Bayesian merging of opinions and algorithmic randomness

Francesca Zaffora Blando

Abstract

We study the phenomenon of merging of opinions for computationally limited Bayesian agents from the perspective of algorithmic randomness. When they agree on which data streams are algorithmically random, two Bayesian agents beginning the learning process with different priors may be seen as having compatible beliefs about the global uniformity of nature. This is because the algorithmically random data streams are of necessity globally regular: they are precisely the sequences that satisfy certain important statistical laws. By virtue of agreeing on what data streams are algorithmically random, two Bayesian agents can thus be taken to concur on what global regularities they expect to see in the data. We show that this type of compatibility between priors suffices to ensure that two computable Bayesian agents will reach inter-subjective agreement with increasing information. In other words, it guarantees that their respective probability assignments will almost surely become arbitrarily close to each other as the number of observations increases. Thus, when shared by computable Bayesian learners with different subjective priors, the beliefs about uniformity captured by algorithmic randomness provably lead to merging of opinions.

1 Introduction

Bayesian learning encompasses a family of probabilistic inference methods that crucially rely on prior probability distributions, which are meant to encapsulate the learner's background knowledge and inductive assumptions before the beginning of the learning process—for instance, before performing an experiment. This reliance on priors is often taken to be a cause for concern: when the available background knowledge does not suffice to reach inter-subjective agreement about which prior should be adopted, how can one be possibly guaranteed that the inferences drawn on the basis of one's own subjective prior provide any objective epistemic warrant? Perhaps most alarmingly, does the use of Bayesian methods—and, thus, of prior probability distributions—in the sciences threaten the objectivity of scientific inquiry?

According to objective Bayesians, such as Jaynes (1968) and Rosenkrantz (1981), this problem can be overcome by singling out the class of rationally permissible priors, the adoption of which ensures the objectivity of the conclusions derived from them. For instance, some objective Bayesians might contend that symmetry considerations play a crucial role in fixing the collection of rationally permissible priors, or that priors should be calibrated with known frequencies. On the other hand, subjective Bayesians such as Ramsey (1931), de Finetti (1937), Savage (1954), and Jeffrey (1977) maintain that the only requirement that rationality imposes on prior probability distributions is probabilistic coherence, and that there is no principled way of arguing for the superiority of any particular prior over another.

To rebuke accusations of excessive subjectivity, subjective Bayesians often appeal to various results from probability theory and measure theory that are meant to show that a Bayesian agent's initial beliefs or assumptions, in the form of a prior probability distribution, are *de facto* immaterial for the purpose of successful inquiry: the dynamics of Bayesian conditioning by themselves ensure that priors are eventually washed out by the shared evidence. Suppes, for instance, claims that

> It is of fundamental importance to any deep appreciation of the Bayesian viewpoint to realize the particular form of the prior distribution expressing beliefs held before the experiment is conducted is not a crucial matter. [...] The well-designed experiment is one that will swamp divergent prior distributions with the clarity and sharpness of its results, and thereby render insignificant the diversity of prior opinion. (Suppes 1966, p. 204)

Similarly, Edwards, Lindman, and Savage maintain that

> Although your initial opinion about future behavior [...] may differ radically from your neighbor's, your opinions and his will ordinarily be so transformed by application of Bayes' theorem [...] as to become nearly indistinguishable. This approximate merging of initially divergent opinions is, we think, one reason why empirical research is called 'objective'. (Edwards et al. 1963, p. 197)

The theorems used to argue that initial diversity of opinions is immaterial have their roots in Savage's work (Savage 1954), and they fall under the umbrella of 'Bayesian merging-of-opinions theorems'. Roughly put, these results establish that, provided that their respective subjective priors are sufficiently compatible, two Bayesian agents beginning the learning process with different beliefs are guaranteed to almost surely reach a consensus: as the number of observations increases, their beliefs, in the form of their posterior probability distributions, will become arbitrarily close to each other with probability one (relative to the agents' priors).[1]

In this article, we explore the phenomenon of merging of opinions in the context of more realistic, less-than-ideal agents. We do so by bringing into play the theory of computation: that is, by focusing on computationally limited Bayesian agents whose subjective priors are computable probability measures. The key idea is that merging of opinions for computationally limited Bayesian agents—computable Bayesian agents, for short—can be studied through the prism of algorithmic randomness: a branch of computability theory aimed at formalizing the notion of an individual mathematical object (such as a real number or a binary string—and, in our case, a sequence of observations, a data stream) displaying no patterns or regularities discernible using algorithmic means. In particular, we will see that algorithmic randomness can be employed to define refined notions of compatibility between priors. Given an algorithmic randomness notion R, the beliefs of two computable Bayesian agents will be said to be compatible relative to R if their respective computable priors agree on which data streams are R-random. More precisely, given two computable priors μ and ν, ν will be said to be compatible with μ with respect to R if the collection of R-random data streams relative to μ is a subset of the collection of R-random data streams relative to ν.

The rationale for using algorithmic randomness to define notions of compatibility between priors is that the algorithmically random data streams, though maximally irregular and patternless when considered locally, bit by bit, are of necessity globally regular. As will be explained in §3, in spite of being unpredictable (observing a finite initial segment of an algorithmically random data stream does not provide any useful information for predicting what the next observation is going to be), the algorithmically random sequences must nonetheless satisfy various effectively specifiable statistical laws (such as the Strong Law of Large Numbers when the underlying probability measure is, for instance, a Bernoulli measure). From this perspective, different notions of algorithmic randomness may be seen as encoding different beliefs about the global uniformity of nature: each algorithmic randomness notion corresponds, from the viewpoint of the computable Bayesian agent with respect to whom that randomness notion is defined, to a precise class of effectively specifiable global regularities. So, when two computable Bayesian agents agree on what data streams are algorithmically random, they can be seen as having compatible inductive assumptions, as having compatible beliefs (or commitments) about which effectively specifiable statistical properties they expect to see in the data.

[1] As shown by Schervish and Seidenfeld (1990), these results are generalizable to the case where there are multiple Bayesian agents.

In other words, they can be seen as concurring on the extent of nature's global uniformity, where the type of uniformity over which they agree amounts to the satisfaction of certain effectively specifiable statistical laws.

We will then see that agreeing on which data streams are algorithmically random provably leads to merging of opinions. Our main results establish that, when shared by computable Bayesian agents with differing subjective priors, the inductive assumptions pertaining to the global uniformity of nature encoded by algorithmic randomness notions guarantee the eventual (almost-sure) attainment of inter-subjective agreement: in other words, for any two computable priors μ and ν and canonical algorithmic randomness notion R, ν being compatible with μ with respect to R ensures that ν will merge with μ as the learning process unfolds.

The study of the equivalence relations between probability measures induced by algorithmic randomness notions[2] has already received some attention in the literature (see (Muchnik et al. 1998) and (Bienvenu and Merkle 2009)). We will review and make use of some of these results in what follows. Our aim in this article is to bridge the theory of algorithmic randomness and the literature on merging of opinions by connecting algorithmic randomness to the study of canonical notions of compatibility (and incompatibility) between subjective priors and, most importantly, by showing that the notions of compatibility induced by algorithmic randomness entail merging of opinions. In doing so, we hope to lay the foundations for the systematic study of the fruitful interactions between, on the one hand, the theory of algorithmic randomness and, on the other, work in Bayesian epistemology and probability theory on merging of opinions, Bayesian learning, and their philosophical ramifications.

The remainder of this article is structured as follows. In §2.1, we review some canonical notions of agreement and disagreement between priors. In §2.2, we discuss merging of opinions in the classical (non-effective) setting and, in particular, what is arguably the most prominent merging-of-opinions result: the Blackwell-Dubins Theorem (Blackwell and Dubins 1962). We also consider the phenomenon opposite to merging of opinions: namely, polarization of opinions. Our main results are in §3. We begin with a brief overview of the theory of algorithmic randomness in §3.1. In §3.2, we explore the relations between standard notions of compatibility/incompatibility and the notions of agreement induced by algorithmic randomness. Then, we show that all core algorithmic randomness notions except for weak 1-randomness generate notions of compatibility that lead to merging in the sense of Blackwell and Dubins. In §3.3, we conclude with some simple observations about how disagreement on which data streams are algorithmically random leads to polarization of opinions.

[2]Two computable probability measures μ and ν are said to be equivalent relative to some algorithmic randomness notion R if the collection of R-random sequences relative to μ coincides with the collection of R-random sequences relative to ν—in other words, μ and ν are compatible with each other relative to R.

2 Merging of opinions

Recall that a probability space $(\Omega, \mathcal{E}, \mu)$ is a triple consisting of a set Ω, a σ-algebra $\mathcal{E}$ on Ω, and a probability measure μ on $\mathcal{E}$ (since the only measures that we will be dealing with are probability measures, we will simply call them measures from now on). The set Ω is the sample space: the collection of all possible basic outcomes of a given experiment—or, more generally, the collection of all possible observational data associated with the inductive problem under consideration. The σ-algebra $\mathcal{E}$, on the other hand, corresponds to the collection of all events (involving observational data) that get assigned a probability: intuitively, all of the events that a Bayesian learner entertains in the given situation.

We are interested in situations where the same space comes equipped with two measures μ and ν. These measures are amenable to multiple interpretations. In this article, we will mostly be concerned with the case where both μ and ν represent subjective priors. Of course, measures admit a non-personalist interpretation, as well: they can be taken to encode objective distributions. For instance, a measure may be seen as representing the true chance distribution governing some stochastic process (such as a game of chance). Merging-of-opinions theorems apply in this context, too. When one of the measures involved is an objective chance distribution while the other one is the subjective prior of a Bayesian agent, these results establish that, with increasing information, the agent's beliefs will asymptotically align with the true chances with objective probability one, provided that, to begin with, the agent's beliefs are sufficiently compatible with the truth. So, in what follows, μ may also be taken to represent the true distribution governing some process, and ν to be the subjective prior of a Bayesian agent trying to approximate that distribution.

In the setting of merging-of-opinions theorems, bodies of evidence are usually modelled in terms of σ-algebras. In particular, increasing bodies of evidence can be naturally represented as filtrations on $(\Omega, \mathcal{E})$: namely, as sequences $\{\mathcal{E}_n\}_{n\in\mathbb{N}}$ of sub-σ-algebras of $\mathcal{E}$ such that $\mathcal{E}_n \subseteq \mathcal{E}_{n+1}$ for all $n \in \mathbb{N}$—where the latter condition ensures that, for each n, the information embodied by $\mathcal{E}_{n+1}$ refines the information embodied by $\mathcal{E}_n$. Given a filtration $\{\mathcal{E}_n\}_{n\in\mathbb{N}}$ on $(\Omega, \mathcal{E})$, let $\mathcal{E}^\infty$ denote the σ-algebra $\sigma(\bigcup_{n\in\mathbb{N}} \mathcal{E}_n)$ generated by the union of the $\mathcal{E}_n$'s. If $\mathcal{E}^\infty = \mathcal{E}$, then the filtration is complete: the cumulating evidence will eventually settle the truth of every event that a Bayesian agent can entertain.

Learning occurs by conditionalizing on the (total) available evidence. Since, in this setting, the growing evidence is encapsulated by a filtration, we need to define conditional probabilities given a sub-σ-algebra. Fix a filtration $\{\mathcal{E}_n\}_{n\in\mathbb{N}}$ on $(\Omega, \mathcal{E})$, an event $A \in \mathcal{E}$, and a prior μ. The conditional probability $\mu(A \mid \mathcal{E}_n)$ of A given $\mathcal{E}_n$ is an $\mathcal{E}_n$-measurable function $\mu(A \mid \mathcal{E}_n) : \Omega \to \mathbb{R}$ (a random variable) such that, for all $B \in \mathcal{E}_n$, $\mu(A \cap B) = \int_B \mu(A \mid \mathcal{E}_n)\, d\mu$. By the Radon-Nikodym Theorem, such a function exists for any sub-σ-algebra and is unique up to sets of μ-measure zero.[3]

[3]The standard definition of conditional probability given by Bayes' formula only applies to

In what follows, as often done in the literature, we shall focus on the space $(2^{\mathbb{N}}, \mathcal{B}(2^{\mathbb{N}}))$, where $2^{\mathbb{N}}$ is the set of one-way countably infinite binary sequences and $\mathcal{B}(2^{\mathbb{N}})$ is the Borel σ-algebra on $2^{\mathbb{N}}$. We will think of sequences in $2^{\mathbb{N}}$ as data streams. The set $2^{\mathbb{N}}$ comes equipped with a natural topology: the topology of pointwise convergence. A basis for this topology is given by the clopen cylinders $[\sigma]$, where σ denotes a finite binary string in $2^{<\mathbb{N}}$ and $[\sigma] \subseteq 2^{\mathbb{N}}$ is the set of all infinite sequences that extend σ. The resulting topological space is called Cantor space. The Borel σ-algebra $\mathcal{B}(2^{\mathbb{N}})$ is the smallest σ-algebra on $2^{\mathbb{N}}$ containing all open sets from this topology. For each $\omega \in 2^{\mathbb{N}}$, $\omega \upharpoonright n$ will denote the initial segment of ω of length n and $[\omega \upharpoonright n]$ the cylinder generated by the string $\omega \upharpoonright n$.

We will work with measures μ, ν on $\mathcal{B}(2^{\mathbb{N}})$. Let ε denote the empty string, so that $[\varepsilon] = 2^{\mathbb{N}}$. By Carathéodory's Extension Theorem,[4] any function m defined on cylinders that takes values in $[0, 1]$, and such that $m([\varepsilon]) = 1$ and, for all $\sigma \in 2^{<\mathbb{N}}$, $m([\sigma]) = m([\sigma 0]) + m([\sigma 1])$ can be uniquely extended to a measure on $\mathcal{B}(2^{\mathbb{N}})$. Hence, from now on, measures on $\mathcal{B}(2^{\mathbb{N}})$ will be identified with their restriction to cylinders. A canonical measure on $\mathcal{B}(2^{\mathbb{N}})$ is the uniform (or Lebsegue) measure λ, given by $\lambda([\sigma]) = 2^{-|\sigma|}$ for all $\sigma \in 2^{<\mathbb{N}}$, where $|\sigma|$ denotes the length of σ. We will also make use of the notion of a continuous semimeasure (semimeasure, for short): namely, a function δ defined on cylinders and taking values in $[0, 1]$ such that $\delta([\varepsilon]) \leq 1$ and, for all $\sigma \in 2^{<\mathbb{N}}$, $\delta([\sigma]) \geq \delta([\sigma 0]) + \delta([\sigma 1])$.

