Philosophy of Science (2023), **90**, 288–306
doi:10.1017/psa.2022.72

PSA PHILOSOPHY OF SCIENCE ASSOCIATION

PHILOSOPHY OF SCIENCE

ARTICLE

A Dilemma for Solomonoff Prediction

Sven Neth

University of California, Berkeley, Berkeley, CA, US
Email: nethsven@berkeley.edu

(Received 09 October 2020; revised 30 December 2021; accepted 15 May 2022; first published online 13 June 2022)

Abstract

The framework of *Solomonoff prediction* assigns prior probability to hypotheses inversely proportional to their Kolmogorov complexity. There are two well-known problems. First, the Solomonoff prior is relative to a choice of universal Turing machine. Second, the Solomonoff prior is not computable. However, there are responses to both problems. Different Solomonoff priors *converge* with more and more data. Further, there are *computable approximations* to the Solomonoff prior. I argue that there is a tension between these two responses. This is because computable approximations to Solomonoff prediction do *not* always converge.

1. Introduction

We are often interested in how to make predictions on the basis of observed data. This question is at the heart of scientific inference and statistics. It is also important for the project of building artificial intelligence (AI) that can make inferences from observed data and act accordingly. Thus, there are many good reasons to be concerned about the right framework for predictive inference.

One way to tackle this question is the *Bayesian approach*, which uses a prior probability distribution over all relevant hypotheses and then updates this prior by conditionalization on the observed data (Earman 1992). The resulting posterior distribution can be used to make predictions and guide action. The Bayesian approach gives us a unified framework for thinking about predictive inference and has been successfully applied across many fields, from astronomy to finance. However, the Bayesian approach requires us to start with a prior probability distribution over all relevant hypotheses. How should we select such a prior probability distribution? This is the *problem of the priors*.

A natural response to the problem of the priors is to say that we should assign a higher prior probability to *simpler* hypotheses. This idea is often known as *Ockham's razor* and seems intuitively appealing to many people. However, how do we measure the simplicity of hypotheses? A possible answer to this question is provided by the framework of *Solomonoff prediction*, which formalizes the simplicity of hypotheses

using tools from algorithmic information theory (Solomonoff 1964; Hutter 2007; Sterkenburg 2016; Li and Vitányi 2019). The *Solomonoff prior* assigns a higher probability to hypotheses that are simpler in this sense. Because the Solomonoff prior is defined for a very broad range of hypotheses, it provides a very general response to the problem of the priors. Moreover, proponents of Solomonoff prediction argue that the Solomonoff prior is an "objective" and "universal" prior. Thus, the framework of Solomonoff prediction potentially sheds light on the foundations of scientific inference, the problem of induction, and our prospects for building "universal AI" (Hutter 2004).

There are two well-known problems for Solomonoff prediction. First, the Solomonoff prior is relative to a choice of universal Turing machine, which means that different choices of universal Turing machine lead to different priors and different predictions. It is natural to worry that this undermines the ambition of Solomonoff prediction to provide an "objective" and "universal" prior. Second, the Solomonoff prior is not computable, which means that no scientist or AI system could actually use the Solomonoff prior to make predictions.

There are well-known responses to both objections. Although it is true that the Solomonoff prior is relative to a choice of universal Turing machine, it can be shown that different Solomonoff priors *converge* with more and more data (in a sense that will be made precise later in the article). Further, although the Solomonoff prior is not computable, there are *computable approximations* to it.

I argue that there is a deep tension between these two responses. This is because different computable approximations to Solomonoff prediction do *not* always converge. Therefore, if we care about universal convergence, computable approximations to Solomonoff prediction do not give us what we want. Thus, proponents of Solomonoff prediction face a pressing dilemma. Either they have to give up universal convergence, which leads to problems of language dependence and subjectivity, or they have to accept that Solomonoff prediction is essentially uncomputable and so cannot be of any help in guiding the inferences of human and artificial agents. Therefore, Solomonoff prediction does *not* solve the problem of finding a universal prior probability distribution that can be used as a foundation for scientific inference and AI.

2. Solomonoff prediction

I start by giving a brief introduction to Solomonoff prediction (Solomonoff 1964; Hutter 2007; Sterkenburg 2016; Li and Vitányi 2019).[1]

Suppose you are given this initial segment of a binary string:

00000000…

Given this initial segment, what is your prediction for the next bit?

In a Bayesian framework, we can answer this question by consulting a *prior probability measure* over the set of all binary strings. To make this answer precise, we first need to introduce some notation. Let $\mathcal{B}^\infty$ be the set of all infinite binary strings and $\mathcal{B}^*$ be the set of all finite binary strings. If $x \in \mathcal{B}^*$ and $y \in \mathcal{B}^* \cup \mathcal{B}^\infty$, we write xy to

[1] For more discussion, see Ortner and Leitgeb (2009), Rathmanner and Hutter (2011), Vallinder (2012), Fulop and Chater (2013), Icard (2017), and Sterkenburg (2018).

https://doi.org/10.1017/psa.2022.72 Published online by Cambridge University Press

denote the *concatenation* of x and y, the (finite or infinite) binary string that starts with x and continues with y. We say that x is a (proper) *prefix* of y if $y = xz$ for some string z (and z is not the empty string).

First, we focus on a particular kind of set of infinite binary strings:

Definition 1. For every $x \in \mathcal{B}^*$, the *cylinder* $\Gamma_x \subseteq \mathcal{B}^\infty$ is defined by $\Gamma_x = \{x\omega : \omega \in \mathcal{B}^\infty\}$ (Li and Vitányi 2019, 265).

Intuitively, a cylinder is a set of binary strings that begin with the same string and then diverge. For example, $\Gamma_1 = \{1\omega : \omega \in \mathcal{B}^\infty\}$ is the set of all binary strings that begin with 1. We write ϵ for the empty string. Therefore, Γ_ϵ is the set of all binary strings that begin with the empty string, which is just the set of all binary strings. We write $\mathfrak{C}$ for the set of all cylinders.

With this framework in place, we can define a probability measure as follows. First, we define:

Definition 2. A *pre-measure* is a function $p : \mathfrak{C} \to [0, 1]$ such that

1. $p(\Gamma_\epsilon) = 1$, and
2. $p(\Gamma_x) = p(\Gamma_{x0}) + p(\Gamma_{x1})$ for all $x \in \mathcal{B}^*$.

