. Then, there is some $\sigma \in \{0,1\}^{<\mathbb{N}}$ with $\left(\bigcap_{n\in\mathbb{N}} \mathscr{V}_n\right) \cap [\sigma] = \emptyset$. Hence, $[\sigma] \subseteq \mu$-MLR. Take a computable sequence in $[\sigma]$ that is not a μ-atom. We can find such a sequence as follows. Since μ is computable, we can computably find $\tau_1 \in \{0,1\}^{<\mathbb{N}}$ with $\mu([\sigma\tau_1]) < \frac{1}{2}$ by dovetailing through the cylinders contained in $[\sigma]$, approximating their respective measures from above. And, given $\tau_1, \ldots, \tau_n \in \{0,1\}^{<\mathbb{N}}$ with $\mu([\sigma\tau_1 \ldots \tau_n]) < 2^{-n}$, we can computably find $\tau_{n+1} \in \{0,1\}^{<\mathbb{N}}$ with $\mu([\sigma\tau_1 \ldots \tau_{n+1}]) < 2^{-(n+1)}$ by following the same procedure inside $[\sigma\tau_1 \ldots \tau_n]$. Then, $\omega = \sigma\tau_1\tau_2 \ldots$ is a computable sequence with $\mu(\{\omega\}) = 0$. Since the only way for a

[31] A real number r is computable if it is computably approximable: if there is a computable sequence $q_0, q_1, q_2, \ldots$ of rational numbers such that $|q_n - r| \leq 2^{-n}$ for all $n \in \mathbb{N}$. A probability measure μ on $\mathfrak{B}$ is computable if, for any $\sigma \in \{0,1\}^{<\mathbb{N}}$, $\mu([\sigma])$ is a computable real number, uniformly in σ. This means that there is a computable function that, on input $\sigma \in \{0,1\}^{\mathbb{N}}$ and $n \in \mathbb{N}$, returns the n-th rational in a computable approximation of $\mu([\sigma])$. For a simple example of a computable probability measure, take any Bernoulli measure with a computable bias.

[32] See (Downey and Hirschfeldt, 2010, Theorem 6.2.5).

https://doi.org/10.1017/psa.2023.177 Published online by Cambridge University Press

computable sequence to be μ-Martin-Löf random is to be a μ-atom, ω is not μ-Martin-Löf random. But this contradicts the fact that $[\sigma] \subseteq \mu\text{-MLR}$. Hence, $\bigcap_{n\in\mathbb{N}} \mathscr{V}_n$ is indeed dense. Clearly, $\mu\text{-MLR} = \bigcup_{n\in\mathbb{N}} \overline{\mathscr{V}_n}$, where each set $\overline{\mathscr{V}_n}$ is Π^0_1 and, so, closed. By the density of $\bigcap_{n\in\mathbb{N}} \mathscr{V}_n$, no $\overline{\mathscr{V}_n}$ contains any non-empty open sets. Therefore, each $\overline{\mathscr{V}_n}$ is such that its closure has empty interior: i.e., each $\overline{\mathscr{V}_n}$ is nowhere dense. Hence, μ-MLR is meagre. The indicator function $\mathbb{1}_{\mu\text{-MLR}}$ of μ-MLR is a Baire class 2 integrable random variable. And since $\mu(\mu\text{-MLR}) = 1$ and μ has full support, the set of sequences $\alpha \in \{0, 1\}^{\mathbb{N}}$ with $\lim_{n\to\infty} \mathbb{E}_\mu[\mathbb{1}_{\mu\text{-MLR}} \mid \mathfrak{F}_n](\alpha) = \mathbb{1}_{\mu\text{-MLR}}(\alpha)$ coincides with μ-MLR. Therefore, the set of sequences along which Lévy's Upward Theorem holds for $\mathbb{1}_{\mu\text{-MLR}}$ is meagre. For $n > 2$, we can then reason as follows.[33] The notion of μ-Kurtz randomness (Definition 3.3) can be generalized to arbitrary levels of the arithmetical hierarchy: for any $n \geq 1$, a sequence is *μ-weakly n-random* if and only if it belongs to every Σ^0_n set of μ-measure one (so, μ-Kurtz randomness coincides with μ-weak 1-randomness). The set of μ-weakly n-random sequences is in $\mathbf{\Pi}^0_{n+1}$ and has μ-measure one. Moreover, μ-weak $(n+1)$-randomness entails μ-weak n-randomness and μ-weak 2-randomness entails μ-Martin-Löf randomness.[34] Hence, for each $n > 2$, by the same argument used for $\mathbb{1}_{\mu\text{-MLR}}$, the indicator function of the set of μ-weakly $(n-1)$-random sequences is an example of a Baire class n integrable random variable for which convergence to the truth occurs on a meagre set. □

Proposition 3.9 and Corollary 3.13 circumscribe the reach of Belot's objection: they establish that, at least for relatively simple inductive problems, convergence to the truth is topologically typical, in addition to being probabilistically typical. Proposition 3.14 pulls in the opposite direction. Not only does it show that, for more complex classes of inductive problems, co-meagre success is not always achievable, it also reveals that the dichotomy problematized by Belot is perhaps more pervasive than one might have initially thought. Past the level of Baire class 1 integrable random variables, co-meagre failure can be easily found at every level of the Borel hierarchy.

Computable and Almost Everywhere Computable Functions. Though well-behaved, all of the functions considered so far were allowed to be arbitrarily computationally complex. We will now concentrate on effective functions—functions whose values are in some sense calculable or approximable—and provide several examples of effective random variables for which convergence to the truth is topologically typical.