In addition, we will restrict attention to the filtration $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$, where, for each n, $\mathcal{F}_n$ is the sub-σ-algebra of $\mathcal{B}(2^{\mathbb{N}})$ generated by the cylinders $[\sigma]$ with $\sigma \in 2^{<\mathbb{N}}$ a string of length n. Each such algebra is induced by a finite partition of $2^{\mathbb{N}}$ (for instance, $\mathcal{F}_1$ is the algebra $\{\emptyset, [0], [1], 2^{\mathbb{N}}\}$ induced by the finite partition $\{[0], [1]\}$). Intuitively, $\mathcal{F}_n$ represents all possible evidential situations that a Bayesian agent may find themselves in at the n-th stage of the learning process, after having made n observations (after having observed the first n digits of the true data stream). Since $\sigma(\bigcup_{n\in\mathbb{N}} \mathcal{F}_n) = \mathcal{B}(2^{\mathbb{N}})$, this filtration is complete. We will thus assume throughout that the evidence is both increasing and complete.

Given that the $\mathcal{F}_n$'s are generated by finite partitions, learning in this setting essentially proceeds by standard Bayesian conditioning. We can in fact almost surely recover the familiar definition of conditional probability as follows: for any $\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})$, $n \in \mathbb{N}$, and μ-almost every $\omega \in 2^{\mathbb{N}}$,

$$\mu(\mathcal{S} \mid [\omega \upharpoonright n]) = \frac{\mu(\mathcal{S} \cap [\omega \upharpoonright n])}{\mu([\omega \upharpoonright n])} = \frac{1}{\mu([\omega \upharpoonright n])} \int_{[\omega \upharpoonright n]} \mu(\mathcal{S} \mid \mathcal{F}_n)\, d\mu = \mu(\mathcal{S} \mid \mathcal{F}_n)(\omega),$$

---

cases where the conditioning event has positive probability. Defining conditional probabilities in this more general setting allows to define conditionalization with respect to probability zero events, as well. This definition also ensures that $\mu(A \mid \mathcal{E}_n)$ is a version of the conditional expectation $\mathbb{E}_\mu[\chi_A \mid \mathcal{E}_n]$ of the indicator function χ_A of $A \in \mathcal{E}$.

[4]See, for instance, (Williams 1991, Theorem 1.7, p. 20). This result is also known as the Hahn-Kolmogorov Extension Theorem.

where the second identity follows from the definition of $\mu(\mathcal{S} \mid \mathcal{F}_n)$ and the last identity from the fact that the value of $\mu(\mathcal{S} \mid \mathcal{F}_n)$ is constant within the partition cells generating $\mathcal{F}_n$—and so, in particular, within the cylinder $[\omega \upharpoonright n]$.

2.1 Classical notions of compatibility/incompatibility between measures

We begin by reviewing some classical notions of compatibility (and incompatibility) between measures, which will serve as a springboard for our study of compatibility notions induced by algorithmic randomness.

Arguably, the most well-studied form of compatibility between measures is absolute continuity:

Definition 2.1 (Absolute continuity)**.** *Given measures μ and ν, μ is absolutely continuous with respect to ν ($\mu \ll \nu$) if, for every event $\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})$, $\mu(\mathcal{S}) > 0$ entails that $\nu(\mathcal{S}) > 0$.*

If μ and ν encode the subjective priors of two Bayesian agents, then μ being absolutely continuous with respect to ν intuitively means that μ is at least as dogmatic as ν. All of the events that are *a priori* 'excluded' by the agent with prior ν (by virtue of having been assigned probability zero before any observations are made) are also 'excluded' by the agent with prior μ. In other words, the agent with prior ν cannot be surprised by any event to which the agent with prior μ assigns positive probability. It is however possible for μ to be strictly more dogmatic than ν: the agent with prior μ may assign probability zero to some events to which the agent with prior ν assigns positive probability. On the other hand, if μ represents the true distribution governing some stochastic process while ν is the subjective prior of a Bayesian agent, then μ being absolutely continuous with respect to ν means that the agent with prior ν assigns probability zero only to events that truly have probability zero.

Here is an example to elucidate this notion.

Example 2.2. *If ν is a non-trivial convex combination of μ_1 and μ_2, then $\mu_1 \ll \nu$ and $\mu_2 \ll \nu$. Take, for instance, the uniform measure λ and let $\nu = \frac{1}{4}\lambda + \frac{3}{4}\mu_{\frac{1}{3}}$, where $\mu_{\frac{1}{3}}$ is the Bernoulli measure given by $\mu_{\frac{1}{3}}([\sigma]) = \frac{1}{3}^k \cdot \frac{2}{3}^{n-k}$, with n the length of σ, k the number of 0's occurring in σ and $n-k$ the number of 1's occurring in σ. Measure ν is a convex combination of λ and $\mu_{\frac{1}{3}}$. Now, let $\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})$ with $\lambda(\mathcal{S}) > 0$. Then, $\frac{1}{4}\lambda(\mathcal{S}) > 0$ and $\frac{3}{4}\mu_{\frac{1}{3}}(\mathcal{S}) \geq 0$, which together entail that $\nu(\mathcal{S}) > 0$. Hence, $\lambda \ll \nu$. An analogous argument establishes that $\mu_{\frac{1}{3}} \ll \nu$.*

Next, we consider a canonical notion of incompatibility between measures, as well as its dual notion, which yields a very minimal form of compatibility.

Definition 2.3 (Orthogonality)**.** *Two measures μ and ν are orthogonal ($\mu \perp \nu$) if there is an event $\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})$ such that $\mu(\mathcal{S}) = 1$ but $\nu(\mathcal{S}) = 0$. If there is no such event, then μ and ν are said to be non-orthogonal ($\mu \not\perp \nu$).*

Orthogonality is diametrically opposed to absolute continuity. If μ and ν are orthogonal, then absolute continuity fails in the most extreme way possible: the event with (without loss of generality) ν-measure zero and positive μ-measure witnessing the failure of absolute continuity has in fact μ-measure one. Orthogonality thus captures a radical type of disagreement.

Below is an example of two orthogonal measures, followed by an example featuring non-orthogonal measures.

Example 2.4. *Take the uniform measure λ and the Bernoulli measure $\mu_{\frac{1}{3}}$ from Example 2.2. Given a sequence $\omega \in 2^{\mathbb{N}}$, let $\frac{\#0(\omega\upharpoonright n)}{n}$ denote the relative frequency of 0 in the first n digits of ω. By the Strong Law of Large Numbers, $\lambda\Big(\Big\{\omega \in 2^{\mathbb{N}} : \lim_{n\to\infty} \frac{\#0(\omega\upharpoonright n)}{n} = \frac{1}{2}\Big\}\Big) = 1$ and $\mu_{\frac{1}{3}}\Big(\Big\{\omega \in 2^{\mathbb{N}} : \lim_{n\to\infty} \frac{\#0(\omega\upharpoonright n)}{n} = \frac{1}{3}\Big\}\Big) = 1$. So, $\mu_{\frac{1}{3}}\Big(\Big\{\omega \in 2^{\mathbb{N}} : \lim_{n\to\infty} \frac{\#0(\omega\upharpoonright n)}{n} = \frac{1}{2}\Big\}\Big) = 0$, which shows that λ and $\mu_{\frac{1}{3}}$ are orthogonal.*

Example 2.5. *Let $\nu = \alpha\mu_1 + (1-\alpha)\mu_2$, with $\alpha \in (0,1)$. Let $\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})$. If $\nu(\mathcal{S}) = \alpha\mu_1(\mathcal{S}) + (1-\alpha)\mu_2(\mathcal{S}) = 1$, then $\mu_1(\mathcal{S}) = 1$ and $\mu_2(\mathcal{S}) = 1$. If $\mu_1(\mathcal{S}) = 1$, then $\nu(\mathcal{S}) \geq \alpha > 0$, and if $\mu_2(\mathcal{S}) = 1$, then $\nu(\mathcal{S}) \geq (1-\alpha) > 0$. Hence, ν and μ_1 are non-orthogonal, and ν and μ_2 are non-orthogonal.*

If $\mu \ll \nu$, then μ and ν are non-orthogonal. The converse, however, does not hold: non-orthogonality is a much weaker form of compatibility than absolute continuity.[5]

We conclude our review of classical compatibility notions by discussing a weaker form of absolute continuity: local absolute continuity.

Definition 2.6 (Local absolute continuity)**.** *Given measures μ and ν, μ is locally absolutely continuous with respect to ν ($\mu \ll_{loc} \nu$) if, for every $n \in \mathbb{N}$ and every $\mathcal{S} \in \mathcal{F}_n$, $\mu(\mathcal{S}) > 0$ entails that $\nu(\mathcal{S}) > 0$.*[6]

Since the filtration $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$ represents the possible evidence the agents may obtain, having that $\mu \ll_{loc} \nu$ means that ν agrees with μ about which evidence they expect to see. In other words, ν cannot be surprised by any piece of evidence to which μ assigns positive probability. Another way to think about local absolute continuity is that it amounts to absolute continuity restricted to finite-horizon events—that is, events that can be settled by a finite amount of evidence.

A measure μ is strictly positive if it assigns positive probability to every basic open set: namely, if $\mu([\sigma]) > 0$ for all $\sigma \in 2^{<\mathbb{N}}$. Intuitively, strictly positive measures embody a certain

[5]For an example, let $\mu = \frac{1}{2}\lambda(\cdot \mid [0]) + \frac{1}{2}\lambda(\cdot \mid [11])$ and $\nu = \frac{1}{2}\lambda(\cdot \mid [1]) + \frac{1}{2}\lambda(\cdot \mid [00])$. Now, for any $\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})$ with $\mu(\mathcal{S}) = 1$, $\lambda(\mathcal{S} \mid [11]) = 1$. Hence, $\lambda(\mathcal{S} \mid [1]) > 0$ and, so, $\nu(\mathcal{S}) > 0$. Similarly, for any $\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})$ with $\nu(\mathcal{S}) = 1$, $\lambda(\mathcal{S} \mid [00]) = 1$. Hence, $\lambda(\mathcal{S} \mid [0]) > 0$ and, so, $\mu(\mathcal{S}) > 0$. Therefore, μ and ν are non-orthogonal. However, neither $\mu \ll \nu$ nor $\nu \ll \mu$, since $\nu([01]) = 0$ while $\mu([01]) = \frac{1}{4}$, and $\mu([10]) = 0$ while $\nu([10]) = \frac{1}{4}$.

[6]Given the definition of the filtration $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$, in our setting this condition is equivalent to the following: for any $\sigma \in 2^{<\mathbb{N}}$, $\mu([\sigma]) > 0$ entails that $\nu([\sigma]) > 0$.

type of open-mindedness: they do not *a priori* rule out any finite string of observations. Clearly, any two strictly positive measures μ and ν are mutually locally absolutely continuous (namely, $\mu \ll_{loc} \nu$ and $\nu \ll_{loc} \mu$)—for a concrete example, take once again the uniform measure λ and the Bernoulli measure $\mu_{\frac{1}{3}}$, which are both strictly positive. While absolute continuity entails local absolute continuity, the reverse implication does not hold: as shown in Example 2.4, λ and $\mu_{\frac{1}{3}}$ are orthogonal and, so, neither of them is absolutely continuous with respect to the other.

Example 2.4 also shows that local absolute continuity does not entail non-orthogonality: two measures can be locally absolutely continuous, and yet there can be an infinite-horizon event (an event that can only be settled by an infinite amount of evidence) on which these two measures maximally disagree. As a matter of fact, non-orthogonality and local absolute continuity are independent notions: neither of them entails the other.[7]

2.2 Merging and polarization in the classical framework

We are now ready to turn our attention to the phenomenon of merging of opinions—and to the other extreme: polarization of opinions.

The most well-studied notion of merging of opinions was introduced in a seminal article by Blackwell and Dubins (1962):

Definition 2.7 (Merging)**.** *Given measures μ and ν, ν is said to merge with μ ($\nu \xrightarrow{\text{M}} \mu$) if, for μ-almost every $\omega \in 2^{\mathbb{N}}$,*

$$\lim_{n\to\infty} \sup_{\mathcal{S}\in\mathcal{B}(2^{\mathbb{N}})} \Big|\nu(\mathcal{S} \mid [\omega \upharpoonright n]) - \mu(\mathcal{S} \mid [\omega \upharpoonright n])\Big| = 0.$$[8]

The distance $\sup_{\mathcal{S}\in\mathcal{B}(2^{\mathbb{N}})} \big|\nu(\mathcal{S}) - \mu(\mathcal{S})\big|$ between μ and ν is called the total variation distance and it essentially amounts to the largest possible difference between the probabilities that μ and ν can assign to the same event in $\mathcal{B}(2^{\mathbb{N}})$. As a result, the quantity $\sup_{\mathcal{S}\in\mathcal{B}(2^{\mathbb{N}})} \big|\nu(\mathcal{S} \mid [\omega \upharpoonright n]) - \mu(\mathcal{S} \mid [\omega \upharpoonright n])\big|$ intuitively represents the maximum possible disagreement between μ and ν after having observed the outcomes of the first n experiments.