Intuitively, a pre-measure assigns probabilities to all cylinder sets. Once we have defined probabilities for all cylinder sets, we can extend our assignment of probabilities to more complicated sets. Let $\mathfrak{F}$ be the result of closing $\mathfrak{C}$ under complementation and countable union. Thus, $\mathfrak{F}$ is a σ-algebra. By Carathéodory's extension theorem, every pre-measure $p : \mathfrak{C} \to [0, 1]$ determines a unique probability measure $p : \mathfrak{F} \to [0, 1]$ that satisfies the standard Kolmogorov axioms.[2] In light of this, we will abuse notation in what follows and sometimes refer to a pre-measure $p : \mathfrak{C} \to [0, 1]$ as a probability measure. If $x \in \mathcal{B}^*$, we will often write $p(x)$ to abbreviate $p(\Gamma_x)$.

Now, the basic idea of Solomonoff prediction is that we should assign a higher prior probability to *simpler* binary strings. However, what do we mean by "simplicity" or "complexity"? We can formalize the complexity of a string as its *Kolmogorov complexity*: the length of the shortest program in some universal programming language which outputs that string. We can model a universal programming language as a monotone universal Turing machine U (Li and Vitányi 2019, 303). A monotone universal Turing machine has a one-way read-only input tape and a one-way write-only output tape. The input tape contains a binary string that is the *program* to be executed, and the output tape contains a binary string that is the *output*. The Turing machine must further be *universal*, which means that it can emulate any computable function. Finally, to say that the Turing machine is *monotone* means that the output tape is write-only, so the machine cannot edit its previous outputs.[3]

[2] Sterkenburg (2018, 64) sketches a more detailed version of this argument. A similar application of Carathéodory's extension theorem is discussed by Earman (1992, 61).

[3] The focus on monotone machines is to ensure, via Kraft's inequality, that the sum in equation (1) is less than or equal to 1 (Li and Vitányi 2019, 275). See also definition 2 in Wood et al. (2013).

https://doi.org/10.1017/psa.2022.72 Published online by Cambridge University Press

Then, we define the *Solomonoff prior*, which assigns prior "probability" to binary strings inversely proportional to their Kolmogorov complexity. For every finite binary string $b \in \mathcal{B}^*$, we have:

$$\lambda_U(b) = \sum_{\rho \in D_{U,b}} 2^{-\ell(\rho)}, \tag{1}$$

where $D_{U,b}$ is the set of minimal programs that lead U to output a string starting with b, and $\ell(\rho)$ is the length of program ρ. To say that $D_{U,b}$ is the set of minimal programs that lead U to output a string starting with b means that (i) upon reading any program in $D_{U,b}$, U will output a string starting with b, and (ii) no proper prefix of any program in $D_{U,b}$ leads U to output a string starting with b.[4] As a rough heuristic, we can think of $\lambda_U(b)$ as the "probability" of producing the string b by feeding random bits to the universal Turing machine U on its input tape. (As we will see in a moment, the Solomonoff prior is not a probability measure, so this is not quite correct.)

As a simple example, consider a binary string that consists of a very long sequence of zeros:

000000000…

Here, $D_{U,b}$ is the set of minimal programs that output a very long sequence of zeros. In Python, one of these might be the following program ρ:[5]

```
while True:
  print(0)
```

In this example, $\ell(\rho)$ is the Kolmogorov complexity of our string because it is the length of one of the minimal programs that outputs our string. To find the Solomonoff prior of our string, we start by computing $2^{-\ell(\rho)}$. However, there might be more than one minimal program that outputs our string. To take this into account, we take the sum over *all* such minimal programs, resulting in the formula in equation (1). As this example shows, there are two assumptions built into this framework. First, strings that are produced by *simpler* programs should get a higher prior probability. Second, strings that are produced by *more* programs should get a higher prior probability.

Each Solomonoff prior $\lambda_U(\cdot)$ induces a Solomonoff predictor, which we can write as follows for every $x \in \mathcal{B}^*$:

$$\lambda_U(x1 \mid x) = \frac{\lambda_U(x1)}{\lambda_U(x)}, \lambda_U(x0 \mid x) = 1 - \lambda_U(x1 \mid x). \tag{2}$$

Intuitively, $\lambda_U(x1 \mid x)$ tells us the probability that the next bit is 1, given that we observed a string starting with x. So if we fix a universal Turing machine U, this answers our earlier question of what we should predict about the next bit after seeing some initial sequence. The hope is that we can encode all real-world inference problems as problems about predicting the next bit of a binary sequence. If this is possible, we can use the Solomonoff predictor to predict any kind of real-word event: the

[4] See Li and Vitányi (2019, 307), Sterkenburg (2016, 466), and Wood et al. (2013, definition 5).

[5] Both here and later in the article, I do *not* claim that these are actually minimal programs but merely use them as simple toy examples.

https://doi.org/10.1017/psa.2022.72 Published online by Cambridge University Press

probability that the sun will rise tomorrow, the probability that the stock market will go up next month, and so on.[6]

As suggested earlier, the Solomonoff prior is *not* a pre-measure on $\mathfrak{C}$. In particular, we only have

1. $\lambda_U(\epsilon) \leq 1$, and
2. $\lambda_U(x) \geq \lambda_U(x0) + \lambda_U(x1)$

for $x \in \mathcal{B}^*$. However, sometimes these inequalities will be strict (Wood et al. 2013, lemma 15). Therefore, the Solomonoff prior is only a *semi-measure*, which we can think of as a "defective" probability measure. This is a problem because there are good reasons to think that rationality requires adherence to the axioms of probability. There are *dutch book arguments*, going back to de Finetti (1937), showing that probabilistically incoherent credences lead agents to accept a sequence of bets that are jointly guaranteed to yield a sure loss. Further, there are *accuracy dominance arguments* showing that probabilistically incoherent credences are guaranteed to be less accurate than some probabilistically coherent credences.[7] Therefore, from a Bayesian point of view, the Solomonoff prior is arguably a nonstarter if it does not satisfy the axioms of probability. Call this the *semi-measure problem*.

To fix this problem, we can define the *normalized Solomonoff prior* Λ_U as follows (Li and Vitányi 2019, 308). We have $\Lambda_U(\epsilon) = 1$, and for every $x \in \mathcal{B}^*$, we recursively define:

$$\Lambda_U(x1) = \Lambda_U(x)\left(\frac{\lambda_U(x1)}{\lambda_U(x0) + \lambda_U(x1)}\right), \Lambda_U(x0) = 1 - \Lambda_U(x1). \tag{3}$$

Λ_U is a pre-measure on $\mathfrak{C}$ and so determines a unique probability measure on $\mathfrak{F}$.[8]

Alternatively, we can interpret the (unnormalized) Solomonoff prior λ_U as a probability measure on the set of infinite and *finite* binary strings (Sterkenburg 2019, 641). From this perspective, cases in which $\lambda_U(x) > \lambda_U(x0) + \lambda_U(x1)$ represent a situation in which λ_U assigns positive probability to the possibility that the binary string ends after the initial segment x.