The very same taxonomy we discussed in the context of the arithmetical hierarchy of subsets of Cantor space also applies to the arithmetical subsets of $\mathbb{R}$. Here, the Σ^0_1 sets are those that can be expressed as a computably enumerable union of open intervals with rational endpoints, while the other levels of the hierarchy are defined in the same way as in the Cantor space setting. Computable functions from $\{0, 1\}^{\mathbb{N}}$ to $\mathbb{R}$ are defined as follows:

---

[33] Of course, $\mathbb{1}_{\mu\text{-MLR}}$ is a Baire class n integrable random variable for every $n \geq 2$, so the above argument already suffices to establish the claim. In the remainder of the proof, we will however show that it is possible to identify a different Baire class n integrable random variable with a meagre success set for each $n \geq 2$. The random variables we shall consider are different from each other for every non-degenerate probability measure for which the algorithmic randomness hierarchy does not collapse.

[34] See (Downey and Hirschfeldt, 2010, §7.2).

https://doi.org/10.1017/psa.2023.177 Published online by Cambridge University Press

Definition 3.15 (Computable function). A function $f : \{0,1\}^{\mathbb{N}} \to \mathbb{R}$ is *computable* if and only if, for every Σ_1^0 subset $\mathscr{U}$ of $\mathbb{R}$, $f^{-1}(\mathscr{U})$ is a Σ_1^0 subset of $\{0,1\}^{\mathbb{N}}$, uniformly in a code for $\mathscr{U}$.[35]

It is not difficult to see that the computable functions are precisely those functions whose values can be computably approximated to any degree of precision (via a computable sequence of rational-valued step functions).[36]

Every open set in the standard topology on the reals can be expressed as a (countable) union of Σ_1^0 sets. Moreover, every Σ_1^0 subset of Cantor space is open. Therefore, every computable function is continuous. Consequently, Proposition 3.9 holds for computable random variables, as well. And when μ has full support, the set of sequences along which Lévy's Upward Theorem holds is co-meagre. So, a Bayesian agent with a prior with full support trying to estimate a computable quantity is guaranteed to be inductively successful on a topologically typical collection of data streams (in fact, along every data stream).

Much like its classical counterpart—the concept of a continuous function—the notion of a computable function is rather demanding. The following example, taken from (Ackerman et al., 2019), nicely illustrates this point. Let $\theta \in [0,1]$. A $\{0,1\}$-valued random variable $f : \{0,1\}^{\mathbb{N}} \to \mathbb{R}$ relative to a probability measure μ is a *θ-Bernoulli random variable* if $\mu(\{\omega \in \{0,1\}^{\mathbb{N}} : f(\omega) = 1\}) = \theta$ (i.e., if the probability that f takes value 1 is θ). Ackerman et al. (2019) show that, for any $\theta \in [0,1]$ that is not a dyadic rational, every θ-Bernoulli random variable fails to be continuous and, as a result, is not computable in the sense of Definition 3.15. At the same time, for any computable $\theta \in [0,1]$, there are many θ-Bernoulli random variables that, though not computable, are "very close" to being computable. The following, more permissive notion is thus generally regarded as the more natural one to focus on in the context of (computable) probability theory:[37]

Definition 3.16 (Almost everywhere computable function). Let μ a probability measure. A partial function $f : \{0,1\}^{\mathbb{N}} \to \mathbb{R}$ is *μ-almost everywhere computable* if and only if it is computable on a Π_2^0 subset of $\{0,1\}^{\mathbb{N}}$ of μ-measure one: namely, if there is a Π_2^0 set $\mathscr{D} \subseteq \{0,1\}^{\mathbb{N}}$ with $\mu(\mathscr{D}) = 1$ such that f is defined on every $\omega \in \mathscr{D}$ and, for every Σ_1^0 subset $\mathscr{U}$ of $\mathbb{R}$, $f^{-1}(\mathscr{U}) \cap \mathscr{D} = \mathscr{U}' \cap \mathscr{D}$, where $\mathscr{U}'$ is a Σ_1^0 subset of $\{0,1\}^{\mathbb{N}}$, uniformly in a code for $\mathscr{U}$.[38]

[35] Let $\mathscr{S}_0, \mathscr{S}_1, \mathscr{S}_2, \ldots$ be a fixed effective enumeration of all the Σ_1^0 subsets of $\mathbb{R}$ and $\mathscr{E}_0, \mathscr{E}_1, \mathscr{E}_2, \ldots$ a fixed effective enumeration of all the Σ_1^0 subsets of $\{0,1\}^{\mathbb{N}}$. The uniformity condition in Definition 3.15 means that there is a computable function $g : \mathbb{N} \to \mathbb{N}$ such that, for all $n \in \mathbb{N}$, $f^{-1}(\mathscr{S}_n) = \mathscr{E}_{g(n)}$.

[36] See, e.g., (Li and Vitányi, 2019, 35-36).

[37] See (Hoyrup, 2008), (Hoyrup and Rojas, 2009), and (Ackerman et al., 2019). Importantly, all standard operations on random variables (such as addition, multiplication, composition, and Cartesian products) preserve almost everywhere computability.

[38] The restriction to Π_2^0 sets might seem surprising at first. We could have defined a μ-almost everywhere computable function as one that is computable on a μ-measure one subset of $\{0,1\}^{\mathbb{N}}$, without requiring that this set also be Π_2^0. However, by effectivizing a classical result due to Kuratowski (see (Kechris, 1995, Theorem 3.8), (Hoyrup, 2008, Theorem 1.6.2.1) or (Ackerman et al., 2019, Remark 2.11)), one can show that, for any such function $f : \{0,1\}^{\mathbb{N}} \to \mathbb{R}$, there is a function $f' : \{0,1\}^{\mathbb{N}} \to \mathbb{R}$ that is μ-almost everywhere computable in the sense of Definition 3.16 and agrees with f on all the sequences over which f is computable. Moreover, a code for the Π_2^0 set over which f' is computable can be computed uniformly from

https://doi.org/10.1017/psa.2023.177 Published online by Cambridge University Press

The argument from the proof of Proposition 3.9 can once again be employed to establish the following fact (since almost everywhere computable random variables are integrable):

Proposition 3.17. Let μ a probability measure and $f : \{0,1\}^{\mathbb{N}} \to \mathbb{R}$ a μ-almost everywhere computable random variable, with $\mathscr{D}$ the Π^0_2 set of μ-measure one on which f is computable. Then, for all $\omega \in supp(\mu) \cap \mathscr{D}$, $\lim_{n\to\infty} \mathbb{E}_\mu[f \mid \mathfrak{F}_n](\omega) = f(\omega)$.