A crucial feature of this type of merging (and what makes it such a strong notion of consensus) is that it requires that the agent with prior ν be eventually able to forecast the prob-

[7]To see that non-orthogonality fails to entail local absolute continuity, note that the example given in footnote 5 of two non-orthogonal measures μ and ν such that $\mu \not\ll \nu$ and $\nu \not\ll \mu$ is also a case where $\mu \not\ll_{loc} \nu$ and $\nu \not\ll_{loc} \mu$.

[8]In order for ν to merge with μ in the sense of Definition 2.7, it has to be the case that, for μ-almost every $\omega \in 2^{\mathbb{N}}$, $\mu([\omega \upharpoonright n]) > 0$ and $\nu([\omega \upharpoonright n]) > 0$ for all $n \in \mathbb{N}$. In other words, the support of μ has to be included in the support of ν. As noted earlier, for each ω for which $\mu([\omega \upharpoonright n]) > 0$ and $\nu([\omega \upharpoonright n]) > 0$ for all $n \in \mathbb{N}$, $\mu(\mathcal{S} \mid [\omega \upharpoonright n]) = \mu(\mathcal{S} \mid \mathcal{F}_n)(\omega)$ and $\nu(\mathcal{S} \mid [\omega \upharpoonright n]) = \nu(\mathcal{S} \mid \mathcal{F}_n)(\omega)$ for all $n \in \mathbb{N}$, all $\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})$, and all versions of $\mu(\mathcal{S} \mid \mathcal{F}_n)$ and $\nu(\mathcal{S} \mid \mathcal{F}_n)$.

abilities of every event in agreement with μ, including the probabilities of infinite-horizon events—more precisely, of events in the tail σ-algebra $\mathcal{G}_\infty = \bigcap_{n\in\mathbb{N}} \mathcal{G}_n$, where, for each $n \in \mathbb{N}$, $\mathcal{G}_n = \sigma(\bigcup_{i\geq n} \mathcal{F}_i)$.

The following result, known as the Blackwell-Dubins Theorem, is a central result in the foundations of probability, statistics, and Bayesian epistemology. It establishes that absolute continuity is sufficient for merging: in other words, if ν is no more dogmatic than μ, then, with μ-probability one, ν will eventually agree with μ on the probability of all events, as the evidence accumulates.

Theorem 2.8 (Blackwell and Dubins 1962)**.** *Given measures μ and ν, if $\mu \ll \nu$, then $\nu \xrightarrow{\text{M}} \mu$.*

It is easy to see that merging, as defined above, entails local absolute continuity.[9] But this is not all. As shown by Kalai and Lehrer (1994), the Blackwell-Dubins Theorem admits a converse: in our setting, it is also the case that merging entails absolute continuity. Therefore, merging of opinions in the sense of Blackwell and Dubins and absolute continuity are equivalent notions.

Theorem 2.9 (Kalai and Lehrer 1994)**.** *Given measures μ and ν, if $\nu \xrightarrow{\text{M}} \mu$, then $\mu \ll \nu$.*

The Blackwell-Dubins Theorem is philosophically significant because it has been argued to provide a vindication of subjective Bayesianism by way of demonstrating that divergent initial opinions should not be a cause for concern. Disagreement over priors does not threaten the objectivity of learning and scientific inquiry, the argument goes, because objectivity can be recovered in the form of inter-subjective agreement.

Yet, merging of opinions is not guaranteed to occur in all circumstances: as we have seen, it is attained when the agents' initial beliefs are sufficiently compatible. When the agents' priors are not compatible enough, disagreement may persist, even as the evidence accumulates. For instance, it is not difficult to see that merging entails non-orthogonality, which means that if two measures are orthogonal, then they fail to merge.[10]

The most radical failure of merging is polarization of opinions, which occurs when disagreement, rather than being gradually eliminated by the shared evidence, becomes maximal as the available information increases.

---

[9]Suppose that $\nu \xrightarrow{M} \mu$ but $\mu \not\ll_{loc} \nu$. Then, there is some $\sigma \in 2^{<\mathbb{N}}$ with $\mu([\sigma]) > 0$ but $\nu([\sigma]) = 0$. Let $\mathcal{M} \in \mathcal{B}(2^{\mathbb{N}})$ denote the set of data streams along which ν merges with μ. For all $\omega \in \mathcal{M}$, $\mu([\omega \upharpoonright n]) > 0$ and $\nu([\omega \upharpoonright n]) > 0$ for all $n \in \mathbb{N}$. Since $\mu(\mathcal{M}) = 1$, $\mu(\mathcal{M} \cap [\sigma]) > 0$. Hence, $\mathcal{M} \cap [\sigma]$ is non-empty. Take $\omega \in \mathcal{M} \cap [\sigma]$. Then, $\nu([\omega \upharpoonright n]) = 0$ for all $n \geq |\sigma|$, which is a contradiction.

[10]Suppose that $\mu \perp \nu$. Then, there is some $C \in \mathcal{B}(2^{\mathbb{N}})$ with $\mu(C) = 0$ and $\nu(C) = 1$. As a result, for every $\omega \in 2^{\mathbb{N}}$ with $\mu([\omega \upharpoonright n]) > 0$ and $\nu([\omega \upharpoonright n]) > 0$ for all $n \in \mathbb{N}$, $\mu(C \mid [\omega \upharpoonright n]) = 0$ and $\nu(C \mid [\omega \upharpoonright n]) = 1$ for all $n \in \mathbb{N}$. Hence, ν does not merge with μ.

Definition 2.10 (Polarization)**.** *Given measures μ and ν, ν is said to become polarized with respect to μ ($\nu \parallel \mu$) if, for μ-almost every $\omega \in 2^{\mathbb{N}}$,*

$$\lim_{n\to\infty} \sup_{\mathcal{S}\in\mathcal{B}(2^{\mathbb{N}})} \left|\nu(\mathcal{S} \mid [\omega \upharpoonright n]) - \mu(\mathcal{S} \mid [\omega \upharpoonright n])\right| = 1.$$

Orthogonality and local absolute continuity together entail polarization of opinions: if two agents agree on what evidence is possible but their priors are orthogonal, then, even as they obtain more and more information, their beliefs remain maximally divergent.

Observation 2.11 (Folklore)**.** *Given measures μ and ν, if $\mu \perp \nu$ and $\mu \ll_{loc} \nu$, then $\nu \parallel \mu$.*

Proof. Suppose there is some $C \in \mathcal{B}(2^{\mathbb{N}})$ with $\mu(C) = 0$ and $\nu(C) = 1$, and that $\mu \ll_{loc} \nu$. Let $\mathcal{U}$ be the set $\{\omega \in 2^{\mathbb{N}} : (\forall n)\, \mu([\omega \upharpoonright n]) > 0\}$. Clearly, $\mu(\mathcal{U}) = 1$ ($\mathcal{U}$ is the support of μ). Moreover, for all $\omega \in \mathcal{U}$ and all $n \in \mathbb{N}$, $\nu([\omega \upharpoonright n]) > 0$, since $\mu \ll_{loc} \nu$. Hence, for all $\omega \in \mathcal{U}$ and all $n \in \mathbb{N}$, $\mu(C \mid [\omega \upharpoonright n]) = 0$ and $\nu(C \mid [\omega \upharpoonright n]) = 1$, so that $\left|\nu(C \mid [\omega \upharpoonright n]) - \mu(C \mid [\omega \upharpoonright n])\right| = 1$. Then, for all $\omega \in \mathcal{U}$ and $n \in \mathbb{N}$, $\sup_{\mathcal{S}\in\mathcal{B}(2^{\mathbb{N}})} \left|\nu(\mathcal{S} \mid [\omega \upharpoonright n]) - \mu(\mathcal{S} \mid [\omega \upharpoonright n])\right| = 1$, from which it follows that $\nu \parallel \mu$. □

Since not all priors are guaranteed to merge, understanding under what conditions merging of opinions occurs and under what conditions it fails to occur has been a central goal of much work in Bayesian epistemology, the foundations of statistics, as well as game theory. Determining exactly which types of compatibility lead to merging—and how reasonable these notions of compatibility are—would in fact help elucidate the philosophical implications of the Blackwell-Dubins Theorem and other related merging-of-opinions results. Our results in the remainder of this article contribute a further step in this direction by shedding light on the phenomenon of merging of opinions in the setting of computable Bayesian agents.

3 Merging of opinions for computable Bayesian agents

Classical merging-of-opinions results, such as the Blackwell-Dubins Theorem discussed above, are proven for arbitrary (probability) measures. In what follows, we will restrict attention to computable measures—and, at times, we will also make use of the more general concept of a lower semi-computable measure. This restriction stems from the fact that our aim is to elucidate the phenomenon of merging of opinions in the context of computationally limited Bayesian learners, who may be identified with agents whose initial credences are given by computable priors.

A measure μ on $\mathcal{B}(2^{\mathbb{N}})$ is computable if the function $\sigma \mapsto \mu([\sigma])$ is computable: that is, if $\mu([\sigma])$ is a computable real, uniformly in σ.[11] Analogously, μ is lower semi-computable if

[11] A real number r is computable if there is a computable sequence $q_0, q_1, q_2, \ldots$ of rationals such that $|r - q_n| \leq 2^{-n}$ for all $n \in \mathbb{N}$. In other words, the q_n's approximate r at a computable

$\mu([\sigma])$ is a left-computably enumerable (left-c.e.) real, uniformly in σ.[12] The uniform measure λ on $\mathcal{B}(2^{\mathbb{N}})$ defined earlier is a simple example of a computable measure, as is every other Bernoulli measure with a computable bias. We will indicate it explicitly when the given measures are merely lower semi-computable; otherwise, from now on, all mentioned measures should be assumed to be computable.

While focusing on effective measures of course means losing some generality, it also allows to draw distinctions that were previously beyond reach. Notably, this computability-theoretic perspective allows to investigate more fine-grained notions of compatibility between priors, and to thereby represent the corresponding agents' inductive assumptions in a more detailed way. From a methodological point of view, this is significant because, as mentioned earlier, the Blackwell-Dubins Theorem and the philosophical lesson standardly drawn from it crucially rely on absolute continuity: merging of opinions cannot be gotten for free, it follows when the agents' initial beliefs are sufficiently similar. Absolute continuity, however, is not without detractors (see, for instance, (Earman 1992) and (Miller and Sanchirico 1999)). It is therefore useful to investigate alternative forms of compatibility that lead to merging and the rationale behind them, and, more generally, to gain a deeper understanding of exactly how similar the initial credences of two agents have to be in order for their posterior credences to eventually align in natural ways.

In what follows, we pursue this approach by defining various notions of compatibility induced by algorithmic randomness—a perspective that goes hand in hand with the computability-theoretic restrictions imposed on priors. Then, we show that agreeing on which data streams are algorithmically random indeed leads to merging of opinions between computable Bayesian agents, and that disagreeing on which data streams are algorithmically random leads to polarization of opinions.

3.1 Algorithmic randomness

Algorithmic randomness combines measure-theoretic and computability-theoretic tools to specify what it means for an individual mathematical object (in our case, an individual data stream) to be random relative to a given probability measure. According to a prominent paradigm for defining algorithmic randomness, the measure-theoretic typicality paradigm, randomness is to be equated with typicality: a sequence is random if it is a typical, or representative, outcome from the perspective of the underlying measure. In a nutshell, a sequence is random if it does

rate, uniformly in n. Saying that $\mu([\sigma])$ is a computable real uniformly in σ means that there is a computable function $f : 2^{<\mathbb{N}} \times \mathbb{N} \to \mathbb{Q}$ which, on inputs $\sigma \in 2^{<\mathbb{N}}$ and $n \in \mathbb{N}$, outputs the n-th rational q_n in the approximation that witnesses the computability of $\mu([\sigma])$.

[12]A real number r is left-c.e. if there is a computable non-decreasing sequence of rationals that converges to r in the limit: that is, if r can be approximated from below via a computable sequence of rationals.

not possess any identifying patterns or regularities that would make it stand out, that would render it an atypical and unexpected outcome relative to the underlying measure. The only patterns displayed by a random data stream are the ones that most sequences possess, which is why a random data stream constitutes a representative sample.

To build some intuition, imagine tossing a fair coin an infinite number of times. By the Strong Law of Large Numbers, we know that, with probability one relative to the uniform measure, the limiting relative frequency of 0 in the resulting sequence is $\frac{1}{2}$. Many other results in probability theory have this shape: they establish that almost every sequence satisfies a given property of interest—namely, that the collection of sequences satisfying that property has probability one relative to the underlying measure. For instance, it follows from the Law of the Iterated Logarithm relative to the uniform measure that almost every sequence is such that the relative frequency of 0 is infinitely often above $\frac{1}{2}$ and infinitely often below it. We will refer to properties like these that hold of almost every sequence as 'statistical laws' and say that a data stream satisfies a given statistical law if it belongs to the probability-one set of sequences displaying the corresponding property. For instance, a data stream satisfies the Strong Law of Large Numbers relative to the uniform measure if it belongs to the set of sequences along which the limiting relative frequency of 0 is $\frac{1}{2}$. Now, it stands to reason that a sequence is measure-theoretically typical if it possesses all the properties that almost every sequence possesses and no property that almost no sequence possesses. So, as a first pass, one may be tempted to say that a sequence is random relative to a given measure if it satisfies all statistical laws relative to that measure.

While intuitive, this definition leaves much to be desired. For instance, for the uniform measure (and, in fact, every atomless measure), it leads to a vacuous definition of randomness. Since every sequence $\omega \in 2^{\mathbb{N}}$ belongs to its own singleton set $\{\omega\}$, and every singleton set has measure zero according to the uniform measure, no sequence satisfies all statistical laws: every ω fails to possess the probability-one property corresponding to the set $2^{\mathbb{N}} \setminus \{\omega\}$. As a result, no sequence is random according to this definition.