Does it matter which of these strategies we pick? It turns out that there is an interesting connection between normalization and the approximation reply, to be discussed later in the article. In particular, normalizing the Solomonoff prior makes it *harder* to maintain the approximation reply. But the point of this article is that there is a tension between the approximation reply and the convergence reply,

[6] In any concrete application, our predictions will depend not only on the Solomonoff prior but also on how we encode a given real-world inference problem as a binary sequence. There are many different ways to represent, say, the state of the stock market as a binary sequence. Thus, there is a worry about language dependence here. However, I will bracket this worry because it turns out that there is another, more direct worry about language dependence, to be discussed in section 3.

[7] Standard accuracy arguments are formulated in a setting with a finite algebra of events (Predd et al. 2009; Pettigrew 2016). However, there are extensions of these arguments to infinite algebras (Kelley, forthcoming).

[8] There are different ways to normalize λ_U, which is a potential source of subjectivity and arbitrariness. I will not pursue this line of criticism here. Li and Vitányi (2019, sec. 4.7) provide a great historical overview of the different approaches to the semi-measure problem by Solomonoff, Levin, and others.

https://doi.org/10.1017/psa.2022.72 Published online by Cambridge University Press

and this tension will arise no matter how we deal with the semi-measure problem. Therefore, my main argument is not much affected by this choice.

3. Relativity and convergence

We have defined the Solomonoff prior with reference to a universal Turing machine U. Because there are infinitely many universal Turing machines, there are infinitely many Solomonoff priors. Furthermore, these priors will often disagree in their verdicts. How much of a problem is this? Let us take a closer look.

Consider our previous example. Suppose you are given the initial segment of a binary string:

0000000000…

Given this initial segment, what is your prediction for the next bit?

You might hope that Solomonoff prediction can vindicate the intuitive verdict that the next bit is likely to be a zero. There is an intuitive sense in which a string consisting entirely of zeros is "simple," and you might hope that our formal framework captures this intuition. After all, it seems like the shortest program that outputs a string of all zeros is shorter than the shortest program that outputs a string of ten zeros followed by ones.

In Python, for example, one of the shortest programs to output a string of all zeros might be the following:

```
while True:
  print(0)
```

In contrast, one of the shortest programs to output a string of ten zeros followed by ones might be the following more complicated program:

```
i = 0
while True:
  while i <= 9:
    print(0)
    i = i + 1
  print(1)
```

Thus, it seems reasonable to expect that our Solomonoff predictor should assign a high probability to the next bit being zero.

If you find this kind of reasoning compelling, you might also hope that Solomonoff prediction helps us to handle the "new riddle of induction" and tells us why, after observing a number of green emeralds, we should predict that the next emerald will be green rather than *grue* (either green and already observed, or blue and not yet observed) (Goodman 1955).[9] Both the hypothesis that all emeralds are green and the hypothesis that all emeralds are grue fit our data equally well, but perhaps the all-green hypothesis is simpler and so should get a higher prior probability.[10]

[9] See Elgin (1997) for a collection of classic papers on the "new riddle of induction."

[10] A similar line of argument is suggested by Vallinder (2012, 42).

https://doi.org/10.1017/psa.2022.72 Published online by Cambridge University Press

However, such hopes are quickly disappointed. This is because different universal Turing machines differ in how they measure the Kolmogorov complexity of strings. Relative to a "natural" universal Turing machine, a string with all zeros is simpler than a string with some zeros first and ones after. However, relative to a "gruesome" universal Turing machine, a string with some zeros first and ones afterward is simpler. If we think about the issue in terms of programming languages, this is quite obvious—it all depends on which operations in our programming language are taken to be primitive. Thus, different Solomonoff priors will license different predictions: some will predict that a sequence of zeros will continue with a zero, whereas others will predict that a sequence of zeros will continue with a one. Thus, if we use one of the Solomonoff priors, there is *no guarantee whatsoever* that after observing a long sequence of zeros, we assign a high probability to the next bit being zero.

The argument just sketched is a variant of the familiar point that simplicity is language dependent. Therefore, different choices of language (universal Turing machine) will lead to different priors.[11] Without a principled reason for why a "natural" universal Turing machine should be preferred over a "gruesome" universal Turing machine, the framework of Solomonoff prediction does not give us any reason for why, given an initial sequence of zeros, we should predict that the next bit is a zero rather than a one. Therefore, it does not look like the framework of Solomonoff prediction is any help in distinguishing "normal" and "gruesome" inductive behavior. As a consequence, it does not look like the framework of Solomonoff prediction gives a satisfying solution to the problem of the priors.

However, proponents of Solomonoff predictions can respond to this argument. According to them, the relativity of the Solomonoff prior to a choice of universal Turing machine is not too worrying because one can prove that all Solomonoff priors eventually *converge* toward the same verdicts when given more and more data. Thus, although different choices of universal Turing machine lead to different predictions in the short run, these differences "wash out" eventually. So although there is an element of subjectivity in the choice of universal Turing machine, this subjective element disappears in the limit. Call this the *convergence reply*.[12]

Why is it true that different Solomonoff priors converge in their verdicts? To show this, we can invoke a standard convergence result from Bayesian statistics. To get this result on the table, we first need to introduce a bit more notation. Let p and p' be two probability measures on $\mathfrak{F}$. We define:

Definition 3. *p is absolutely continuous with respect to p'* if for all $A \in \mathfrak{F}$,

$$p(A) > 0 \Rightarrow p'(A) > 0.$$

We now need a way of measuring the difference between two probability functions. Let p and p' be two probability functions on $\mathfrak{F}$. We define:

[11] Readers familiar with Goodman (1955) will recognize that a version of this argument was leveled by Goodman against the idea that "green" is more simple than "grue"—it all depends on your choice of primitives.

[12] This reply is discussed by Rathmanner and Hutter (2011, 1133), Vallinder (2012, 32), and Sterkenburg (2016, 473).

https://doi.org/10.1017/psa.2022.72 Published online by Cambridge University Press

Definition 4. The *total variational distance* between p and p' is

$$sup_{A\in\mathfrak{F}} \mid p(A) - p'(A) \mid .$$

Intuitively, the total variational distance between two probability functions defined on the same domain is the "maximal disagreement" between them. We are interested in what happens after learning more and more data. To capture this, we define:

Definition 5. $E_n : \mathcal{B}^\infty \to \mathfrak{C}$ is the function that, given an infinite binary string $b \in \mathcal{B}^\infty$, outputs the cylinder set of strings that agree with b in the first n places.