Proof. Let $\omega \in \mathrm{supp}(\mu) \cap \mathscr{D}$. Then, for all $n \in \mathbb{N}$, $\mathbb{E}[f \mid \mathfrak{F}_n](\omega) = \int_{[\omega\restriction n]} f\, d\mu / \mu([\omega\restriction n])$. Let $\varepsilon > 0$. Since f is computable on $\mathscr{D}$, f is continuous on $\mathscr{D}$. Hence, there is $m \in \mathbb{N}$ such that, for all $\alpha \in \mathscr{D}$, if $\alpha \in [\omega \restriction m]$, then $|f(\omega) - f(\alpha)| < \varepsilon$. Therefore, for all $n \geq m$,

$$\begin{aligned}\left|\frac{\int_{[\omega\restriction n]} f\, d\mu}{\mu([\omega\restriction n])} - f(\omega)\right| &\leq \frac{\int_{[\omega\restriction n]} |f - f(\omega)|\, d\mu}{\mu([\omega\restriction n])} = \frac{\int_{[\omega\restriction n]\cap\mathscr{D}} |f - f(\omega)|\, d\mu}{\mu([\omega\restriction n])}\\ &< \frac{\int_{[\omega\restriction n]\cap\mathscr{D}} \varepsilon\, d\mu}{\mu([\omega\restriction n])} \qquad = \varepsilon,\end{aligned}$$

where both identities follow from the fact that $\mu(\mathscr{D}) = 1$. This establishes the claim. □

For any probability measure μ with full support, the Π^0_2 collection of sequences over which a μ-almost everywhere computable function is computable is co-meagre:

Proposition 3.18. Let μ a probability measure with full support and $\mathscr{D} \subseteq \{0,1\}^{\mathbb{N}}$ a Π^0_2 set of μ-measure one. Then, $\mathscr{D}$ is co-meagre.

Proof. Let $\{\mathscr{D}_n\}_{n\in\mathbb{N}}$ a computable sequence of Σ^0_1 sets with $\mathscr{D} = \bigcap_{n\in\mathbb{N}} \mathscr{D}_n$. Then, $\mu(\mathscr{D}_n) = 1$ for all $n \in \mathbb{N}$ and, so, each $\mathscr{D}_n$ is dense. For, suppose not. Then, there is some $n \in \mathbb{N}$ and $\sigma \in \{0,1\}^{<\mathbb{N}}$ such that $\mathscr{D}_n \cap [\sigma] = \emptyset$. Since $\mu([\sigma]) > 0$, $\mu(\mathscr{D}_n) \leq 1 - \mu([\sigma]) < 1$, which yields a contradiction. Let $\mathscr{U} \subseteq \{0,1\}^{\mathbb{N}}$ be an arbitrary open set. Since each $\mathscr{D}_n$ is dense, $\mathscr{D}_n \cap \mathscr{U} \neq \emptyset$ for all n. Moreover, given that each $\mathscr{D}_n$ is open, each set $\mathscr{D}_n \cap \mathscr{U}$ is open in the subspace topology on $\mathscr{U}$. For all n, since $(\overline{\mathscr{D}_n} \cap \mathscr{U}) \cap (\mathscr{D}_n \cap \mathscr{U}) = \emptyset$, $\overline{\mathscr{D}_n} \cap \mathscr{U}$ is not dense in $\mathscr{U}$. Hence, each $\overline{\mathscr{D}_n}$ is nowhere dense. Therefore, $\overline{\mathscr{D}} = \bigcup_{n\in\mathbb{N}} \overline{\mathscr{D}_n}$ is meagre and $\mathscr{D}$ is co-meagre. □

Proposition 3.17 and Proposition 3.18 entail the following:

Corollary 3.19. Let μ a probability measure with full support and $f : \{0,1\}^{\mathbb{N}} \to \mathbb{R}$ a μ-almost everywhere computable random variable. Then, the collection of all $\omega \in \{0,1\}^{\mathbb{N}}$ with $\lim_{n\to\infty} \mathbb{E}_\mu[f \mid \mathfrak{F}_n](\omega) = f(\omega)$ is co-meagre.

Hence, a Bayesian reasoner whose prior has full support has a co-meagre success set for any random variable whose values are computably approximable.

Randomness and Genericity at Work. While in and of themselves significant, convergence to the truth with probability one and convergence to the truth on a co-meagre

a code for the family of Σ^0_1 sets witnessing the μ-almost everywhere computability (in the weaker sense defined above) of f. Hence, without loss of generality, we can always assume that the μ-measure-one set of sequences over which a μ-almost everywhere computable function is computable is Π^0_2.

https://doi.org/10.1017/psa.2023.177 Published online by Cambridge University Press

set remain somewhat elusive notions. For one, in its classical form, Lévy's Upward Theorem is silent as to which data streams belong to the probability-one set of sequences along which convergence to the truth provably occurs. More generally, proving that convergence to the truth happens with probability one or on a co-meagre set provides little information about the composition of the success set. It also does not tell us how the composition of this set varies depending on the particular quantity the agent is trying to estimate, nor does it indicate whether the data streams that ensure eventual convergence to the truth share any property that might explain their conduciveness to learning—that is, any significant property that sets them apart from the data streams along which learning fails. In what follows, we will address these worries from the vantage point of computability theory. In particular, we will see that the theories of algorithmic randomness and effective genericity can be put to use to identify specific topologically typical collections of data streams along which convergence to the truth in the sense of Lévy's Upward Theorem is achieved for several classes of effective random variables.