The key idea behind the theory of algorithmic randomness is that this problem can be resolved by appealing to computability theory: in particular, by restricting attention to the statistical laws that can be effectively specified—roughly, the statistical laws that can be defined in the language of computability theory. More precisely, to be effectively specifiable, a property has to coincide with a subset of Cantor space with a classification in the arithmetical hierarchy. Within the arithmetical hierarchy, a set is assigned classifications of the form Π^0_n, Σ^0_n, or Δ^0_n, with $n \geq 1$. A set $C \subseteq 2^{\mathbb{N}}$ is a Π^0_n class if it is definable by a Π^0_n formula: that is, if there is a computable relation R such that $C = \{\omega \in 2^{\mathbb{N}} : (\forall k_1)(\exists k_2)...(Qk_n) R(\omega \upharpoonright k_1, \omega \upharpoonright k_2, ..., \omega \upharpoonright k_n)\}$, where $Q = \forall$ if n is odd and $Q = \exists$ otherwise. A Σ^0_n class, on the other hand, is the complement of a Π^0_n class. Equivalently, it is a set $C \subseteq 2^{\mathbb{N}}$ definable by a Σ^0_n formula, which means that there exists a computable relation R such that $C = \{\omega \in 2^{\mathbb{N}} : (\exists k_1)(\forall k_2)...(Qk_n) R(\omega \upharpoonright k_1, \omega \upharpoonright k_2, ..., \omega \upharpoonright k_n)\}$, where $Q = \exists$ if n is odd and $Q = \forall$ otherwise. Lastly, a Δ^0_n class is a set $C \subseteq 2^{\mathbb{N}}$ that is both a

Π^0_n class and a Σ^0_n class.[13]

All canonical statistical laws, such as the ones discussed above, are effective in this sense. For instance, the Strong Law of Large Numbers relative to the uniform measure corresponds to the Π^0_3 class

$$\left\{\omega \in 2^{\mathbb{N}} : (\forall k)(\exists m)(\forall n \geq m) \left|\frac{\#0(\omega \upharpoonright n)}{n} - \frac{1}{2}\right| < 2^{-k}\right\},$$

since checking whether $\left|\frac{\#0(\omega\upharpoonright n)}{n} - \frac{1}{2}\right| < 2^{-k}$ can be done computably. In light of this observation, equating randomness with effective measure-theoretic typicality—that is, with the satisfaction of statistical laws that are effectively specifiable—seems to be a natural solution. As mentioned above, this is exactly the path taken by the theory of algorithmic randomness: given a computable measure μ, a data stream ω is algorithmically μ-random if it is an effectively typical sequence of outcomes relative to μ. As we will see, focusing on the satisfaction of effectively specifiable statistical laws ensures that, for each measure μ, the collection of μ-random sequences is itself a μ-measure one set. This means that almost every sequence is μ-random, so that random sequences are indeed typical.

Algorithmic randomness is not the theory of a single randomness notion. Rather, it studies an infinite hierarchy of randomness concepts, each of which corresponds to the satisfaction of a different class of effectively specifiable statistical laws. Below, we provide the definitions of some of the most well-studied and well-behaved algorithmic randomness notions in the literature.[14] We will then see how each of them gives rise to a different type of compatibility between priors.

We begin with one of the most prominent notions of algorithmic randomness: Martin-Löf randomness (Martin-Löf 1966).

Definition 3.1 (Martin-Löf randomness)**.**

(a) A sequence $\{\mathcal{U}_n\}_{n\in\mathbb{N}}$ of Σ^0_1 classes is a sequence of uniformly Σ^0_1 classes if $\bigcap_{n\in\mathbb{N}} \mathcal{U}_n$ is a Π^0_2 class. Let $\{\mathcal{U}_n\}_{n\in\mathbb{N}}$ be a sequence of uniformly Σ^0_1 classes satisfying $\mu(\mathcal{U}_n) \leq 2^{-n}$ for all $n \in \mathbb{N}$. Such a sequence is called a μ-Martin-Löf test. Since, as n goes to infinity, $\mu(\mathcal{U}_n)$ converges to 0 at a computable rate, $\bigcap_{n\in\mathbb{N}} \mathcal{U}_n$ is said to be a set of effective μ-measure zero.

(b) A sequence $\omega \in 2^{\mathbb{N}}$ is μ-Martin-Löf random if and only if there is no μ-Martin-Löf test $\{\mathcal{U}_n\}_{n\in\mathbb{N}}$ such that $\omega \in \bigcap_{n\in\mathbb{N}} \mathcal{U}_n$.

[13]Every Π^0_n class is also a Δ^0_{n+1}, a Π^0_{n+1} and a Σ^0_{n+1} class, and the same holds for Σ^0_n classes. A Σ^0_1 class is the Cantor space analogue of a computably enumerable set of natural numbers, a Π^0_1 class is the analogue of a co-computably enumerable set of natural numbers, and a Δ^0_1 class is the analogue of a computable set of natural numbers.

[14]For an exhaustive treatment of the theory of algorithmic randomness, see, for instance, (Nies 2009) or (Downey and Hirschfeldt 2010).

In short, a sequence is μ-Martin-Löf random if it does not possess any Π^0_2 properties of effective μ-measure zero in the sense of Definition 3.1(a)—equivalently, if it possesses all Σ^0_2 properties of effective μ-measure one. Crucially, this definition entails that, for all computable Bernoulli measures, the Martin-Löf random sequences satisfy the Strong Law of Large Numbers, the Law of the Iterated Logarithm, as well as many other canonical statistical laws.[15]

The collection of μ-Martin-Löf random sequences, which will be denoted by μ-MLR, is itself a set of μ-measure one ($\mu(\mu\text{-MLR}) = 1$). This is because there are only countably many μ-Martin-Löf tests: as a result, the set of all sequences that fail at least one μ-Martin-Löf test is a countable collection of μ-null sets and, so, itself a μ-null set (essentially the same argument establishes that $\mu(\mu\text{-R}) = 1$ for every algorithmic randomness notion R).

Another central algorithmic randomness notion, Schnorr randomness (Schnorr 1971a; Schnorr 1971b), is obtained by considering a more restricted family of randomness tests:

Definition 3.2 (Schnorr randomness)**.**

(a) Let $\{\mathcal{U}_n\}_{n\in\mathbb{N}}$ be a μ-Martin-Löf test such that the measure $\mu(\mathcal{U}_n)$ of each set $\mathcal{U}_n$ is a computable real, uniformly in n. Then, $\{\mathcal{U}_n\}_{n\in\mathbb{N}}$ is called a μ-Schnorr test.

(b) A sequence $\omega \in 2^{\mathbb{N}}$ is μ-Schnorr random if and only if there is no μ-Schnorr test $\{\mathcal{U}_n\}_{n\in\mathbb{N}}$ such that $\omega \in \bigcap_{n\in\mathbb{N}} \mathcal{U}_n$.

Just as in the case of Martin-Löf randomness, a μ-Schnorr random sequence is one that does not possess any Π^0_2 properties of effective μ-measure zero. The type of effectivity involved in the definition of Schnorr randomness is however more stringent than the one involved in the definition of Martin-Löf randomness. In the case of μ-Schnorr tests $\{\mathcal{U}_n\}_{n\in\mathbb{N}}$, we in fact have that not only $\mu(\mathcal{U}_n)$ converges to 0 at a computable rate, but also that each $\mu(\mathcal{U}_n)$ is itself a computable real number.

The collection of μ-Schnorr random sequences will be denoted by μ-SR. For every computable measure μ, $\mu\text{-MLR} \subseteq \mu\text{-SR}$. The converse, however, does not hold in general: for many measures, counting as Schnorr random requires passing 'fewer' tests than in the context of Martin-Löf randomness, so 'more' sequences are random according to Schnorr's definition.

Next, we define the weak n-randomness hierarchy (Kurtz 1981): a family of algorithmic randomness concepts that, as we shall see, yield notions of compatibility that are very closely connected to absolute continuity.

Definition 3.3 (Weak n-randomness)**.** *Let $n \geq 1$. A sequence $\omega \in 2^{\mathbb{N}}$ is μ-weakly n-random if and only if it belongs to every Σ^0_n class of μ-measure one.*[16]

[15] Given appropriate generalizations of these laws, this also holds for arbitrary (computable) measures.

[16] Equivalently, if it avoids all Π^0_n classes of μ-measure zero.

In other words, a sequence is weakly n-random if it satisfies all statistical laws definable by a Σ^0_n formula (including all the Σ^0_n properties of effective measure one). The collection of μ-weakly n-random sequences will be denoted by μ-WnR. For every computable measure μ, ... $\subseteq \mu$-W3R $\subseteq \mu$-W2R $\subseteq \mu$-MLR $\subseteq \mu$-SR $\subseteq \mu$-W1R.

Algorithmic randomness notions can also be defined in terms of unpredictability: that is, one may take the essence of a random sequence to be that past observations do not provide any information that can be exploited to make better-than-chance predictions about future outcomes. This is the central idea behind the unpredictability paradigm, according to which a sequence is algorithmically random if it is impossible for a gambler to devise an effective betting strategy that would allow them to gain unbounded wealth by successively wagering on the bits of that sequence.

The betting strategies employed to define randomness are called dyadic martingales:[17]

Definition 3.4 (Dyadic martingale)**.** *Given a measure μ, a dyadic μ-martingale is a partial function $d :\subseteq 2^{<\mathbb{N}} \to \mathbb{R}^{\geq 0}$ such that, for all strings $\sigma \in 2^{<\mathbb{N}}$,*

(a) if $d(\sigma)$ is undefined, then $\mu([\sigma]) = 0$ (impossibility condition);

(b) $d(\sigma)\mu([\sigma]) = d(\sigma 0)\mu([\sigma 0]) + d(\sigma 1)\mu([\sigma 1])$ (fairness condition), where a term of the form $d(\tau)\mu([\tau])$ is taken to be equal to 0 if $\mu([\tau]) = 0$ even when $d(\tau)$ is undefined.

A dyadic μ-martingale d is said to be normed if $d(\varepsilon) = 1$. It is said to succeed on a sequence $\omega \in 2^{\mathbb{N}}$ if $\limsup_{n\to\infty} d(\omega \upharpoonright n) = \infty$.

A dyadic martingale intuitively formalizes the capital fluctuations incurred by a gambler as a result of following a certain betting strategy. For each $\sigma \in 2^{<\mathbb{N}}$ on which d is defined, $d(\sigma)$ represents the capital accumulated after betting on the first $n = |\sigma|$ bits of a sequence whose initial segment of length n is σ (so $d(\varepsilon)$ represents the initial capital available to the gambler). The impossibility condition and the convention that $d(\tau)\mu([\tau]) = 0$ whenever $\mu([\tau]) = 0$ and $d(\tau)$ is undefined ensure that the fairness condition is well-defined. In turn, the fairness condition ensures that the game is fair: it requires that, at each round of the game, the gambler's expected winnings equal their current capital. A dyadic martingale is successful on a sequence if the underlying betting strategy wins an unbounded amount of wealth when played against that sequence.

A canonical algorithmic randomness notion defined via martingales is computable randomness:[18]

[17]Here, in generalizing the concept of a dyadic martingale from the uniform measure to arbitrary computable measures, we follow Rute (2016).

[18]Computable randomness was introduced by Schnorr (1971a) and Schnorr (1971b) in the context of the uniform measure. Its generalization to arbitrary computable measures is due to Rute (2016). Computable randomness can also be characterized within the measure-theoretic

Definition 3.5 (Computable randomness)**.** *A sequence $\omega \in 2^{\mathbb{N}}$ is μ-computably random if and only if (a) $\mu([\omega \upharpoonright n]) > 0$ for all $n \in \mathbb{N}$, and (b) $\limsup_{n\to\infty} d(\omega \upharpoonright n) < \infty$ for all almost everywhere computable dyadic μ-martingales d.*[19]

The collection of μ-computably random sequences will be denoted by μ-CR. For every computable measure μ, μ-MRL $\subseteq \mu$-CR $\subseteq \mu$-SR (but the converse inclusions do not hold in general).

The definition of a dyadic martingale can be relaxed as follows: given a measure μ, a function $d :\subseteq 2^{<\mathbb{N}} \to \mathbb{R}^{\geq 0}$ is a dyadic μ-supermartingale if, for all $\sigma \in 2^{<\mathbb{N}}$, d satisfies the impossibility condition from Definition 3.4, and $d(\sigma)\mu([\sigma]) \geq d(\sigma 0)\mu([\sigma 0]) + d(\sigma 1)\mu([\sigma 1])$. Dyadic supermaringales differ from dyadic martingales in that they can be wasteful: they are 'allowed to discard part of [the] capital, such as by buying drinks or tipping the dealer' (Downey and Hirschfeldt 2010, p. 235). Dyadic martingales and supermartingales can be put to use to provide an alternative characterization of Martin-Löf randomness within the unpredictability paradigm:[20]

Theorem 3.6 (Schnorr 1971b)**.** *Let $\omega \in 2^{\mathbb{N}}$. The following are equivalent:*

(1) ω is μ-Martin-Löf random;

(2) $\mu([\omega \upharpoonright n]) > 0$ for all $n \in \mathbb{N}$, and $\limsup_{n\to\infty} d(\omega \upharpoonright n) < \infty$ for all left-c.e. dyadic μ-martingales d.

In (2) above, the requirement that d be left-c.e. can be relaxed to the requirement that d be almost everywhere left-c.e., and the requirement that d be a dyadic μ-martingale can be relaxed to the requirement that d be a dyadic μ-supermartingale.[21]

We conclude our brief overview of the theory of algorithmic randomness with density randomness, a notion that results from a natural modification of the characterization of Martin-Löf randomness given in Theorem 3.6:

Definition 3.7 (Density randomness)**.** *A sequence $\omega \in 2^{\mathbb{N}}$ is μ-density random if and only if (a) $\mu([\omega \upharpoonright n]) > 0$ for all $n \in \mathbb{N}$, and (b) $\lim_{n\to\infty} d(\omega \upharpoonright n)$ exists and is finite for all left-c.e.*

---

typicality paradigm in terms of the satisfaction of effective statistical laws (see, for instance, (Downey and Hirschfeldt 2010)).