Intuitively, E_n is a random variable that tells us the first n digits of the string we are observing.[13] We further define:

Definition 6. A probability function $p : \mathfrak{F} \to [0, 1]$ is *open-minded* if $p(\Gamma_x) > 0$ for all $x \in \mathcal{B}^*$.

This captures the class of probability functions that do not rule out any finite initial sequence by assigning a probability of zero to it.

We want to talk about arbitrary probability functions $p : \mathfrak{F} \to [0, 1]$, so we write $\Delta(\mathfrak{F})$ for the set of all probability functions on $\mathfrak{F}$. Now we define:

Definition 7. For any open-minded probability function $p : \mathfrak{F} \to [0, 1]$, $p(\cdot \mid E_n) : \mathcal{B}^\infty \to \Delta(\mathfrak{F})$ is the function that outputs $p(\cdot \mid E_n(b))$ for each $b \in \mathcal{B}^\infty$.

So $p(\cdot \mid E_n)$ is the result of conditionalizing $p(\cdot)$ on the first n digits of the observed sequence. To make sure that $p(\cdot \mid E_n)$ is always well defined, we restrict our attention to open-minded probability functions.

Now we can invoke the following well-known result in Bayesian statistics (Blackwell and Dubin 1962):[14]

Theorem 1. Let p and p' be two open-minded probability functions on $\mathfrak{F}$ such that p is absolutely continuous with respect to p'. Then, we have

$$\lim_{n\to\infty} sup_{A\in\mathfrak{F}} \mid p(A \mid E_n) - p'(A \mid E_n) \mid = 0,$$

p-almost surely. Therefore, p-almost surely, the total variational distance between p and p' goes to zero as $n \to \infty$.

Let me briefly comment on this result. First, to say that the equality holds "p-almost surely" means that it holds for all binary sequences except perhaps a

[13] One can prove the Bayesian convergence result in a considerably more general setting, working with an abstract probability space and modeling evidence as a sequence of increasingly fine-grained finite partitions (or sub-σ-algebras). However, it is sufficient for our purposes to work with the measurable space $\langle \mathcal{B}^\infty, \mathfrak{F} \rangle$ introduced earlier.

[14] This and related results are discussed extensively by Earman (1992), Huttegger (2015), and Nielsen and Stewart (2018, 2019).

https://doi.org/10.1017/psa.2022.72 Published online by Cambridge University Press

set to which p assigns a probability of zero. Second, as a direct corollary, if p is absolutely continuous with respect to p', and vice versa—so p and p' agree on which events have a prior probability of zero—then p and p' will also agree that, almost surely, their maximal disagreement will converge to zero as they observe more and more data. This captures a natural sense of what it means for p and p' to converge in their verdicts.

With this result in place, the (almost sure) asymptotic equivalence of all Solomonoff priors follows straightforwardly.[15] Let λ_U and $\lambda_{U'}$ be two Solomonoff priors defined relative to two universal Turing machines U and U'. Now $\lambda_{U'}$ is absolutely continuous with respect to λ_U because λ_U *dominates* $\lambda_{U'}$, which means that there is a constant c, depending on U and U', such that for all $x \in \mathcal{B}^*$, we have $\lambda_U(x) \geq c\lambda_{U'}(x)$ (Sterkenburg 2018, 71–72). This is because the shortest programs producing a given string relative to two different universal Turing machines cannot differ by more than a constant, as stated by the *invariance theorem* (Li and Vitányi 2019, 105). Because λ_U and $\lambda_{U'}$ were arbitrary, it follows that all Solomonoff priors are absolutely continuous with respect to each other.

Furthermore, each Solomonoff prior is open-minded. This is because it assigns positive probability to all computable sequences, and every finite sequence is computable. (In the worst case, we can just hard-code the sequence into our program.) Therefore, by Theorem 1, we have

$$\lim_{n \to \infty} sup_{A \in \mathfrak{F}} \mid \lambda_U(A \mid E_n) - \lambda_{U'}(A \mid E_n) \mid = 0,$$

almost surely, so λ_U and $\lambda_{U'}$ converge toward the same verdict. Thus, all the infinitely many Solomonoff priors are (almost surely) asymptotically equivalent.[16]

As another consequence, we can show that any Solomonoff prior converges (almost surely) to optimal predictions on any sequence that is generated by some computable stochastic process (Sterkenburg 2016, 467). This means that we can think about the Solomonoff prior as a "universal pattern detector" that makes asymptotically optimal predictions on the minimal assumption that the data we are observing is generated by some computable process.

There is much more to say about the convergence reply. In particular, worries about subjectivity in the short run remain unaffected by long-run convergence results of the kind just explained (Elga 2016, 314). We still have no argument for why, after observing a finite number of green emeralds, it is more reasonable to predict that the next emerald will be green rather than grue. However, for the sake of argument, I am happy to grant that long-run convergence endows Solomonoff prediction with some kind of desirable objectivity. The focus of my argument is how the emphasis on long-run convergence interacts with another problematic feature of

[15] For the purpose of stating the convergence result, I will assume that the Solomonoff priors are normalized to be probability measures on $\mathfrak{F}$. It is possible to obtain convergence results with the weaker assumption that Solomonoff priors are semi-measures, but there are difficulties in interpreting these results (Sterkenburg 2018, 200), so to simplify our discussion, I'll stick with probability measures.

[16] The "almost sure" qualification matters: it is *not* true that different Solomonoff priors are asymptotically equivalent on *all* sequences, as shown by Sterkenburg (2018, 95), drawing on Hutter and Muchnik (2007). However, this is generally true of Bayesian convergence theorems and no particular problem affecting Solomonoff prediction. For this reason, I will continue to say that different Solomonoff priors are "asymptotically equivalent" and sometimes drop the qualifier "almost surely."

https://doi.org/10.1017/psa.2022.72 Published online by Cambridge University Press

Solomonoff prediction: the fact that the Solomonoff priors are themselves *not* computable.

4. Computability and approximation

There is a second problem for Solomonoff prediction: the infinitely many Solomonoff priors are all uncomputable. This means that there is no possible algorithm that will tell us, after finitely many steps, what the Solomonoff prior of a particular binary sequence *is*—even if we have fixed a choice of universal Turing machine.