We will begin by having a second look at the class of almost everywhere computable random variables.

Recall the definition of μ-Kurtz randomness (Definition 3.3). Given μ, let μ-KR denote the set of μ-Kurtz random sequences.

Proposition 3.20. Let μ a probability measure and $f:\{0,1\}^{\mathbb{N}} \to \mathbb{R}$ a μ-almost everywhere computable function, with $\mathscr{D}$ the μ-measure one Π^0_2 set on which f is computable. Then, $\mu\text{-}\mathsf{KR} \subseteq \mathit{supp}(\mu) \cap \mathscr{D}$.

Proof. Let $\omega \in \{0,1\}^{\mathbb{N}}$. First, suppose that $\mu([\omega \restriction n]) = 0$ for some $n \in \mathbb{N}$. Then, there is a Π^0_1 set of μ-measure zero, $[\omega \restriction n]$, that ω belongs to, which entails that $\omega \notin \mu\text{-}\mathsf{KR}$. Now, suppose that $\omega \in \overline{\mathscr{D}}$. Since $\overline{\mathscr{D}}$ is a Σ^0_2 set of μ-measure zero, $\overline{\mathscr{D}} = \bigcup_{n\in\mathbb{N}} \mathscr{A}_n$, where each $\mathscr{A}_n$ is in Π^0_1 and has μ-measure zero. Therefore, there is once again a Π^0_1 set of μ-measure zero that ω belongs to. Hence, $\omega \notin \mu\text{-}\mathsf{KR}$. □

By combining Proposition 3.20 with Proposition 3.17, we can immediately conclude that, for *any* μ-almost everywhere computable random variable (and, a fortiori, any computable random variable), observing a μ-Kurtz random data stream is a sufficient condition for converging to the truth in the limit:

Corollary 3.21. Let μ a probability measure and $f:\{0,1\}^{\mathbb{N}} \to \mathbb{R}$ a μ-almost everywhere computable random variable. Then, for all $\omega \in \mu\text{-}\mathsf{KR}$, $\lim_{n\to\infty} \mathbb{E}[f \mid \mathfrak{F}_n](\omega) = f(\omega)$.

When μ has full support, the set μ-KR is co-meagre. Hence, when the agent's prior has full support, there is a precisely identifiable collection of data streams—one that is typical both probabilistically and topologically—membership in which guarantees convergence to the truth for any inductive problem that can be modelled as a μ-almost everywhere computable random variable.

Just as the computable functions are the effective analogue of the continuous functions, Baire class n functions can be naturally effectivized as follows:

Definition 3.22 (Effective Baire class n function.). Let $n \in \mathbb{N}$. A function $f : \{0,1\}^{\mathbb{N}} \to \mathbb{R}$ is of *effective Baire class n* if and only if, for every Σ^0_1 subset $\mathscr{U}$ of $\mathbb{R}$, $f^{-1}(\mathscr{U})$ is a Σ^0_{n+1} subset of $\{0,1\}^{\mathbb{N}}$, uniformly in a code for $\mathscr{U}$.

For each $n \geq 1$, the indicator functions of Δ^0_n, Σ^0_n, Π^0_n, and Δ^0_{n+1} sets are simple examples of effective Baire class n functions (this, of course, is the effective counterpart of the fact that the indicator functions of $\boldsymbol{\Delta}^0_n$, $\boldsymbol{\Sigma}^0_n$, $\boldsymbol{\Pi}^0_n$, and $\boldsymbol{\Delta}^0_{n+1}$ sets are Baire class n functions in the classical setting). For instance, the indicator functions of Σ^0_1 sets, which intuitively correspond to binary decision problems that can be effectively verified with a finite amount of data, and the indicator functions of Π^0_1 sets, which correspond to binary decision problems that can be effectively refuted with a finite amount of data, are all effective Baire class 1 functions. For level one of the hierarchy, another natural example is the collection of *semicomputable* functions:

Definition 3.23 (Lower and upper semicomputable function). A function $f : \{0,1\}^{\mathbb{N}} \to \mathbb{R}$ is *lower semicomputable* if and only if all sets of the form $f^{-1}((a, +\infty))$, with $a \in \mathbb{Q}$, are Σ^0_1, uniformly in a. A function $f : \{0,1\}^{\mathbb{N}} \to \mathbb{R}$ is *upper semicomputable* if and only if all sets of the form $f^{-1}((-\infty, b))$, with $b \in \mathbb{Q}$, are Σ^0_1, uniformly in b.

Semicomputability is the effective analogue of semicontinuity: a function is computable if and only if it is both lower semicomputable and upper semicomputable. The lower semicomputable functions are those whose values can be computably approximated from below, while the upper semicomputable functions are those whose values can be computable approximated from above.[39] The collection of lower semicomputable functions includes the indicator functions of Σ^0_1 sets, while the collection of upper semicomputable functions includes the indicator functions of Π^0_1 sets.

Given that every effective Baire class 1 function is a Baire class 1 function in the classical sense, Theorem 3.12 and Corollary 3.13 also apply to effective Baire class 1 integrable random variables: when the agent's prior has full support and the quantity to be estimated is an effective Baire class 1 integrable random variable, Lévy's Upward Theorem holds on a co-meagre set of data streams.

But, as with the almost everywhere computable functions, adding effectivity into the mix allows to go beyond the mere observation that we can attain co-meagre success. For effective Baire class 1 functions, 1-genericity (Definition 3.28) can be used to provide a more in-depth analysis of convergence to the truth via the following effectivization of Theorem 3.12:

Theorem 3.24 (Kuyper and Terwijn (2014)). Let $f : \{0,1\}^{\mathbb{N}} \to \mathbb{R}$ a function of effective Baire class 1. Then, f is continuous at every 1-generic sequence.