[19] A dyadic μ-martingale is computable if it is a total computable function and almost everywhere computable if it is a partial computable function. For strictly positive measures (measures that assign positive probability to every cylinder), computable and almost everywhere computable dyadic martingales coincide.

[20] Schnorr (1971b) proved Theorem 3.6 in the context of the uniform measure, but the result is generalizable to all computable measures.

[21] A dyadic μ-martingale or μ-supermartingale is left-c.e. if it is a total left-c.e. function and almost everywhere left-c.e. if it is a partial left-c.e. function.

dyadic μ-martingales d. Once again, we obtain the same randomness notion if the requirement that d be left-c.e. is relaxed to the requirement that d be almost everywhere left-c.e., and if the requirement that d be a dyadic μ-martingale is relaxed to the requirement that d be a dyadic μ-supermartingale.

The collection of μ-density random sequences will be denoted by μ-DR. For every computable measure μ, μ-$\mathsf{DR} \subseteq \mu$-$\mathsf{MLR}$, but the reverse inclusion does not hold in general.[22]

3.2 Algorithmic randomness and merging

Though possibly surprising at first, algorithmic randomness offers a natural framework for defining notions of compatibility between priors. As already suggested by Skyrms, algorithmic randomness notions may in fact be thought of as embodying beliefs in a special version of the principle of the uniformity of nature:

> Without pursuing the matter in detail, I want to note a fact that is invariant over questions of fine tuning the analysis. It is that random sequences must have a limiting relative frequency. This is a rather spicy revelation in view of Reichenbach's taking the existence of limiting relative frequencies as the principle of the uniformity of nature. The most chaotic and disordered alternative to uniformity that we can find *entails* uniformity-in-the-sense-of-Reichenbach! [...] Randomness is indeed a kind of disorder, but it carries with it of necessity a kind of statistical order in the large. (Skyrms 1984, p. 38)

In particular, algorithmic randomness notions may be taken to encode a specific type of inductive assumptions—or commitments (either explicit or implicit)—that result from the subjective prior with respect to which randomness is defined. This is because algorithmic randomness notions embody the effective statistical laws that an agent expects to see in the data by virtue of having a certain prior. For instance, if their initial beliefs are captured by the Bernoulli measure with bias $\frac{2}{3}$ towards 1, the agent is (at least implicitly) making the inductive assumption that the limiting relative frequency of 0 in the true data stream is $\frac{1}{3}$, in the sense that they assign probability one to the set of data streams having this property. So, by believing that every sequence of n observations (or outcomes of the experiment under consideration) featuring k 0's has probability $\frac{1}{3}^k \frac{2}{3}^{(n-k)}$, the agent is also committed to believing in the relevant version of the Strong Law of Large Numbers. But, as we have seen, algorithmic randomness captures inductive assumptions of exactly this type: beliefs in the fact that the data will display certain effective statistical regularities that stem from one's beliefs about events that can be settled with a finite number of observations.

[22]See, for instance, (Miyabe et al. 2016). The fact that $\mu(\mu\text{-}\mathsf{DR}) = 1$ follows from Doob's Martingale Convergence Theorem (see, for instance, (Williams 1991, Theorem 11.5)) and the fact that dyadic μ-martingales are non-negative.

With this motivation in place, let us turn to the properties of the notions of compatibility induced by algorithmic randomness. First, recall that, given two computable priors μ and ν, as well as an algorithmic randomness notion R, ν is compatible with μ with respect to R if μ-$\mathsf{R} \subseteq \nu$-R. Intuitively, this indicates that the agent with prior ν cannot be surprised by a data stream that the agent with prior μ considers typical (in the sense of R). Now, a well-known result due to Muchnik et al. (1998) establishes that, for any two computable measures, agreeing on which data streams are computably random entails agreeing on which data streams are Martin-Löf random:

Proposition 3.8 (Muchnik et al. 1998)**.** *If* μ-$\mathsf{CR} \subseteq \nu$-$\mathsf{CR}$*, then* μ-$\mathsf{MLR} \subseteq \nu$-$\mathsf{MLR}$.

This means that having compatible beliefs about the type of uniformity embodied by computable randomness entails having compatible beliefs about the type of uniformity embodied by Martin-Löf randomness.

As we will now show, agreeing on which data streams are Martin-Löf random in turn entails agreeing on which data streams are density random (again, for all computable measures). Hence, having compatible beliefs about the type of uniformity embodied by Martin-Löf randomness in turn entails having compatible beliefs about the type of uniformity embodied by density randomness.

Proposition 3.9. *If* μ-$\mathsf{MLR} \subseteq \nu$-$\mathsf{MLR}$*, then* μ-$\mathsf{DR} \subseteq \nu$-$\mathsf{DR}$.

Proof. Suppose that μ-$\mathsf{MLR} \subseteq \nu$-$\mathsf{MLR}$ and that $\omega \in \mu$-DR. Then, $\omega \in \mu$-MLR. Therefore, $\omega \in \mu$-$\mathsf{DR} \cap \nu$-MLR. Suppose towards a contradiction that $\omega \notin \nu$-DR. Since $\omega \in \nu$-MLR, for all n, $\nu([\omega \restriction n]) > 0$. Thus, by Theorem A.2 (see the Appendix), there must be a lower semi-computable semi-measure ξ such that either $\lim_{n\to\infty} \dfrac{\xi([\omega \restriction n])}{\nu([\omega \restriction n])} = \infty$ or the sequence $\left\{\frac{\xi([\omega\restriction n])}{\nu([\omega\restriction n])}\right\}_{n\in\mathbb{N}}$ does not have a limit.

Let us consider the second case first. The fact that $\left\{\frac{\xi([\omega\restriction n])}{\nu([\omega\restriction n])}\right\}_{n\in\mathbb{N}}$ does not have a limit entails that $\xi([\omega \restriction n]) > 0$ for all n, and that there are reals a, b with $0 < a < b$ such that the number of upcrossings of the sequence $\left\{\frac{\xi([\omega\restriction n])}{\nu([\omega\restriction n])}\right\}_{n\in\mathbb{N}}$ across the interval $[a, b]$ is infinite. Since $\omega \in \mu$-DR, $\mu([\omega \restriction n]) > 0$ for all n. Moreover, Theorem A.2 entails that $\lim_{n\to\infty} \dfrac{\xi([\omega \restriction n])}{\mu([\omega \restriction n])}$ exists and is finite. Call this limit ℓ. We have two sub-cases to examine. First, suppose that $\ell = 0$. Then, $\lim_{n\to\infty} 1 \Big/ \dfrac{\xi([\omega \restriction n])}{\mu([\omega \restriction n])} = \infty$. For each n,

$$\frac{\mu([\omega \restriction n])}{\nu([\omega \restriction n])} = \frac{\xi([\omega \restriction n])}{\nu([\omega \restriction n])} \Big/ \frac{\xi([\omega \restriction n])}{\mu([\omega \restriction n])}$$

(all three ratios are well-defined and positive because $\mu([\omega \restriction n]) > 0$, $\nu([\omega \restriction n]) > 0$, and $\xi([\omega \restriction n]) > 0$ for all n). Since there are infinitely many n with $\frac{\xi([\omega\restriction n])}{\nu([\omega\restriction n])} \geq b > 0$, it follows that

$$\limsup_{n\to\infty} \frac{\mu([\omega \restriction n])}{\nu([\omega \restriction n])} = \limsup_{n\to\infty} \frac{\xi([\omega \restriction n])}{\nu([\omega \restriction n])} \Big/ \frac{\xi([\omega \restriction n])}{\mu([\omega \restriction n])} \geq \lim_{n\to\infty} \left(b \Big/ \frac{\xi([\omega \restriction n])}{\mu([\omega \restriction n])} \right) = \infty.$$

This, however, contradicts the fact that $\omega \in \nu$-MLR (Theorem A.1 in the Appendix). So, suppose instead that $\ell > 0$. Let $a' = \frac{2}{3}a + \frac{1}{3}b$ and $b' = \frac{1}{3}a + \frac{2}{3}b$. Then, $0 < a' < b'$, and the number of upcrossings of the sequence $\left\{\frac{\xi([\omega\restriction n])}{\nu([\omega\restriction n])}\right\}_{n\in\mathbb{N}}$ across the interval $[a', b']$ is infinite, as well. Moreover, for each of the infinitely many n such that $\frac{\xi([\omega\restriction n])}{\nu([\omega\restriction n])} \le a < a'$, $a' - \frac{\xi([\omega\restriction n])}{\nu([\omega\restriction n])} > \frac{1}{3}(b-a)$, and for each of the infinitely many n such that $\frac{\xi([\omega\restriction n])}{\nu([\omega\restriction n])} \ge b > b'$, $\frac{\xi([\omega\restriction n])}{\nu([\omega\restriction n])} - b' > \frac{1}{3}(b-a)$. We then have that

$$\limsup_{n\to\infty} \frac{\nu([\omega\restriction n])}{\mu([\omega\restriction n])} = \limsup_{n\to\infty} \frac{\xi([\omega\restriction n])}{\mu([\omega\restriction n])} \Bigg/ \frac{\xi([\omega\restriction n])}{\nu([\omega\restriction n])} \ge \frac{\ell}{a'}, \text{ and}$$

$$\liminf_{n\to\infty} \frac{\nu([\omega\restriction n])}{\mu([\omega\restriction n])} = \liminf_{n\to\infty} \frac{\xi([\omega\restriction n])}{\mu([\omega\restriction n])} \Bigg/ \frac{\xi([\omega\restriction n])}{\nu([\omega\restriction n])} \le \frac{\ell}{b'}.$$

Since, ℓ, a' and b' are all positive and $a' < b'$, $\frac{\ell}{a'} > \frac{\ell}{b'}$. Hence, the sequence $\left\{\frac{\nu([\omega\restriction n])}{\mu([\omega\restriction n])}\right\}_{n\in\mathbb{N}}$ fails to converge, which, by Theorem A.2, contradicts the assumption that $\omega \in \mu$-DR.

Let us now consider the first case: that is, suppose that $\lim\limits_{n\to\infty} \frac{\xi([\omega\restriction n])}{\nu([\omega\restriction n])} = \infty$. This entails that $\xi([\omega\restriction n]) > 0$ for all n. Since $\lim\limits_{n\to\infty} \frac{\xi([\omega\restriction n])}{\mu([\omega\restriction n])} < \infty$ by Theorem A.2, it follows that

$$\limsup_{n\to\infty} \frac{\mu([\omega\restriction n])}{\nu([\omega\restriction n])} = \limsup_{n\to\infty} \frac{\xi([\omega\restriction n])}{\nu([\omega\restriction n])} \Bigg/ \frac{\xi([\omega\restriction n])}{\mu([\omega\restriction n])} = \infty,$$

which contradicts the fact that $\omega \in \nu$-MLR (Theorem A.1). Hence, $\omega \in \nu$-DR. □

Next, we will see that agreeing on which data streams are density random entails absolute continuity: if ν is compatible with μ with respect to density randomness, then ν is no more dogmatic than μ. *A fortiori*, by Proposition 3.9 and Proposition 3.8, agreeing on which data streams are Martin-Löf random and agreeing on which data streams are computably random entail absolute continuity, too.[23]

Proposition 3.10. *If* μ-DR $\subseteq \nu$-DR, *then* $\mu \ll \nu$.

Proof. Suppose there is some $\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})$ with $\nu(\mathcal{S}) = 0$ and $\mu(\mathcal{S}) > 0$. Then, there is some rational q with $\mu(\mathcal{S}) > q > 0$. Since ν is regular,[24] $\nu(\mathcal{S}) = \inf\{\nu(\mathcal{U}) : \mathcal{S} \subseteq \mathcal{U}$ and $\mathcal{U} \in \mathcal{B}(2^{\mathbb{N}})$ is an open set$\}$. Hence, for all $n \in \mathbb{N}$, there is an open set $\mathcal{U}_n$ with $\mathcal{S} \subseteq \mathcal{U}_n$ such that $\nu(\mathcal{U}_n) < 2^{-n}$ and $\mu(\mathcal{U}_n) > q$. Every $\mathcal{U}_n$ is of the form $\bigcup_{i\in\mathbb{N}}[\sigma_{n,i}]$—where, without loss of generality, the cylinders $[\sigma_{n,i}]$ can be taken to be pairwise disjoint. For each $\mathcal{U}_n$, there is some K_n such that (1) $\mu(\bigcup_{i\le K_n}[\sigma_{n,i}]) > q$, while (2) $\nu(\bigcup_{i\le K_n}[\sigma_{n,i}]) < 2^{-n}$. Let $V_n = \{\sigma_{n,0}, ..., \sigma_{n,K_n}\}$ and let $\bigcup_{i\le K_n}[\sigma_{n,i}]$ be denoted as $[V_n]$. For each $m \in \mathbb{N}$, let $\mathcal{V}_m = \bigcup_{n>m}[V_n]$. Then, we have that $\nu(\mathcal{V}_m) \le \sum_{n>m} \nu([V_n]) \le \sum_{n>m} 2^{-n} = 2^{-m}$ and, since $\mu([V_n]) > q$ for all n, $\mu(\mathcal{V}_m) > q$, as

[23] The fact that μ-MLR $\subseteq \nu$-MLR entails that $\mu \ll \nu$ was proven by Bienvenu and Merkle (2009). The proof of Proposition 3.10 is analogous to the proof of this fact.