Let us first define what it means for a pre-measure $p : \mathfrak{C} \to [0, 1]$ to be computable, following Li and Vitányi (2019, 365):

Definition 8. $p : \mathfrak{C} \to [0, 1]$ is computable if there exists a computable function $g(x, k) : \mathfrak{C} \times \mathbb{N} \to \mathbb{Q}$ such that for any $\Gamma_x \in \mathfrak{C}$ and $k \in \mathbb{N}$,

$$| p(\Gamma_x) - g(\Gamma_x, k) | \; < \frac{1}{k}.$$

This means that a pre-measure $p : \mathfrak{C} \to [0, 1]$ is computable if there is an algorithm that we can use to approximate $p(\Gamma_x)$ to any desired degree of precision for any cylinder set $\Gamma_x \in \mathfrak{C}$.

Then, we have the following:

Theorem 2. For any universal Turing machine U, λ_U is not computable (Li and Vitányi 2019, 303).

Leike and Hutter (2018) discuss further results on the computability of Solomonoff prediction and related frameworks.

Because it seems plausible that we can only use computable inductive methods, this looks like a big problem. It is impossible for anyone to actually use Solomonoff prediction for inference or decision making. The lack of computability also seems to undermine the intended application of Solomonoff prediction as a foundation for AI because it is impossible to build an AI system that uses Solomonoff prediction. One might worry that for this reason, Solomonoff prediction is *completely useless* as a practical guide for assigning prior probabilities. Further, the lack of computability might cut even deeper. It is unclear whether it is even possible for us, or any AI agent we might build, to "adopt" one of the uncomputable Solomonoff priors. I will return to this issue later in the article.

Again, proponents of Solomonoff prediction can respond to this argument. Although it is true that Solomonoff prediction is not computable, it is *semi-computable*, which means that there are algorithms that get closer to $\lambda_U(x)$ at each step. This means that there are algorithms that *approximate* the Solomonoff prior in some sense. Call this the *approximation reply*.[17]

To see how such approximations could work, let me first explain in a bit more detail *why* the Solomonoff prior is not computable. Recall that the Solomonoff prior of a binary string b is inversely proportional to the Kolmogorov complexity of b: the length of the shortest program that outputs b, given some universal Turing machine.

[17] This reply is discussed by Solomonoff (1964, 11; 2009, 8–9).

https://doi.org/10.1017/psa.2022.72 Published online by Cambridge University Press

However, Kolmogorov complexity is not computable.[18] There is no possible algorithm that, given an arbitrary binary string, outputs the Kolmogorov complexity of that string. As a consequence, the Solomonoff prior is not computable.

However, although Kolmogorov complexity is not computable, there are computable approximations to it. To simplify drastically, we can approximate the Kolmogorov complexity of a given string by stopping the search for the shortest program that outputs that string after a fixed time and considering the shortest program *so far* that outputs the string. Call this *bounded Kolmogorov complexity*.[19] We can define a prior that assigns probability inversely proportional to bounded Kolmogorov complexity. As we let the search time go to infinity, we recover the original Kolmogorov complexity of our string.[20]

Given such approximations, one might hope that Solomonoff prediction is still a useful constraint on priors. It provides an ideal for the prior probabilities of a computationally unbounded reasoner, and in practice, we should do our best to approximate this ideal using our finite computational resources. This attitude is expressed, for example, when Solomonoff (1997, 83) writes that "despite its incomputability, algorithmic probability can serve as a kind of 'gold standard' for induction systems."

As before, there is much more to say about this argument, which raises interesting questions about "ideal theorizing" and the value of approximation.[21] However, for the sake of argument, I am happy to grant that there may be something valuable about an ideal theory that can never be implemented but only approximated.

There are some messy details that I'm ignoring here. First, it turns out that the Solomonoff *predictor* is not even semi-computable (Sterkenburg 2019, 651). Furthermore, the normalized Solomonoff prior is not even semi-computable (Leike and Hutter 2018). Both only satisfy the weaker requirement of *limit computability*: there is an algorithm that will converge to the correct probability value in the limit, but it is *not* guaranteed to get closer at each step. These messy details make it harder to maintain the convergence reply because they make it harder to see how we could have *any* sensible method for approximating Solomonoff prediction. However, the point I will discuss next is an *additional* problem, even if these messy details can somehow be cleaned up.

5. A dilemma

When pressed on the relativity of the Solomonoff prior to a universal Turing machine, it is natural to appeal to asymptotic convergence. When pressed on the uncomputability of the Solomonoff prior, it is natural to appeal to computable approximations. However, there is a deep tension between the convergence reply and the approximation reply.

The tension arises for the following reason. Suppose we accept the approximation reply. We hold that although Solomonoff prediction is not computable, we can use

[18] Chaitin, Arslanov, and Calude (1995) provide a direct proof of this fact by reducing the problem of computing Kolmogorov complexity to the halting problem.

[19] See Li and Vitányi (2019, chap. 7) for a rich discussion.

[20] Veness et al. (2011) provide a concrete approximation to Solomonoff prediction. Also see Schmidhuber (2002).

[21] See Staffel (2019) and Carr (forthcoming) for recent discussions of "ideal" versus "nonideal" theorizing in epistemology and the value of approximation.

https://doi.org/10.1017/psa.2022.72 Published online by Cambridge University Press

some computable approximation of Solomonoff prediction to guide our inductive reasoning and construct AI systems. However, this response undercuts the convergence reply because, for reasons I will explain in a moment, *different computable approximations to Solomonoff prediction are not necessarily asymptotically equivalent*. Therefore, we can no longer respond to the worry about language dependence by invoking long-run convergence.

To see why different computable approximations to Solomonoff prediction are not guaranteed to converge, recall first that different Solomonoff priors *do* converge because they are absolutely continuous with respect to each other. Now consider some computable approximation to Solomonoff prediction. There are different ways to spell out what it means to "approximate" the Solomonoff prior, but for my argument, the details of how we think about our "approximation strategy" will be largely irrelevant. As explained earlier, there are considerable difficulties in whether we can make sense of such an approximation strategy for the Solomonoff predictor and normalized Solomonoff prior because they are only limit computable. I will sidestep these difficulties by treating the approximation strategy as a black box—what matters is just that our computable approximation to the Solomonoff prior is *some computable probability measure*.