It then immediately follows that, for any agent whose prior has full support, observing a 1-generic data stream is conducive to learning no matter which effective Baire class 1 integrable random variable that agent is trying to estimate:

[39] See (Li and Vitányi, 2019, 35-36).

https://doi.org/10.1017/psa.2023.177 Published online by Cambridge University Press

Corollary 3.25. Let μ a probability measure with full support and $f:\{0,1\}^{\mathbb{N}} \to \mathbb{R}$ an effective Baire class 1 integrable random variable. Then, for every 1-generic $\omega \in \{0,1\}^{\mathbb{N}}$, $\lim_{n\to\infty} \mathbb{E}_{\mu}[f \mid \mathfrak{F}_n](\omega) = f(\omega)$.

Since the collection of 1-generic sequences is co-meagre, Corollary 3.25 allows to pinpoint a single co-meagre collection of data streams that guarantee convergence to the truth for all inductive problems that can be modelled as effective Baire class 1 integrable random variables.

$\emptyset^{(k)}$-Computable Functions and $\emptyset^{(k)}$-Effective Baire Class n Functions. We conclude our discussion of co-meagre convergence to the truth with a more technical note: by considering two collections of functions that rely on oracle computation—the $\emptyset^{(k)}$*-computable* functions and the $\emptyset^{(k)}$*-effective Baire class n* functions.

The arithmetical hierarchy can be relativized to sequences $\omega \in \{0,1\}^{\mathbb{N}}$ (taken to represent the indicator function of a set of natural numbers) by letting the relation R be ω-computable (that is, computable with oracle ω). In this way, one obtains the notions of a $\Sigma_n^{0,\omega}$, $\Pi_n^{0,\omega}$ and $\Delta_n^{0,\omega}$ set. From the perspective of computability theory, an especially useful collection of oracles is the class of *Turing jumps* of the empty set $\emptyset$. The zero-th jump $\emptyset^{(0)}$ of $\emptyset$ is simply $\emptyset$ itself, which of course does not provide any additional computational power. The first jump $\emptyset^{(1)}$ of $\emptyset$ is the *halting set*: namely, the set $\{n \in \mathbb{N} : \varphi_n(n)\downarrow\}$ of all natural numbers n such that the n-the partial computable function φ_n (equivalently, $\varphi_n^{\emptyset^{(0)}}$) is defined on n (the Turing machine computing φ_n halts on input n). For $k > 1$, the k-th jump $\emptyset^{(k)}$ of $\emptyset$ is the halting set relativized to $\emptyset^{(k-1)}$: i.e., the set $\{n \in \mathbb{N} : \varphi_n^{\emptyset^{(k-1)}}(n)\downarrow\}$ of all natural numbers n such that the n-th $\emptyset^{(k-1)}$-partial computable function $\varphi_n^{\emptyset^{(k-1)}}$ is defined on n. Using sets of the form $\emptyset^{(k)}$ (or, rather, the infinite sequences corresponding to these sets) as oracles, one obtains the classes $\Sigma_n^{0,\emptyset^{(k)}}$, $\Pi_n^{0,\emptyset^{(k)}}$ and $\Delta_n^{0,\emptyset^{(k)}}$. For instance, a set $\mathscr{S}$ is $\Sigma_1^{0,\emptyset^{(k)}}$ if and only if there is a Σ_{k+1}^{0} set $S \subseteq \{0,1\}^{<\mathbb{N}}$—namely, a set of strings S that is computably enumerable relative to the k-th jump $\emptyset^{(k)}$ of $\emptyset$—such that $\mathscr{S} = [S]$.[40]

The notion of a $\emptyset^{(k)}$-computable function is defined as follows:

Definition 3.26 ($\emptyset^{(k)}$-Computable function). Let $k \in \mathbb{N}$. A function $f:\{0,1\}^{\mathbb{N}} \to \mathbb{R}$ is $\emptyset^{(k)}$*-computable* if and only if, for every Σ_1^0 subset $\mathscr{U}$ of $\mathbb{R}$, $f^{-1}(\mathscr{U})$ is a $\Sigma_1^{0,\emptyset^{(k)}}$ subset of $\{0,1\}^{\mathbb{N}}$, uniformly in a code for $\mathscr{U}$.

For each $k \geq 1$, the indicator functions of $\Delta_1^{0,\emptyset^{(k)}}$ sets are simple instances of $\emptyset^{(k)}$-computable functions. Functions of this type can be thought of as binary decision problems that are not in themselves decidable, but which become decidable having access to the background information encapsulated by $\emptyset^{(k)}$ (or any other problem of the same complexity). More generally, a $\emptyset^{(k)}$-computable function is one whose values can be computably approximated to any degree of precision with background information $\emptyset^{(k)}$.

[40] See (Soare, 2016, Chapter 3 and Chapter 4).

https://doi.org/10.1017/psa.2023.177 Published online by Cambridge University Press

Crucially, for all $k \in \mathbb{N}$, the $\Sigma_1^{0,\emptyset^{(k)}}$ subsets of $\{0,1\}^{\mathbb{N}}$ are open. Hence, just like the computable functions, the $\emptyset^{(k)}$-computable functions are continuous. By Proposition 3.9, the set of data streams that satisfy Lévy's Upward Theorem relative to a $\emptyset^{(k)}$-computable random variable is thus co-meagre, as long as the agent's prior has full support. And since, for any probability measure μ, every μ-Kurtz random sequence is in the support of μ, observing a μ-Kurtz random data stream suffices to converge to the truth for any $\emptyset^{(k)}$-computable random variable. So, the collection of μ-Kurtz random sequences once again provides a crisp example of a set of data streams along which convergence to the truth is guaranteed to occur.