[24] Regularity follows from the fact that ν is a Borel probability measure and Cantor space is a locally compact Hausdorff space with a countable base.

well. Note that sets V_n with the above properties can be chosen in such a way that $\{\mathcal{V}_m\}_{m\in\mathbb{N}}$ is a sequence of uniformly Σ^0_1 classes. Given n, simply enumerate the strings in $2^{<\mathbb{N}}$ (for instance, in the length-lexicographic order) until conditions (1) and (2) are met. By the above, we are guaranteed that a finite, prefix-free collection of cylinders satisfying these conditions will eventually be found, effectively. Hence, $\{\mathcal{V}_m\}_{m\in\mathbb{N}}$ is a ν-Martin-Löf test. This entails that $\bigcap_{m\in\mathbb{N}} \mathcal{V}_m \cap \nu\text{-}\mathsf{MLR} = \emptyset$. Given that $\nu\text{-}\mathsf{DR} \subseteq \nu\text{-}\mathsf{MLR}$, $\bigcap_{m\in\mathbb{N}} \mathcal{V}_m \cap \nu\text{-}\mathsf{DR} = \emptyset$. However, since $\mu(\mathcal{V}_m) > q$ for all m and the sequence $\{\mathcal{V}_m\}_{m\in\mathbb{N}}$ is nested, $\mu(\bigcap_{m\in\mathbb{N}} \mathcal{V}_m) \geq q > 0$. Due to the fact that μ-DR has μ-measure one, we therefore have that $\bigcap_{m\in\mathbb{N}} \mathcal{V}_m \cap \mu\text{-}\mathsf{DR} \neq \emptyset$. It then follows that $\mu\text{-}\mathsf{DR} \not\subseteq \nu\text{-}\mathsf{DR}$. □

So, agreeing on which data streams are computably random, Martin-Löf random, or density random all entail absolute continuity. This is epistemologically significant because, by the Blackwell-Dubins Theorem, we then have that these three forms of compatibility induced by algorithmic randomness all ensure asymptotic merging of opinions. In other words, the inductive assumptions encoded by these core algorithmic randomness notions—the commitments to the global uniformity of nature that they each represent—when shared by computable Bayesian agents, guarantee the attainment of inter-subjective agreement.

Corollary 3.11. *If* $\mu\text{-}\mathsf{DR} \subseteq \nu\text{-}\mathsf{DR}$, *then* $\nu \xrightarrow{\mathrm{M}} \mu$. A fortiori, *if* $\mu\text{-}\mathsf{MLR} \subseteq \nu\text{-}\mathsf{MLR}$, *then* $\nu \xrightarrow{\mathrm{M}} \mu$, *and if* $\mu\text{-}\mathsf{CR} \subseteq \nu\text{-}\mathsf{CR}$, *then* $\nu \xrightarrow{\mathrm{M}} \mu$.

While it does not entail the other notions of compatibility induced by randomness, agreeing on which data streams are Schnorr random entails absolute continuity, too:

Proposition 3.12 (Bienvenu and Merkle 2009)**.** *If* $\mu\text{-}\mathsf{SR} \subseteq \nu\text{-}\mathsf{SR}$, *then* $\mu \ll \nu$.

Thus, the inductive assumptions encapsulated by Schnorr randomness lead to almost-sure inter-subjective agreement, as well.

Corollary 3.13. *If* $\mu\text{-}\mathsf{SR} \subseteq \nu\text{-}\mathsf{SR}$, *then* $\nu \xrightarrow{\mathrm{M}} \mu$.

Note that the implications between the notions of compatibility induced by algorithmic randomness presented above cannot be reversed: that is, for each of them, it is possible to find two computable measures for which the converse implication fails.[25] This means that these notions of compatibility are all different from each other, as well as from absolute continuity. Hence, approaching the question of when inter-subjective agreement is attainable from the perspective of algorithmic randomness indeed affords a richer, more fine-grained analysis of the type of commitments and inductive assumptions that a (computable) Bayesian agent can have. The compatibility notions generated by algorithmic randomness give rise to a novel hierarchy of notions of agreement between priors that refine the type of agreement captured by absolute continuity.

[25] See (Bienvenu and Merkle 2009) and (Zaffora Blando 2020).

Next, we turn our attention to the weak n-randomness hierarchy. In the effective setting, it is natural to consider the following version of absolute continuity, which only applies to Π^0_n classes:

Definition 3.14 (Π^0_n-absolute continuity)**.** *Given measures μ and ν, μ is Π^0_n-absolutely continuous with respect to ν ($\mu \ll_{\Pi^0_n} \nu$) if, for every Π^0_n class $\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})$, $\mu(\mathcal{S}) > 0$ entails that $\nu(\mathcal{S}) > 0$.*[26]

If μ is Π^0_n-absolutely continuous with respect to ν, then the agent with prior ν cannot be surprised by any event definable by a Π^0_n formula to which the agent with prior μ assigns positive probability. This, however, leaves open the possibility that there might be more complex events that have μ-positive probability and, yet, ν-probability zero.

For each n, it is easy to see that the notion of compatibility yielded by weak n-randomness coincides with Π^0_n-absolute continuity:

Observation 3.15. *For all $n \geq 1$, $\mu \ll_{\Pi^0_n} \nu$ if and only if μ-WnR $\subseteq \nu$-WnR.*

Proof. For the left-to-right direction, suppose that $\omega \in \mu$-WnR and let C be a Σ^0_n class of ν-measure one. Then, $\overline{C}$ is a Π^0_n class of ν-measure zero. Since $\mu \ll_{\Pi^0_n} \nu$, it follows that $\mu(\overline{C}) = 0$ and $\mu(C) = 1$. Then, given that $\omega \in \mu$-WnR, $\omega \in C$. But since C was an arbitrary Σ^0_n class of ν-measure one, we can conclude that $\omega \in \nu$-WnR. So, μ-WnR $\subseteq \nu$-WnR.
For the right-to-left direction, suppose that there is a Π^0_n class $\mathcal{A}$ such that $\nu(\mathcal{A}) = 0$, but $\mu(\mathcal{A}) > 0$. Then, $\overline{\mathcal{A}}$ is a Σ^0_n class of ν-measure one, which entails that ν-WnR $\subseteq \overline{\mathcal{A}}$. However, $\mu(\overline{\mathcal{A}}) < 1$, so μ-WnR $\nsubseteq \overline{\mathcal{A}}$, because the collection of μ-weakly n-random sequences has μ-measure one. Hence, μ-WnR $\nsubseteq \nu$-WnR. □

The type of compatibility induced by weak n-randomness is thus but a variant of absolute continuity—arguably, the most canonical notion of agreement between priors.

As a matter of fact, the connections between weak n-randomness and absolute continuity are deeper than the above observation lets out. Recall that a Π^0_2 class is the effective analogue of a G_δ subset of Cantor space—where a G_δ set is a countable intersection of open sets. The proposition below is the effective version (in the context of computable measures) of the well-known equivalence of absolute continuity and absolute continuity restricted to G_δ sets.[27]

[26] One could also consider the notion of Σ^0_n-absolute continuity: namely, absolute continuity restricted to Σ^0_n classes. Without loss of generality, we can however focus on Π^0_n-absolute continuity, since Π^0_n-absolute continuity is equivalent to Σ^0_{n+1}-absolute continuity.

[27] The proof of the non-trivial direction of Proposition 3.16 is the same as the first part of the proof of Proposition 3.10. Again, Proposition 3.16 is but an effective analogue of a canonical alternative characterization of absolute continuity in terms of a simple continuity condition: $\mu \ll \nu$ if and only if, for all $\epsilon > 0$, there is $\delta > 0$ such that, for all $\mathcal{S} \in \mathcal{B}(2^{\mathbb{N}})$, $\mu(\mathcal{S}) < \delta$ entails that $\nu(\mathcal{S}) < \epsilon$ (see, for instance, (Royden and Fitzpatrick 2010, Proposition 19, p. 381)).

Proposition 3.16 (Folklore)**.** *Given measures μ and ν, $\mu \ll \nu$ if and only if $\mu \ll_{\Pi^0_2} \nu$.*

Proposition 3.16 establishes that, past the second level, the Π^0_n-absolute continuity hierarchy collapses: that is, for all $n \geq 2$, Π^0_n-absolute continuity is not a weaker form of absolute continuity, it actually coincides with it. So, for $n \geq 2$, ν is no more dogmatic than μ when it comes to events definable by Π^0_n formulas if and only if ν is no more dogmatic than μ *simpliciter*. By Observation 3.15 and Proposition 3.16, we then have that, for every $n \geq 2$, having compatible beliefs about which data streams are weakly n-random is the same as absolute continuity.

Corollary 3.17. *For all $n \geq 2$, $\mu \ll \nu$ if and only if μ-$\mathsf{WnR} \subseteq \nu$-$\mathsf{WnR}$.*

An immediate consequence of this equivalence is that, for all $n \geq 2$, the type of compatibility yielded by weak n-randomness guarantees merging of opinions and, vice-versa, merging of opinions guarantees agreement on the weakly n-random data streams.

Corollary 3.18. *For all $n \geq 2$, μ-$\mathsf{WnR} \subseteq \nu$-$\mathsf{WnR}$ if and only if $\nu \xrightarrow{\text{M}} \mu$.*

In light of the above, it is natural to ask whether Π^0_1-absolute continuity coincides with absolute continuity, as well. This question, originally raised by Gaifman and Snir (1982), was given a negative answer by Bienvenu and Merkle (2009):

Proposition 3.19 (Bienvenu and Merkle 2009)**.** *There exist measures μ and ν such that μ-$\mathsf{W1R}$ $\subseteq \nu$-$\mathsf{W1R}$, but $\mu \not\ll \nu$.*

This establishes that the notion of compatibility induced by weak 1-randomness is the only one, among the compatibility concepts generated by the weak n-randomness hierarchy, that is strictly weaker than absolute continuity.

In spite of not entailing absolute continuity, agreement on weak 1-randomness does entail local absolute continuity—this simply follows from the fact that cylinders, being Δ^0_1 classes, are also Π^0_1 classes. The entailment is strict, as shown by the following example:

Example 3.20. *Take the uniform measure λ and the Bernoulli measure $\mu_{\frac{1}{3}}$. Then, $\lambda \ll_{loc} \mu_{\frac{1}{3}}$, since both measures are strictly positive. Let $C = \{\omega \in 2^{\mathbb{N}} : (\exists m)(\forall n \geq m) \frac{\#0(\omega\restriction n)}{n} > \frac{7}{20}\}$. Clearly, $\mu_{\frac{1}{3}}(C) = 0$, while $\lambda(C) = 1$. Since C is a Σ^0_2 class, we then have that $\lambda \not\ll_{\Sigma^0_2} \mu_{\frac{1}{3}}$ and, consequently, that λ-$\mathsf{W1R} \not\subseteq \mu_{\frac{1}{3}}$-$\mathsf{W1R}$ (by Observation 3.15 and the fact that Σ^0_2-absolute continuity is equivalent to Π^0_1-absolute continuity).*[28]

[28]A more general reason for why local absolute continuity does not entail agreement on weak 1-randomness is that the latter entails non-orthogonality (see Proposition 40 and Corollary 41 in (Bienvenu and Merkle 2009)), while, as seen earlier, local absolute continuity does not. Bienvenu and Merkle (2009) also show that non-orthogonality does not entail agreement on weak 1-randomness (see Proposition 54).

Crucially, the fact that agreement on weak 1-randomness is strictly weaker than absolute continuity allows to conclude that the type of compatibility induced by weak 1-randomness is too weak to entail merging of opinions in the sense of Blackwell and Dubins:

Corollary 3.21. *There exist measures μ and ν such that μ-$\mathsf{W1R} \subseteq \nu$-$\mathsf{W1R}$, and yet $\nu \overset{\mathrm{M}}{\not\longrightarrow} \mu$.*

Therefore, while agreement on algorithmic randomness generally entails merging, weak 1-randomness is the exception: making compatible inductive assumptions about the global regularities encoded by weak 1-randomness does not suffice to attain inter-subjective agreement in the strong sense of Blackwell and Dubins. This observation is in itself interesting because weak 1-randomness is a bit of an outlier within the algorithmic randomness hierarchy. In particular, weak 1-randomness does not entail several fundamental statistical laws, such as the Strong Law of Large Numbers and the Law of the Iterated Logarithm discussed earlier. Hence, it is perhaps not so surprising that agreeing on which data streams are weakly 1-random does not ensure merging. In fact, this failure, together with Corollary 3.18, can be taken to corroborate our explanation for why it is reasonable to use algorithmic randomness to define notions of compatibility, in that it shows that having compatible inductive assumptions about the global uniformity of nature—about sufficiently many statistical laws—is not only sufficient but also necessary for merging.

Since the type of compatibility induced by weak 1-randomness does not entail merging in the sense of Blackwell and Dubins, an immediate question is whether there is some weaker type of merging that agreement on weak 1-randomness might nonetheless guarantee. After all, the notion of merging of opinions introduced by Blackwell and Dubins, which requires eventual agreement on all events, including tail events, may be deemed excessively demanding, especially in the context of computationally limited agents. From this perspective, a natural, less demanding alternative is the following notion of merging, first investigated by Kalai and Lehrer (1993) and Kalai and Lehrer (1994), which requires alignment of opinions only on finite-horizon events:

Definition 3.22 (Weak merging)**.** *Given measures μ and ν, ν is said to weakly merge with μ ($\nu \overset{\mathrm{WM}}{\longrightarrow} \mu$) if, for μ-almost every $\omega \in 2^{\mathbb{N}}$ and all $k \in \mathbb{N}$,*

$$\lim_{n\to\infty} \sup_{\mathcal{S}\in\mathcal{F}_{n+k}} \left|\nu(\mathcal{S} \mid [\omega \restriction n]) - \mu(\mathcal{S} \mid [\omega \restriction n])\right| = 0.$$

The type of compatibility induced by weak 1-randomness targets statistical laws that correspond to Σ^0_1 classes: namely, statistical laws whose satisfaction can be verified with a finite number of observations. This is because a Σ^0_1 class is an (effectively) open set: that is, a countable union of cylinders—and membership in a cylinder can be decided after a finite number of observations. The fact that weak 1-randomness does not entail some crucial statistical regularities prevents it from yielding a type of compatibility strong enough to ensure merging in the sense of Blackwell and Dubins; yet, the inductive assumptions captured by Σ^0_1 classes of measure one might nonetheless suffice for weak merging, which only requires the attainment of inter-subjective agreement on short-run events. We leave it as an open question whether this is indeed the case.