Why should it be a probability measure, as opposed to a semi-measure? For standard Bayesian reasons: to avoid dutch books and accuracy dominance. Why should it be computable? Because the whole point of the approximation reply is that we can actually use the approximation to make inferences and guide decisions, so we should be able to compute, in a finite time, what the probability of a given event is. Otherwise, the approximation reply seems like a nonstarter.

So let us consider some approximation to Solomonoff prediction, which is some computable probability measure. I claim that this computable approximation must assign probability zero to some computable sequence. This is because *every computable probability measure assigns probability zero to some computable sequence*:

Theorem 3. Let $p : \mathfrak{F} \to [0, 1]$ be a computable probability measure. Then, there is some computable $b \in \mathcal{B}^\infty$ such that $p(b) = 0$.

This result is originally by Putnam (1963), who gives a beautiful "diagonal argument" for it.[22] Consider some computable prior p. Here is how to construct a "diagonal sequence" D for our prior p, where D_i denotes the ith bit of D, and E_n denotes the first n bits of D:

$$D_1 = 0$$

$$D_{n+1} = \begin{cases} 1 & \text{if } \; p(1 \mid E_n) < \frac{1}{2} \\ 0 & \text{if } \; p(1 \mid E_n) \geq \frac{1}{2}. \end{cases}$$

We arbitrarily start our sequence with a zero. To determine the next digit, we first check what our prior p predicts after observing a zero. Then, we do the opposite.

[22] For a wide-ranging discussion of Putnam's argument, see Earman (1992, chap. 9). In statistics, a similar result was shown by Oakes (1985), which is explicitly connected to Putnam's argument by Dawid (1985). See also Schervish (1985).

https://doi.org/10.1017/psa.2022.72 Published online by Cambridge University Press

We iterate this procedure infinitely many times, and our binary sequence D is finished. Because we have assumed that p is computable, D must be computable as well.

Now, why must p assign probability zero to D? Because by construction, $p(D_{n+1} \mid E_n)$ can never go above $1/2$. Therefore, even though the sequence we are observing is generated by a deterministic computable process, our computable prior cannot predict the next bit better than random guessing. However, if $p(D)$ were greater than zero, then $p(D_{n+1} \mid E_n)$ would eventually climb above $1/2$, which contradicts our assumption.

Sterkenburg (2019) discusses the relationship between Solomonoff prediction and Putnam's diagonal argument and concludes that "Putnam's argument stands" (Sterkenburg 2019, 653). In particular, Putnam's argument provides an alternative way to prove that the Solomonoff prior is not computable.[23] My argument here is different because my point is that we can use Putnam's argument to highlight a deep tension between the approximation reply and the convergence reply. Although this tension is a relatively straightforward consequence of Putnam's diagonal argument, this particular point has not received any attention in the debate surrounding Solomonoff prediction. I conjecture that this is because the convergence reply and the approximation reply are often discussed separately, and not enough attention is paid to how they interact with each other. The convergence reply inhabits the realm of "ideal theorizing," where we don't really care about the constraints of computability, whereas the approximation reply tries to connect ideal theory to the real world. However, it is important to pay close attention to how these different features of our theory interact. With this article, I hope to take some steps to remedy this cognitive fragmentation.

Now that I've clarified what this article aims to accomplish, let's get into the argument. Suppose we use a computable approximation to Solomonoff prediction. *The key point is that we face a choice between different approximations that are not guaranteed to be asymptotically equivalent.*

Consider two different computable priors p and p' that approximate Solomonoff prediction in some sense. Note that this could mean two different things: it could mean that we fix a given Solomonoff prior λ_U and use two different "approximation strategies." Alternatively, it could mean that we fix an "approximation strategy" and apply it to two different Solomonoff priors λ_U and $\lambda_{U'}$ based on different universal Turing machines. The second possibility is closely related to the kind of language dependence discussed earlier—we might face the choice between a "natural" and a "gruesome" universal Turing machine. The first possibility seems a bit different; it is best characterized as a kind of "approximation dependence." My argument will work with either of these options.

So we have two computable approximations p and p'. This means, as I have argued earlier, that both p and p' are computable probability measures. By Putnam's argument, both p and p' assign a probability of zero to some computable sequences. Call these sequences D and D'. Note, first, that both p and p' rule out some computable hypotheses and so seem to make substantive assumptions about the world *beyond* computability. For those who hold that Solomonoff prediction gives us a "universal

[23] Further, Sterkenburg (2019, 651) points out that we can use Putnam's argument to show that the Solomonoff predictor is not semi-computable but only limit-computable.

https://doi.org/10.1017/psa.2022.72 Published online by Cambridge University Press

pattern detector" that can find any computable pattern, this is already a problem because the approximations p and p' cannot find *every* computable pattern. This is a first hint that the asymptotic properties that make Solomonoff prediction great are *not* preserved in computable approximations to Solomonoff prediction.

Now, the key point for my argument is that if p and p' are different, then D and D' might be different as well. So p might assign a positive probability to D'. Conversely, p' might assign a positive probability to D. The crucial observation is that although each prior p is forced to assign probability zero to its "own" diagonal sequence D on pain of inconsistency, no inconsistency arises when some prior p assigns positive probability to the diagonal sequence D' for some *other* prior p'.[24]

In the case just discussed, p and p' fail to be absolutely continuous with respect to each other because they differ in what events are assigned a probability of zero. Therefore, it is *not* guaranteed that p and p' are (almost surely) asymptotically equivalent. They might yield different verdicts forever. This means that if there is a subjective element in the choice between p and p', this subjective element is *not* guaranteed to "wash out" in the long run.

To bring this out more clearly, we can draw on a recent result by Nielsen and Stewart (2018). They relax the assumption of absolute continuity and study what happens to Bayesian convergence results in this more general setting. What they show is the following: if prior p is *not* absolutely continuous with respect to prior p', then p must assign some positive probability to the event that p and p' *polarize*, which means that the total variational distance between them converges to 1 as they learn an increasing sequence of shared evidence.[25] So if two priors fail to be absolutely continuous with respect to each other, they must assign positive probability to the event that learning shared evidence drives them toward maximal disagreement.

I have argued earlier that two computable approximations of the Solomonoff prior might fail to be absolutely continuous with respect to each other. In combination with the result by Nielsen and Stewart (2018), this means that two computable approximations of the Solomonoff prior might assign positive probability to polarization in the limit: further evidence drives them toward maximal disagreement. This gives us a clear sense in which, when we consider computable approximations to the Solomonoff prior, subjectivity is *not* guaranteed to "wash out" as we observe more evidence. This, in turn, means that the choice between our two approximations introduces a significant subjective element that is *not* guaranteed to wash out but might, with positive probability, persist indefinitely. This looks like bad news for the convergence reply.