A $\emptyset^{(k)}$-effective Baire class n function is instead defined as follows:

Definition 3.27 ($\emptyset^{(k)}$-Effective Baire class n function). Let $k, n \in \mathbb{N}$. A function $f : \{0,1\}^{\mathbb{N}} \to \mathbb{R}$ is of *$\emptyset^{(k)}$-effective Baire class n* if and only if, for every Σ_1^0 subset $\mathscr{U}$ of $\mathbb{R}$, $f^{-1}(\mathscr{U})$ is a $\Sigma_{n+1}^{0,\emptyset^{(k)}}$ subset of $\{0,1\}^{\mathbb{N}}$, uniformly in a code for $\mathscr{U}$.

For each $k \geq 1$, the indicator functions of $\Delta_n^{0,\emptyset^{(k)}}$, $\Sigma_n^{0,\emptyset^{(k)}}$, $\Pi_n^{0,\emptyset^{(k)}}$, and $\Delta_{n+1}^{0,\emptyset^{(k)}}$ sets are all of $\emptyset^{(k)}$-effective Baire class n. For instance, the indicator functions of $\Sigma_1^{0,\emptyset^{(k)}}$ sets, which correspond to binary decision problems that can be effectively verified having access to $\emptyset^{(k)}$, are all of $\emptyset^{(k)}$-effective Baire class 1, while the indicator functions of $\Sigma_2^{0,\emptyset^{(k)}}$ sets, which correspond to binary decision problems that can be effectively verified in the limit having access to $\emptyset^{(k)}$, are all of $\emptyset^{(k)}$-effective Baire class 2.

Once again, since every $\emptyset^{(k)}$-effective Baire class 1 function f is a Baire class 1 function, Theorem 3.12 and Corollary 3.13 both apply: the points of continuity of f form a co-meagre $\mathbf{\Pi}_2^0$ set and, for any prior with full support, the collection of data streams along which Lévy's Upward Theorem holds with respect to f (when f is an integrable random variable) is co-meagre. As with effective Baire class 1 functions, we can however say more.

First, note that the notion of 1-genericity introduced earlier can be generalized as follows to any positive natural number:

Definition 3.28 (n-Genericity). Let $n \geq 1$. A sequence $\omega \in \{0,1\}^{\mathbb{N}}$ is *n-generic* if and only if ω meets every Σ_n^0 set $S \subseteq \{0,1\}^{<\mathbb{N}}$ that is dense along ω.

For every $n \geq 1$, $(n+1)$-genericity entails n-genericity, but the reverse implication does not hold. Moreover, the fact that the set of 1-generic sequences is topologically typical generalizes to the entire hierarchy:[41]

Proposition 3.29. Let $n \geq 1$. The set of n-generic sequences is co-meagre.

The following is proven analogously to Theorem 3.24:

Theorem 3.30. Let $k \geq 1$ and $f : \{0,1\}^{\mathbb{N}} \to \mathbb{R}$ a function of $\emptyset^{(k-1)}$-effective Baire class 1. Then, f is continuous at every k-generic sequence.

[41] See (Kurtz, 1981), (Nies, 2009), or (Downey and Hirschfeldt, 2010).

https://doi.org/10.1017/psa.2023.177 Published online by Cambridge University Press

Proof. Let $\mathscr{S}_0, \mathscr{S}_1, \mathscr{S}_2, \ldots$ be a fixed effective enumeration of the Σ^0_1 subsets of $\mathbb{R}$. Recall that f being discontinuous at $\omega \in \{0,1\}^{\mathbb{N}}$ means that there is an open subset $\mathscr{O}$ of $\mathbb{R}$ with $f(\omega) \in \mathscr{O}$ but, for all open $\mathscr{U} \subseteq f^{-1}(\mathscr{O})$, $\omega \notin \mathscr{U}$. In other words, $\omega \in f^{-1}(\mathscr{O})$, but $\omega \notin \mathrm{Int}(f^{-1}(\mathscr{O}))$. Every open set is a union of Σ^0_1 sets, so $\{\omega \in \{0,1\}^{\mathbb{N}} : f \text{ is discontinuous at } \omega\} = \bigcup_{n\in\mathbb{N}} \left(f^{-1}(\mathscr{S}_n) \setminus \mathrm{Int}(f^{-1}(\mathscr{S}_n))\right)$. Let $\omega \in \{0,1\}^{\mathbb{N}}$ such that f is discontinuous at ω. Then, there is $m \in \mathbb{N}$ with $\omega \in f^{-1}(\mathscr{S}_m) \setminus \mathrm{Int}(f^{-1}(\mathscr{S}_m))$. Given that f is of $\emptyset^{(k-1)}$-effective Baire class 1, by definition, $f^{-1}(\mathscr{S}_m)$ is a $\Sigma^{0,\emptyset^{(k-1)}}_2$ subset of $\{0,1\}^{\mathbb{N}}$. Hence, $f^{-1}(\mathscr{S}_m) = \bigcup_{i\in\mathbb{N}} \mathscr{C}_i$, where each $\mathscr{C}_i$ is a $\Pi^{0,\emptyset^{(k-1)}}_1$ set. So, $f^{-1}(\mathscr{S}_m) \setminus \mathrm{Int}(f^{-1}(\mathscr{S}_m)) = \bigcup_{i\in\mathbb{N}} \mathscr{C}_i \setminus \mathrm{Int}(\bigcup_{i\in\mathbb{N}} \mathscr{C}_i) \subseteq \bigcup_{i\in\mathbb{N}} (\mathscr{C}_i \setminus \mathrm{Int}(\mathscr{C}_i))$. Thus, there is $j \in \mathbb{N}$ with $\omega \in \mathscr{C}_j \setminus \mathrm{Int}(\mathscr{C}_j)$, which means that $\omega \in \mathscr{C}_j \cap \mathrm{Cl}(\overline{\mathscr{C}_j})$. Since $\overline{\mathscr{C}_j}$ is a $\Sigma^{0,\emptyset^{(k-1)}}_1$ set, $\overline{\mathscr{C}_j} = [C]$ for some Σ^0_k set $C \subseteq \{0,1\}^{<\mathbb{N}}$. And since $\omega \in \mathrm{Cl}([C])$, C is dense along ω. However, ω does not meet C, as $\omega \notin [C] = \overline{\mathscr{C}_j}$. Therefore, ω is not k-generic. □