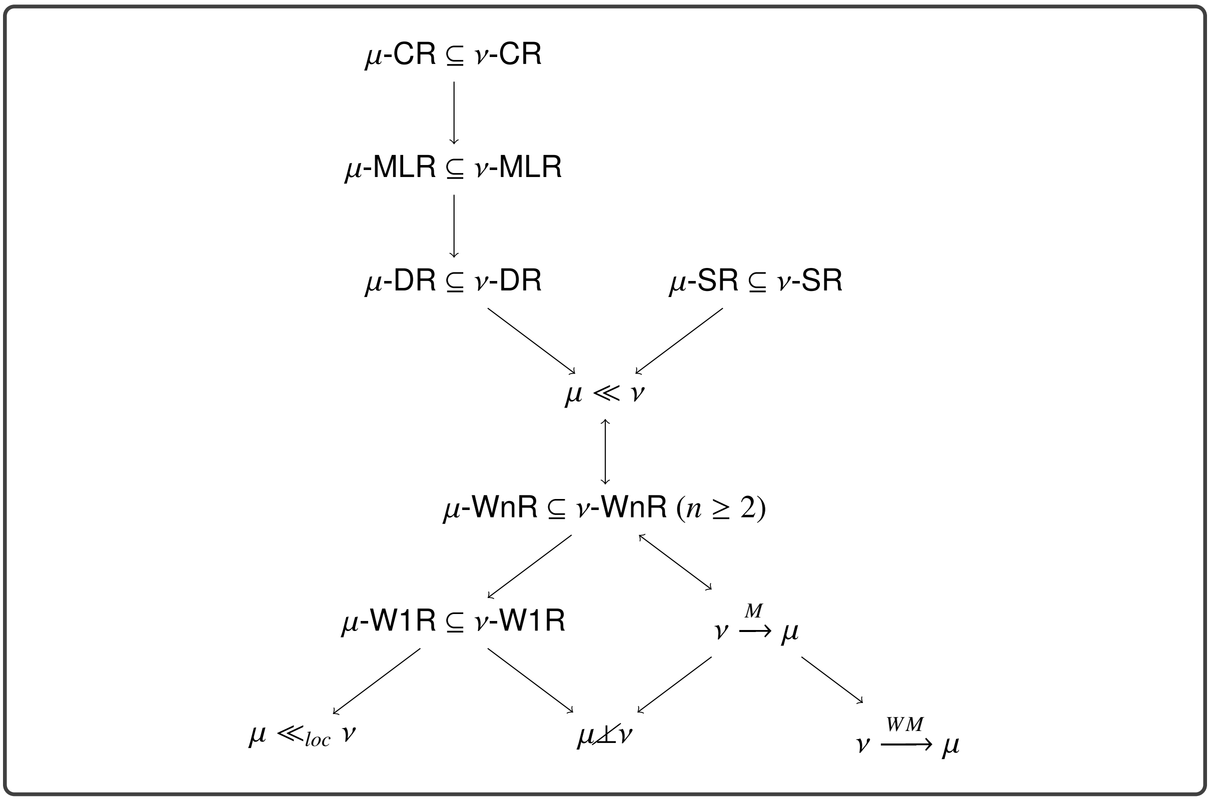

Figure 1: Logical dependencies between the notions of compatibility induced by algorithmic randomness, some classical notions of compatibility, merging in the sense of Blackwell and Dubins (1962), and weak merging in the sense of Kalai and Lehrer (1993).

3.3 Algorithmic randomness and polarization

Algorithmic randomness can be used to define not only notions of compatibility between computable priors, but also notions of incompatibility or disagreement. Just as orthogonality corresponds to the most radical failure of absolute continuity, for any algorithmic randomness notion R, two computable measures μ and ν are radically incompatible with respect to R when μ-R $\cap$ ν-R $= \emptyset$.

Surprisingly, as we have seen, the logical dependencies between the compatibility notions yielded by algorithmic randomness are very different from the logical relations that hold among the underlying randomness concepts. We conclude by noting that the logical dependencies between the types of incompatibility induced by algorithmic randomness are instead a mirror image of the algorithmic randomness hierarchy.

Observation 3.23. *Given measures μ and ν, the following hold:*

(i) if μ-W1R $\cap$ *ν*-W1R $= \emptyset$*, then μ*-SR $\cap$ *ν*-SR $= \emptyset$*;*

(ii) if μ-SR $\cap$ ν-SR $= \emptyset$*, then* μ-CR $\cap$ ν-CR $= \emptyset$*;*

(iii) if μ-CR $\cap$ ν-CR $= \emptyset$*, then* μ-MLR $\cap$ ν-MLR $= \emptyset$*;*

(iv) if μ-MLR $\cap$ ν-MLR $= \emptyset$*, then* μ-DR $\cap$ ν-DR $= \emptyset$*;*

(v) if μ-MLR $\cap$ ν-MLR $= \emptyset$*, then* μ-W2R $\cap$ ν-W2R $= \emptyset$*;*

(vi) if μ-WnR $\cap$ ν-WnR $= \emptyset$*, then* μ-Wn+1R $\cap$ ν-Wn+1R $= \emptyset$ *for all* $n \geq 1$.

Proof. All of the above cases are proved in the same way: they rely on the logical dependencies between the algorithmic randomness notions involved. As an example, consider case (i). Suppose that μ-W1R $\cap$ ν-W1R $= \emptyset$. Since μ-SR $\subseteq$ μ-W1R and ν-SR $\subseteq$ ν-W1R, we can immediately conclude that μ-SR $\cap$ ν-SR $= \emptyset$. □

It is also immediate to see that, for any algorithmic randomness notion R and computable measures μ, ν, if μ and ν are radically incompatible with respect to R (if μ-R $\cap$ ν-R $= \emptyset$), then μ and ν are orthogonal. This follows from the fact that $\mu(\mu\text{-R}) = \nu(\nu\text{-R}) = 1$, which, together with the fact that μ-R $\cap$ ν-R $= \emptyset$, entails that $\nu(\mu\text{-R}) = 0$ and $\mu(\nu\text{-R}) = 0$. Hence, the type of incompatibility induced by algorithmic randomness entails one of the most canonical classical notions of incompatibility between priors: orthogonality. As a result, if μ and ν are radically incompatible with respect to R, then ν does not merge with μ and μ does not merge with ν (in the sense of Blackwell and Dubins). Moreover, if μ and ν are radically incompatible relative to R and μ is locally absolutely continuous with respect to ν, then ν becomes polarized with respect to μ: radical disagreement over which data streams are algorithmically random and local absolute continuity entail the most radical failure of merging of opinions. For an example, take once again λ, $\mu_{\frac{1}{3}}$, and Martin-Löf randomness. Since λ-MLR is a subset of the set of sequences along which the limiting relative frequency of 0 is $\frac{1}{2}$ and $\mu_{\frac{1}{3}}$-MLR is a subset of the set of sequences along which the limiting relative frequency of 0 is $\frac{1}{3}$, it follows that λ-MLR $\cap$ $\mu_{\frac{1}{3}}$-MLR $= \emptyset$. And since λ and $\mu_{\frac{1}{3}}$ are mutually locally absolutely continuous, both agents (the one with prior λ and the one with prior $\mu_{\frac{1}{3}}$) expect their beliefs to be and remain maximally divergent with probability one.

4 Conclusion

The key idea behind the present work is that two computationally limited Bayesian agents beginning the learning process with different priors, but who nonetheless agree on which data streams are algorithmically random, may be thought of as having compatible inductive assumptions about the uniformity of nature. This is because the algorithmically random data streams, while individually chaotic and patternless, display of necessity important global regularities: they have to satisfy various effectively specifiable statistical laws—where the class of effectively specifiable statistical laws to be satisfied varies depending on the algorithmic randomness notion under consideration. As a result, by virtue of agreeing on which data streams are algorithmically random, two computable Bayesian agents may be seen as concurring on what global, effectively specifiable regularities they expect to see in the data.

We saw that using algorithmic randomness to define notions of agreement between computable priors leads to a hierarchy of compatibility notions that provably lead to merging of opinions. More precisely, we focused on the strong notion of merging of opinions introduced by Blackwell and Dubins (1962), and we showed that, apart from weak 1-randomness, all core algorithmic randomness notions give rise to forms of doxastic compatibility that ensure this type of merging (and, *a fortiori*, the weaker type of merging studied by Kalai and Lehrer (1993) and Kalai and Lehrer (1994)). We also saw that disagreement on which data streams are algorithmically random leads to polarization of opinions. This suggests that the theory of algorithmic randomness provides a fruitful framework for identifying and classifying the inductive assumptions of computable Bayesian agents, and for understanding how said inductive assumptions contribute to or hinder the attainment of successful learning in the form of inter-subjective agreement.

We take these results to be but the first step in the systematic study of merging and polarization of opinions through the prism of computability theory and algorithmic randomness. A natural next step is to consider notions of compatibility intermediate between absolute continuity and agreement on weak 1-randomness, to determine exactly what kind of merging of opinions is achievable with weaker assumptions.

Appendix A

In addition to its characterizations in terms of tests and dyadic martingales/supermartingales, Martin-Löf randomness can also be characterized in terms of semi-measures as follows (note that we will once again use the term 'measure' to refer to probability measures):

Theorem A.1 (Folklore)**.** *Let $\omega \in 2^{\mathbb{N}}$ and μ a computable measure. The following are equivalent:*

(1) ω is μ-Martin-Löf random;

(2) $\mu([\omega \upharpoonright n]) > 0$ for all $n \in \mathbb{N}$, and $\limsup_{n\to\infty} \frac{\xi([\omega \upharpoonright n])}{\mu([\omega \upharpoonright n])} < \infty$ for all lower semi-computable semi-measures ξ.

In what follows, we will prove that density randomness has a natural characterization in terms of semi-measures, as well:

Theorem A.2. *Let $\omega \in 2^{\mathbb{N}}$ and μ a computable measure. The following are equivalent:*

(1) ω is μ-density random;

(2) $\mu([\omega \upharpoonright n]) > 0$ *for all* $n \in \mathbb{N}$*, and* $\lim_{n\to\infty} \frac{\xi([\omega \upharpoonright n])}{\mu([\omega \upharpoonright n])}$ *exists and is finite for all lower semi-computable semi-measures* ξ*.*

The proof of Theorem A.2 relies on the following auxiliary lemma:

Lemma A.3. *For all computable measures* μ*, the following holds.*

(i) If d *is a normed dyadic* μ*-martingale, then* $\xi([\sigma]) = d(\sigma)\mu([\sigma])$ *defines a measure. If* d *is left-c.e. or almost-everywhere left-c.e., then* ξ *is lower semi-computable, uniformly from* d*.*

(ii) If ξ *is a semi-measure, then*

$$d(\sigma) = \begin{cases} \frac{\xi([\sigma])}{\mu([\sigma])} & \text{if } \mu([\sigma]) > 0, \\ \text{undefined} & \text{if } \mu([\sigma]) = 0 \end{cases}$$

is a dyadic μ*-supermartingale. If* ξ *is a lower semi-computable semi-measure, then* d *is an almost-everywhere left-c.e dyadic* μ*-supermaringale. If* ξ *is lower semi-computable and* μ *is strictly positive, then* d *is a left-c.e dyadic* μ*-supermaringale.*

Proof. (i) First, note that ξ is well-defined: if $d(\sigma)$ is undefined, then $\mu([\sigma]) = 0$ and $d(\sigma)\mu([\sigma]) = 0$. Now, d being normed means that $d(\varepsilon) = 1$. Hence, $\xi([\varepsilon]) = d(\varepsilon)\mu([\varepsilon]) = 1$. Moreover, for all $\sigma \in 2^{<\mathbb{N}}$,

$$\begin{aligned} \xi([\sigma]) &= d(\sigma)\mu([\sigma]) \\ &= d(\sigma 0)\mu([\sigma 0]) + d(\sigma 1)\mu([\sigma 1]) \\ &= \xi([\sigma 0]) + \xi([\sigma 1]), \end{aligned}$$

where the second identity follows from the fairness condition. Given that d and μ are both non-negative, so is ξ. So, all that we have left to show is that $\xi([\sigma]) \leq 1$ for all $\sigma \in 2^{<\mathbb{N}}$. This follows from a simple argument by induction. We already know that $\xi([\varepsilon]) = 1$. Now, suppose that $\xi([\sigma]) \leq 1$. Then, $\xi([\sigma 0]) = \xi([\sigma]) - \xi([\sigma 1]) \leq \xi([\sigma]) \leq 1$. The reasoning is analogous in the case of $\xi([\sigma 1])$. Hence, ξ is a measure.
Next, suppose that d is left-c.e. (and, thus, total). Let $h : 2^{<\mathbb{N}} \times \mathbb{N} \to \mathbb{Q}$ be a total computable function such that, for all $(\sigma, n) \in 2^{<\mathbb{N}} \times \mathbb{N}$, the sequence $\{h(\sigma, n)\}_{n\in\mathbb{N}}$ is non-decreasing and $\lim_{n\to\infty} h(\sigma, n) = d(\sigma)$. Without loss of generality, h can be assumed to be non-negative. Since μ is computable, it is also lower semi-computable. Therefore, for each $\sigma \in 2^{<\mathbb{N}}$, $\mu([\sigma])$ is a left-c.e. real, uniformly in σ. For each $\sigma \in 2^{<\mathbb{N}}$, let $\{q_{\sigma,n}\}_{n\in\mathbb{N}}$ be a uniformly computable non-decreasing sequence of rationals with $\lim_{n\to\infty} q_{\sigma,n} = \mu([\sigma])$. Without loss of generality, the $q_{\sigma,n}$'s can be assumed to be non-negative. For each $\sigma \in 2^{<\mathbb{N}}$, $\{h(\sigma, n) \cdot q_{\sigma,n}\}_{n\in\mathbb{N}}$ is thus a uniformly computable non-decreasing sequence of rational numbers such that $\lim_{n\to\infty} h(\sigma, n) \cdot q_{\sigma,n} = d(\sigma)\mu([\sigma]) = \xi([\sigma])$.