Let me add an important clarification. My argument shows that for two computable approximations p and p' of the Solomonoff prior, it is not guaranteed that p and p' will converge *without making further assumptions*. We might add additional requirements on "acceptable approximations" that rule out such cases by forcing all

[24] Here is a simple example. Let p' be generated by the uniform measure that assigns probability 2^{-n} to each binary sequence of length n. Applying Putnam's construction, the diagonal sequence D' for this prior is the sequence s_0 consisting of all zeros. However, we can easily find *another* (computable) prior p that assigns positive probability to s_0: just let $p(\{s_0\}) = 1$.

[25] See their theorem 3, which generalizes the classic merging-of-opinion results by Blackwell and Dubin (1962).

https://doi.org/10.1017/psa.2022.72 Published online by Cambridge University Press

computable approximations to the Solomonoff prior to be absolutely continuous with respect to each other. However, any such strategy faces a deep problem. Because each computable prior must assign a probability of zero to some computable sequence, this would mean that our set of approximations to the Solomonoff prior rules out some computable sequences a priori. However, this looks incompatible with the motivation behind Solomonoff prediction. The Solomonoff prior is supposed to be a "universal pattern detector" that can learn any computable pattern. So the price for forcing asymptotic agreement among different approximations to the Solomonoff prior would be to make substantive assumptions *beyond* computability, which is exactly what Solomonoff prediction was designed to avoid.

Thus, there is a deep tension between the convergence reply and the approximation reply. If we accept the approximation reply, this means that we should use some computable approximation to the Solomonoff prior to guide our inductive reasoning. However, the move to computable approximations undercuts the convergence reply because different computable approximations are *not* necessarily asymptotically equivalent. They might, with positive probability, yield different verdicts forever and *never* converge to the same predictions. Therefore, we can no longer dismiss the worry about language dependence by invoking long-run convergence. For example, if two different approximations arise from two different universal Turing machines, the difference between "natural" and "gruesome" universal Turing machines is *not* guaranteed to wash out in the long run but might stay with us forever. Therefore, we better come up with some good reasons for why we should use a "natural" rather than a "gruesome" universal Turing machine.[26] More generally, we have to face the problem of subjectivity in the choice of universal Turing machine head-on and cannot downplay the significance of this choice by invoking asymptotic convergence. In fact, the situation is even more bleak: even if we find convincing arguments for why some universal Turing machine is the "correct" or "natural" one, we might still face the choice between different "approximation strategies," which introduces a persistent subjective element. So when we consider computable approximations to Solomonoff prediction, both language dependence and approximation dependence introduce subjective elements that are *not* guaranteed to wash out.

Suppose, on the other hand, that we are convinced by the convergence reply. In this case, we think that what makes Solomonoff prediction great is that different choices of universal Turing machine lead to priors that are (almost surely) asymptotically equivalent and that assign positive probability to all computable sequences. However, in this case, we have to embrace that Solomonoff prediction is essentially uncomputable. This is because there is no computable prior that assigns positive probability to all computable sequences, so the emphasis on convergence undercuts the approximation reply. From this perspective, what makes Solomonoff prediction great is its asymptotic behavior. *However, no computable approximation to Solomonoff prediction preserves this great asymptotic behavior.* Therefore, it is not clear why there is any point in using a computable approximation to Solomonoff prediction to guide our inductive inferences or as a foundation for AI.

[26] Rathmanner and Hutter (2011, 1113) inconclusively explore the issue of whether some universal Turing machines might be more "natural" than others.

https://doi.org/10.1017/psa.2022.72 Published online by Cambridge University Press

You might object to my argument as follows: "Suppose I adopt the Solomonoff prior. In response to the charge that it is not objective, I invoke convergence. In response to the charge that the Solomonoff prior is not computable, I invoke approximation. In response to the charge that these computable approximations need not themselves converge, I simply deny that there is any problem. The computable approximations are not *my probabilities*; they are just useful computational tools that I can use to calculate and report my (approximate) probabilities."[27]

Let me reply to this objection by making clear what the target of my argument is. I grant that if one can really "adopt" one of the Solomonoff priors and use computable approximations merely as a tool to report one's probabilities, this gets around the problem. But is it really possible for us, or an AI agent we build, to adopt an incomputable probability function as a prior? This depends on what makes it the case that an agent has a particular prior, which is a difficult question I cannot fully discuss here. But it seems plausible that any physically implemented agent can only represent and act according to a computable prior. Therefore, it is unclear whether we can really "adopt" an uncomputable prior. The same reasoning holds for any AI system that we might construct. The best we can do is to adopt some approximation to the Solomonoff prior, and my point is that we face some difficult choices in choosing such an approximation.

6. Convergence for subjective Bayesians

Let me finish by briefly discussing how my argument relates to broader questions in Bayesian epistemology. As we have seen at the beginning, one of the big questions for Bayesians is how to choose a prior—the problem of the priors. Solomonoff prediction is an attempt to solve this problem by specifying a "universal" prior. But as I have argued, this ambition ultimately fails because we lose guaranteed convergence if we use computable approximations to the Solomonoff prior.

One might wonder whether this argument poses problems for Bayesian convergence arguments more generally. Bayesians often argue that the choice of prior is not very significant because, given "mild" assumptions, different priors converge as more data are observed.[28] However, the key assumption is absolute continuity: different priors must assign positive probability to the same events. And Putnam's argument shows that every computable prior must assign a probability of zero to some computable hypothesis. Taken together, this suggests that we can only hope for convergence if we agree on substantive assumptions about the world—beyond computability. So the scope of Bayesian convergence arguments is more limited than one might have hoped.[29]

This should not come as a surprise to *subjective* Bayesians who hold that the choice of prior embodies substantive assumptions that reflect the personal beliefs of an agent. Consider, for example, the following passage in Savage (1972) that defends a "personalistic" (subjective Bayesian) view of probability: "The criteria incorporated in the personalistic view do not guarantee agreement on all questions among all

[27] Thanks to an anonymous referee for pressing this objection.

[28] See, for example, the classic discussion by Earman (1992, chap. 6).

[29] This is also the conclusion of Nielsen and Stewart (2018), who argue that Bayesian rationality is compatible with persistent disagreement after learning shared evidence.

https://doi.org/10.1017/psa.2022.72 Published online by Cambridge University Press

honest and freely communicating people, even in principle. That incompleteness, if one will call it such, does not distress me, for I think that at least some of the disagreement we see around us is due neither to dishonesty, to errors in reasoning, nor to friction in communication" (Savage 1972, 67–68).