Theorem 3.30 entails that, for any Bayesian reasoner whose prior has full support, observing a k-generic data stream leads to inductive success for any inductive problem corresponding to a $\emptyset^{(k-1)}$-effective Baire class 1 integrable random variable:

Corollary 3.31. Let $k \geq 1$, μ a probability measure with full support, and $f : \{0,1\}^{\mathbb{N}} \to \mathbb{R}$ a $\emptyset^{(k-1)}$-effective Baire class 1 integrable random variable. Then, for every k-generic $\omega \in \{0,1\}^{\mathbb{N}}$, $\lim_{n\to\infty} \mathbb{E}_{\mu}[f \mid \mathfrak{F}_n](\omega) = f(\omega)$.

As noted earlier, the set of k-generic sequences is co-meagre. Hence, Corollary 3.31 reveals that, for priors with full support, we can once again single out a specific co-meagre collection of data streams along which convergence to the truth happens for all $\emptyset^{(k-1)}$-effective Baire class 1 integrable random variables.

4. Conclusion

According to Belot (2013), Bayesian learners are unavoidably epistemically orgulous: the Bayesian framework, with its convergence-to-the truth results, compels them to be confident in their ability to be inductively successful even when there are co-meagre many data streams along which learning, as a matter of fact, fails.

In this article, we set out to elucidate how pervasive the issue identified by Belot is. We suggested to use descriptive set theory and computability theory to classify the inductive problems (random variables) faced by Bayesian agents in terms of their complexity. Then, relying on this taxonomy, we provided an analysis of the conditions under which inductive success is both probabilistically and topologically typical and the conditions under which these two notions of typicality instead come apart.

We showed that there are several classes of random variables admitting natural epistemic interpretations for which the dichotomy Belot highlights does not arise: for the inductive problems in these classes, Lévy's Upward Theorem holds both with probability one (relative to the agent's prior) and on a co-meagre set of data streams. Specifically, the collections of inductive problems for which we established that success is topologically typical, in addition to being probabilistically typical, are the classes of random variables listed in Table 1 below.

https://doi.org/10.1017/psa.2023.177 Published online by Cambridge University Press

Classical setting	Effective setting
Continuous random variables	Computable random variables Almost everywhere computable random variables $\emptyset^{(k)}$-Computable random variables
Baire class 1 integrable random variables	Effective Baire class 1 integrable random variables $\emptyset^{(k)}$-Effective Baire class 1 integrable random variables

Table 1. Classes of random variables for which convergence to the truth occurs on a co-meagre set.

The random variables for which we proved that convergence to the truth happens on a co-meagre set correspond to natural but relatively simple inductive problems (they include, for instance, all binary decision problems that can be verified or refuted with a finite amount of data and all binary decision problems that can be decided in the limit). For more complex families of inductive problems (in fact, for the entire hierarchy of Baire class n integrable random variables starting at level 2), we showed that there are problems for which convergence to the truth only happens on a meagre set. Hence, the proposed taxonomy can also be leveraged to add to Belot's negative results and reveal that "Bayesian orgulity" is, relative to this classification at least, a pervasive phenomenon.

Classical notions of measure-theoretic and topological typicality can be used to prove that convergence to the truth happens along the "vast majority" of data streams, but they convey little information as to what kind of data streams are conducive to inductive learning, depending on the particular inductive problem at hand. We saw that, in the effective setting, it is possible to get a much crisper understanding of the success sets of Bayesian agents. In particular, we showed that the theories of algorithmic randomness and effective genericity (which are theories of effective measure-theoretic typicality and effective topological typicality, respectively) can be employed to single out specific co-meagre collections of data-streams along which Lévy's Upward Theorem holds, no matter which inductive problem from the classes of effective random variables listed in Table 1 the agent is trying to solve.

Our findings, while preliminary, evince that the taxonomy of inductive problems afforded by descriptive set theory and computability theory is a promising lens through which to probe Bayesian convergence-to-the-to-truth theorems. In particular, they suggest that, by further analyzing the inductive problems faced by Bayesian learners in terms of their complexity, we may be able to come to understand the full reach of Belot's objection. Moreover, quite aside from Belot's concerns, the approach adopted in this article also offers a natural framework within which to investigate the following general question: how does the complexity of a Bayesian learner's success set, understood in either topological or computability-theoretic terms, vary as a function of the complexity of the inductive problem faced by the learner?

Acknowledgements. I am especially indebted to Krzysztof Mierzewski for his extensive and insightful comments on various iterations of this article. I am also grateful to the two anonymous referees for their valuable suggestions, and to Gordon Belot, Kevin Dorst, Simon Huttegger, Kevin Kelly, Teddy Seidenfeld, Sean Walsh, Snow Zhang, and Kevin Zollman for many illuminating questions and discussions. Lastly, I would like to thank the participants of the Logic Seminar at Peking University, the Formal Rationality Forum, the Logic and Philosophy of Science Group at the University of Toronto, and the session "What's

New in Formal Epistemology?" at the *17th Congress on Logic, Methodology, and Philosophy of Science and Technology* for their helpful questions.

References

Ackerman, Nathanael L., Cameron E. Freer, and Daniel M. Roy. 2019. "On the Computability of Conditional Probability." *Journal of the ACM* 66 (3): 1–40. doi:10.1145/3321699.

Belot, Gordon. 2013. "Bayesian Orgulity." *Philosophy of Science* 80 (4): 483–503. doi:10.1086/673249.