Hence, $\xi([\sigma])$ is a left-c.e. real, uniformly in σ, which means that ξ is lower semi-computable. If, on the other hand, d is merely almost-everywhere left-c.e., then it is a partial left-c.e. function. Let $h :\subseteq 2^{<\mathbb{N}} \times \mathbb{N} \to \mathbb{Q}$ be a non-negative partial computable function such that (1) for all $\sigma \in 2^{<\mathbb{N}}$, $d(\sigma)$ is defined if and only if $h(\sigma, n)$ is defined for all $n \in \mathbb{N}$, and (2) for all $\sigma \in 2^{<\mathbb{N}}$ such that $d(\sigma)$ is defined, $\{h(\sigma, n)\}_{n\in\mathbb{N}}$ is a non-decreasing sequence with $\lim_{n\to\infty} h(\sigma, n) = d(\sigma)$. As before, for each $\sigma \in 2^{<\mathbb{N}}$, let $\{q_{\sigma,n}\}_{n\in\mathbb{N}}$ be a uniformly computable non-decreasing sequence of non-negative rationals with $\lim_{n\to\infty} q_{\sigma,n} = \mu([\sigma])$. Now, for each $n \in \mathbb{N}$ and each $\sigma \in 2^{<\mathbb{N}}$, let

$$q'_{\sigma,n} = \begin{cases} h(\sigma, n) \cdot q_{\sigma,n} & \text{if } q_{\sigma,n} > 0; \\ 0 & \text{otherwise.} \end{cases}$$

Then, $\{q'_{\sigma,n}\}_{n\in\mathbb{N}}$ is a uniformly computable non-decreasing sequence of rational numbers such that $\lim_{n\to\infty} q'_{\sigma,n} = d(\sigma)\mu([\sigma]) = \xi([\sigma])$, which establishes that ξ is lower semi-computable.
(ii) If $d(\sigma)$ is undefined, then $\mu([\sigma]) = 0$ by definition. Hence, the impossibility condition is satisfied. If $\mu([\sigma]) = 0$, then $\mu([\sigma 0]) = \mu([\sigma 1]) = 0$. Hence, $d(\sigma)\mu([\sigma]) = 0 = d(\sigma 0)\mu([\sigma 0]) + d(\sigma 1)\mu([\sigma 1])$ (again, recall that we follow the convention that $d(\tau)\mu([\tau]) = 0$ if $\mu([\tau]) = 0$ even when $d(\tau)$ is undefined). If, on the other hand, $\mu([\sigma]) > 0$, then we have two cases to consider. First, suppose that $\mu([\sigma 0]) > 0$ and $\mu([\sigma 1]) > 0$. Then,

$$\begin{aligned} d(\sigma)\mu([\sigma]) &= \frac{\xi([\sigma])}{\mu([\sigma])}\mu([\sigma]) \\ &\geq \xi([\sigma 0]) + \xi([\sigma 1]) \\ &= \frac{\xi([\sigma 0])}{\mu([\sigma 0])}\mu([\sigma 0]) + \frac{\xi([\sigma 1])}{\mu([\sigma 1])}\mu([\sigma 1]) \\ &= d(\sigma 0)\mu([\sigma 0]) + d(\sigma 1)\mu([\sigma 1]), \end{aligned}$$

where the inequality holds because ξ is by assumption a semi-measure. Second, suppose that either $\mu([\sigma 0]) = 0$ or $\mu([\sigma 1]) = 0$. Without loss of generality, assume that $\mu([\sigma 0]) = 0$. Then,

$$\begin{aligned} d(\sigma)\mu([\sigma]) &= \frac{\xi([\sigma])}{\mu([\sigma])}\mu([\sigma]) \\ &\geq \xi([\sigma 0]) + \xi([\sigma 1]) \\ &\geq 0 + \xi([\sigma 1]) \\ &= d(\sigma 0)\mu([\sigma 0]) + d(\sigma 1)\mu([\sigma 1]). \end{aligned}$$

Hence, the version of the fairness condition for dyadic supermartingales is satisfied in all cases. Now, suppose that ξ is lower semi-computable. Define the function $h :\subseteq 2^{<\mathbb{N}} \times \mathbb{N} \to \mathbb{Q}$ as follows. If $\mu([\sigma]) = 0$, let $h(\sigma, n)$ be undefined for all $n \in \mathbb{N}$. If $\mu([\sigma]) > 0$, on the other hand, we do the following. Let $\{q_{\sigma,n}\}_{n\in\mathbb{N}}$ be a uniformly computable non-decreasing sequence of (without loss of generality) non-negative rationals witnessing the fact that $\xi([\sigma])$ is a left-c.e. real, uniformly in σ. Since μ is a computable measure, it is also upper semi-computable, which means that $\mu([\sigma])$ is a right-c.e. real, uniformly in σ.[29] Let $\{q'_{\sigma,n}\}_{n\in\mathbb{N}}$ be a uniformly computable non-increasing sequence of positive rationals witnessing the fact that $\mu([\sigma])$ is right-c.e., uniformly in σ. Then, $\lim_{n\to\infty} q'_{\sigma,n} = \mu([\sigma])$. Hence, $\{\frac{1}{q'_{\sigma,n}}\}_{n\in\mathbb{N}}$ is a uniformly computable non-decreasing sequence of positive rationals that converges to $\frac{1}{\mu([\sigma])}$. Define $h(\sigma, n)$ as $\frac{q_{\sigma,n}}{q'_{\sigma,n}}$ for all n. Then, h is a partial computable function, and the sequence $\{h(\sigma, n)\}_{n\in\mathbb{N}}$ is non-decreasing and converges to $\frac{\xi([\sigma])}{\mu([\sigma])} = d(\sigma)$ for all $\sigma \in 2^{<\mathbb{N}}$ such that $d(\sigma)$ is defined (namely, all $\sigma \in 2^{<\mathbb{N}}$ such that $\mu([\sigma]) > 0$). Hence, d is almost-everywhere left-c.e. (and it is left-c.e. if μ is strictly positive). □

We are now ready to prove Theorem A.2, which offers a useful alternative characterization of density randomness.

Proof of Theorem A.2. For the (1)-to-(2) direction, suppose that ω is μ-density random. Then, $\mu([\omega \upharpoonright n]) > 0$ for all $n \in \mathbb{N}$. Now, let ξ be a lower semi-computable semi-measure. By Lemma A.3(ii), $\frac{\xi}{\mu}$ is an almost everywhere left-c.e. dyadic μ-supermartingale. Since ω is μ-density random, we have that $\lim_{n\to\infty} d(\omega \upharpoonright n)$ exists and is finite for all almost everywhere left-c.e. dyadic μ-supermartingales d. Hence, $\lim_{n\to\infty} \frac{\xi([\omega \upharpoonright n])}{\mu([\omega \upharpoonright n])}$ exists and is finite.
For the (2)-to-(1) direction, suppose that $\omega \notin \mu$-DR. If $\mu([\omega \upharpoonright n]) = 0$ for some $n \in \mathbb{N}$, then we are done. So, suppose that $\mu([\omega \upharpoonright n]) > 0$ for all $n \in \mathbb{N}$. Then, there is a left-c.e. dyadic μ-martingale d that fails to converge to a finite value along ω. Without loss of generality, we can assume d to be normed. For each $\sigma \in 2^{<\mathbb{N}}$, let $\xi([\sigma]) = d(\sigma)\mu([\sigma])$. Then, by Lemma A.3(i), ξ is a lower semi-computable measure. But then the sequence $\left\{\frac{\xi([\omega\upharpoonright n])}{\mu([\omega\upharpoonright n])}\right\}_{n\in\mathbb{N}}$ either does not have a limit or $\lim_{n\to\infty} \frac{\xi([\omega \upharpoonright n])}{\mu([\omega \upharpoonright n])} = \infty$. □

Acknowledgements

I am especially grateful to the two anonymous reviewers for their helpful comments and to Johan van Benthem, Ray Briggs, Persi Diaconis, Thomas Icard, Krzysztof Mierzewski, and Brian Skyrms for their careful comments and suggestions, and for countless stimulating conversations. I would also like to thank the participants of the conferences and seminars where I had the opportunity to present this work at various stages of development. In particular (and in

[29] A real number r is rigth-c.e. (right-computably enumerable) if there is a computable non-increasing sequence of rationals that converges to r in the limit: in other words, if r can be approximated from above via a computable sequence of rationals. A measure is upper semi-computable if the probability of each cylinder $[\sigma]$ is a right-c.e. real, uniformly in σ. It is easy to see that a measure is computable if and only if it is both lower semi-computable and upper semi-computable.

reverse chronological order), I am grateful to the audiences at the *Tenth Workshop on Combining Probability and Logic (PROGIC 2021)*, the *Foundations of Probability Seminar* at Rutgers University, the *Workshop on the Foundations, Applications and Theory of Inductive Logic: Interdisciplinarity of Inductive Logic* at the Munich Center for Mathematical Philosophy, and the *Institute for Mathematical Behavioral Sciences Seminar* at the University of California, Irvine. This work was partly funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, grant no. 432308570).

Francesca Zaffora Blando
Department of Philosophy
Carnegie Mellon University
Baker Hall 161
5000 Forbes Avenue
Pittsburgh, PA 15213, USA
fzaffora@andrew.cmu.edu

References

Bienvenu, L. and W. Merkle (2009). "Constructive equivalence relations on computable probability measures". In: *Annals of Pure and Applied Logic* 160, pp. 238–54.

Blackwell, D. and L. Dubins (1962). "Merging of opinions with increasing information". In: *The Annals of Mathematical Statistics* 33, pp. 882–86.

de Finetti, B. (1937). "La Prévision: Ses Lois Logiques, Ses Sources Subjectives". In: *Annales de l'Institut Henri Poincaré* 7, pp. 1–68.

Downey, R. G. and D. R. Hirschfeldt (2010). *Algorithmic Randomness and Complexity*. New York: Springer.

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

Edwards, W., H. Lindman, and L. J. Savage (1963). "Bayesian statistical inference for psychological research". In: *Psychological Review* 3(70), pp. 193–242.

Gaifman, H. and M. Snir (1982). "Probabilities over rich languages, testing and randomness". In: *The Journal of Symbolic Logic* 47, pp. 495–548.

Jaynes, E. T. (1968). "Prior Probabilities". In: *Institute of Electrical and Electronic Engineers Transactions on Systems Science and Cybernetics*, pp. 227–241.

Jeffrey, R. C. (1977). "Mises Redux". In: *Basic problems in methodology and linguistics. Part Three of the Proceedings of the Fifth International Congress of Logic, Methodology and Philosophy of Science, London, Ontario, Canada-1975*. Ed. by R. E. Butts and J. Hintikka. Vol. 11. The Western Ontario Series in Philosophy of Science. D. Reidel, Dordrecht, pp. 213–222.

Kalai, E. and E. Lehrer (1993). "Rational learning leads to Nash equilibrium". In: *Econometrica* 61, pp. 1019–45.

— (1994). "Weak and strong merging of opinions". In: *Journal of Mathematical Economics* 23, pp. 73–86.

Kurtz, S. A. (1981). *Randomness and Genericity in the Degrees of Unsolvability*. University of Illinois at Urbana-Champaign: Ph.D. Thesis.

Martin-Löf, P. (1966). "The Definition of Random Sequences". In: *Information and Control* 9, pp. 602–19.

Miller, R. I. and C. W. Sanchirico (1999). "The role of absolute continuity in 'merging of opinions' and 'rational learning'". In: *Games and Economic Behavior* 1-2(29), pp. 170–90.

Miyabe, K., A. Nies, and J. Zhang (2016). "Using almost-everywhere theorems from analysis to study randomness". In: *Bulletin of Symbolic Logic* 22(3), pp. 305–331.

Muchnik, A. A., A. L. Semenov, and V. A. Uspensky (1998). "Mathematical metaphysics of randomness". In: *Theoretical Computer Science* 207, pp. 263–317.

Nies, A. (2009). *Computability and Randomness*. Oxford: Oxford University Press.

Ramsey, F. P. (1931). "Truth and Probability". In: *Foundations of Mathematics and Other Logical Essay*. Ed. by R. B. Braithwaite. London: Routledge and Kegan Paul, pp. 156–98.

Rosenkrantz, R. D. (1981). *Foundations and Applications of Inductive Probability*. Atascadero, CA: Ridgeview Publishing.

Royden, H. L. and F. M. Fitzpatrick (2010). *Real Analysis*. 4th. London: Pearson.

Rute, J. (2016). "Computable randomness and betting for computable probability spaces". In: *Mathematical Logic Quarterly* 62(4-5), pp. 335–66.

Savage, L. J. (1954). *The Foundations of Statistics*. New York: John Wiley & Sons.

Schervish, M. and T. Seidenfeld (1990). "An approach to consensus and certainty with increasing evidence". In: *Journal of Statistical Planning and Inference* 3(25), pp. 401–14.

Schnorr, C.-P. (1971a). "A unified approach to the definition of a random sequence". In: *Mathematical Systems Theory* 5, pp. 246–58.

— (1971b). *Zufälligkeit und Wahrscheinlichkeit. Eine algorithmische Begründung der Wahrscheinlichkeitstheorie*. Vol. 218. Lecture Notes in Mathematics. Springer.

Skyrms, B. (1984). *Pragmatics and Empiricism*. New Haven and London: Yale University Press.

Suppes, P. (1966). "A Bayesian approach to the paradoxes of confirmation". In: *Studies in Logic and the Foundations of Mathematics*. Vol. 43. Elsevier, pp. 198–207.

Williams, D. (1991). *Probability with Martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge.

Zaffora Blando, F. (2020). *Patterns and Probabilities: A Study in Algorithmic Randomness and Computable Learning*. Stanford University: Ph.D. Thesis.