Belot, Gordon. 2017. "Objectivity and Bias." *Mind* 126 (503): 655–95. doi:10.1093/mind/fzv185.

Cisewski, Jessi, Joseph B. Kadane, Mark J. Schervish, Teddy Seidenfeld, and Rafael B. Stern. 2018. "Standards for Modest Bayesian Credences." *Philosophy of Science* 85 (1): 53–78. doi:10.1086/694836.

Downey, Rod G., and Dennis R. Hirschfeldt. 2010. *Algorithmic Randomness and Complexity.* New York, NY: Springer. doi:10.1007/978-0-387-68441-3.

Earman, John. 1992. *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory.* Cambridge, MA: The MIT Press.

Edwards, Ward, Harold Lindman, and Leonard J. Savage. 1963. "Bayesian Statistical Inference for Psychological Research." *Psychological Review* 70 (3): 193–242. doi:10.1037/h0044139.

Elga, Adam. 2016. "Bayesian Humility." *Philosophy of Science* 83 (3): 305–23. doi:10.1086/685740.

Freedman, David A. 1963. "On the Asymptotic Behavior of Bayes' Estimates in the Discrete Case." *Annals of Mathematical Statistics* 34 (4): 1386–1403. doi:10.1214/aoms/1177703871.

Freedman, David A. 1965. "On the Asymptotic Behavior of Bayes' Estimates in the Discrete Case II." *Annals of Mathematical Statistics* 36 (2): 454–56. doi:10.1214/aoms/1177700155.

Glymour, Clark. 1980. *Theory and Evidence.* Princeton: Princeton University Press.

Gong, Ruobin, Joseph B. Kadane, Mark J. Schervish, Teddy Seidenfeld, and Rafael B. Stern. 2021. "Deceptive Credences." *Ergo* 7 (28): 757–73. doi:10.3998/ergo.1125.

Hoyrup, Mathieu. 2008. "Randomness and Ergodic Theory on Metric Spaces." PhD diss., Université Paris Diderot (Paris VII).

Hoyrup, Mathieu, and Cristobal Rojas. 2009. "Computability of Probability Measures and Martin-Löf Randomness over Metric Spaces." *Information and Computation* 207 (7): 830–47. doi:10.1016/j.ic.2008.12.009.

Huttegger, Simon M. 2015. "Bayesian Convergence to the Truth and the Metaphysics of Possible Worlds." *Philosophy of Science* 82 (4): 587–601. doi:10.1086/682941.

Huttegger, Simon M. 2017. *The Probabilistic Foundations of Rational Learning.* Cambridge: Cambridge University Press. doi:10.1017/9781316335789.

Huttegger, Simon M., Sean Walsh, and Francesca Zaffora Blando. 2023. "Algorithmic Randomness, Effective Disintegrations, and Rates of Convergence to the Truth." Manuscript in preparation.

Kechris, Alexander S. 1995. *Classical Descriptive Set Theory.* New York, NY: Springer. doi:10.1007/978-1-4612-4190-4.

Kelly, Kevin T. 1996. *The Logic of Reliable Inquiry.* Oxford: Oxford University Press. doi:10.1093/oso/9780195091953.001.0001.

Kjos-Hanssen, Bjørn. 2010. "The Probability Distribution as a Computational Resource for Randomness Testing." *Journal of Logic and Analysis* 2 (10): 1–13. doi:10.4115/jla.2010.2.10.

Kurtz, Stuart A. 1981. "Randomness and Genericity in the Degrees of Unsolvability." PhD diss., University of Illinois at Urbana-Champaign.

Kuyper, Rutger, and Sebastiaan A. Terwijn. 2014. "Effective Genericity and Differentiability." *Journal of Logic and Analysis* 6 (4): 1–14. doi:10.4115/jla.2014.6.4.

Lévy, Paul P. 1937. *Théorie de l'Addition des Variables Aléatoires.* Paris: Gauthier-Villars.

Li, Ming, and Paul M. B. Vitányi. 2019. *An Introduction to Kolmogorov Complexity and Its Applications.* Fourth edition. Cham: Springer. doi:10.1007/978-3-030-11298-1.

Martin-Löf, Per. 1966. "The Definition of Random Sequences." *Information and Control* 9 (6): 602–19. doi:10.1016/S0019-9958(66)80018-9.

Nielsen, Michael, and Rush T. Stewart. 2019. "Obligation, Permission, and Bayesian Orgulity." *Ergo* 6 (3): 58–70. doi:10.3998/ergo.12405314.0006.003.

Nies, André. 2009. *Computability and Randomness.* Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199230761.001.0001.

Oxtoby, John C. 1980. *Measure and Category: A Survey of the Analogies between Topological and Measure Spaces.* New York, NY: Springer. doi:10.1007/978-1-4684-9339-9.

Pomatto, Luciano, and Alvaro Sandroni. 2018. "An Axiomatic Theory of Inductive Inference." *Philosophy of Science* 85 (2): 293–315. doi:10.1086/696386.

Savage, Leonard J. 1954. *The Foundations of Statistics.* New York, NY: John Wiley & Sons.

Skyrms, Brian. 1984. *Pragmatics and Empiricism.* New Haven and London: Yale University Press.

Soare, Robert I. 2016. *Turing Computability: Theory and Applications.* Berlin, Heidelberg: Springer. doi:10.1007/978-3-642-31933-4.

Weatherson, Brian. 2015. "For Bayesians, Rational Modesty Requires Imprecision." *Ergo* 2 (20): 529–45. doi:10.3998/ergo.12405314.0002.020.

Williams, David. 1991. *Probability with Martingales.* Cambridge: Cambridge University Press. doi:10.1017/CBO9780511813658.

Zaffora Blando, Francesca. 2020. "Patterns and Probabilities: A Study in Algorithmic Randomness and Computable Learning." PhD diss., Stanford University.