If you agree that the choice of prior embodies a subjective element, then the fact that we cannot guarantee convergence without shared substantive assumptions should not come as a shock. Thus, my argument does not raise new problems for subjective Bayesians. However, it raises problems for any attempt to define a "universal" or "objective" prior that does not embody substantive assumptions about the world.

7. Conclusion

Proponents of Solomonoff prediction face a dilemma. They cannot simultaneously respond to worries about language dependence by invoking asymptotic convergence while responding to worries about uncomputability by invoking computable approximations. This is because, for very general reasons, no computable approximation to Solomonoff prediction has the same asymptotic behavior as the Solomonoff priors.

In the absence of principled criteria for choosing a universal Turing machine, it looks like Solomonoff prediction is either subject to thorny problems of subjectivity and language dependence or else essentially uncomputable and therefore useless as a guide to scientific inference and the design of optimal artificial agents.

Acknowledgments. I would like to thank Lara Buchak, John MacFarlane, Thomas Icard, and two anonymous referees for helpful comments on earlier drafts. I presented this material at the 27th Biennial Meeting of the Philosophy of Science Association in November 2021 and would like to thank the audience for asking good questions. Further thanks to Snow Zhang, Kshitij Kulkarni, and Reid Dale for helpful discussion and Jürgen Neth for comments on the final manuscript. Special thanks to the editors, who were very helpful and accommodating when I faced some unforeseen challenges in finishing this article. During research, I was supported by a Global Priorities Fellowship from the Forethought Foundation.

References

Blackwell, David, and Lester Dubin. 1962. "Merging of Opinions with Increasing Information." *Annals of Mathematical Statistics* 33 (3):882–86. doi: 10.1214/aoms/1177704456.

Carr, Jennifer Rose. Forthcoming. "Why Ideal Epistemology?" *Mind.* doi: 10.1093/mind/fzab023.

Chaitin, Gregory J., Asat Arslanov, and Cristian Calude. 1995. "*Program-Size Complexity Computes the Halting Problem*." Technical Report CDMTCS-008, Department of Computer Science, University of Auckland, New Zealand.

Dawid, A. Philip. 1985. "Comment: The Impossibility of Inductive Inference." *Journal of the American Statistical Association* 80 (390):340–41. doi: 10.1080/01621459.1985.10478118.

De Finetti, Bruno. 1937. "La Prévision: Ses Lois Logiques, Ses Sources Subjectives." *Annales de l'Institut Henri Poincaré* 17 (1):1–68.

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

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

Elgin, Catherine Z. 1997. *Nelson Goodman's New Riddle of Induction. Vol. 2. The Philosophy of Nelson Goodman.* Milton Park, UK: Taylor & Francis.

Fulop, Sean, and Nick Chater. 2013. "Editors' Introduction: Why Formal Learning Theory Matters for Cognitive Science." *Topics in Cognitive Science* 5 (1):3–12. doi: 10.1111/tops.12004.

Goodman, Nelson. 1955. *Fact, Fiction, and Forecast*. Cambridge, MA: Harvard University Press.

Huttegger, Simon M. 2015. "Merging of Opinions and Probability Kinematics." *Review of Symbolic Logic* 8 (4):611–48. doi: 10.1017/s1755020315000180.
Hutter, Marcus. 2004. *Universal Artificial Intelligence: Sequential Decisions Based on Algorithmic Probability*. Texts in Theoretical Computer Science. New York: Springer. doi: 10.1007/b138233.
Hutter, Marcus. 2007. "On Universal Prediction and Bayesian Confirmation." *Theoretical Computer Science* 384 (1):33–48. doi: 10.1016/j.tcs.2007.05.016.
Hutter, Marcus, and Andrej Muchnik. 2007. "On Semimeasures Predicting Martin-Löf Random Sequences." *Theoretical Computer Science* 382 (3):247–61. doi: 10.1016/j.tcs.2007.03.040.
Icard, Thomas. 2017. "Beyond Almost-Sure Termination." In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*, edited by Glenn Gunzelmann, Andrew Howes, Thora Tenbrink, and Eddy Davelaar, 2255–60. London: Cognitive Science Society.
Kelley, Mikayla. Forthcoming. "On Accuracy and Coherence with Infinite Opinion Sets." *Philosophy of Science.* doi: 10.1017/psa.2021.37.
Leike, Jan, and Marcus Hutter. 2018. "On the Computability of Solomonoff Induction and AIXI." *Theoretical Computer Science* 716:28–49. doi: 10.1016/j.tcs.2017.11.020.
Li, Ming, and Paul Vitányi. 2019. *An Introduction to Kolmogorov Complexity and Its Applications*, vol. 4. New York: Springer. doi: 10.1007/978-3-030-11298-1.
Nielsen, Michael, and Rush Stewart. 2018. "Persistent Disagreement and Polarization in a Bayesian Setting." *British Journal for the Philosophy of Science* 72 (1):51–78. doi: 10.1093/bjps/axy056.
Nielsen, Michael, and Rush Stewart. 2019. "Another Approach to Consensus and Maximally Informed Opinions with Increasing Evidence." *Philosophy of Science* 86 (2):236–54. doi: 10.1086/701954.
Oakes, David. 1985. "Self-Calibrating Priors Do Not Exist." *Journal of the American Statistical Association* 80 (390):339. doi: 10.1080/01621459.1985.10478117.
Ortner, Ronald, and Hannes Leitgeb. 2009. "Mechanizing Induction." In *Handbook of the History of Logic: Inductive Logic*, edited by Dov M. Gabby, Stephan Hartmann, and John Woods, 719–72. Amsterdam: Elsevier. doi: 10.1016/b978-0-444-52936-7.50018-5.
Pettigrew, Richard. 2016. *Accuracy and the Laws of Credence*. Oxford: Oxford University Press.
Predd, Joel, Robert Seiringer, Elliott H. Lieb, Daniel N. Osherson, H. Vincent Poor, and Sanjeev R. Kulkarni. 2009. "Probabilistic Coherence and Proper Scoring Rules." *IEEE Transactions on Information Theory* 55 (10):4786–92. doi: 10.1109/TIT.2009.2027573.
Putnam, Hilary. 1963. "Degree of Confirmation and Inductive Logic." In *The Philosophy of Rudolf Carnap*, edited by Paul Arthur Schilpp, 761–83. La Salle, IL: Open Court.
Rathmanner, Samuel, and Marcus Hutter. 2011. "A Philosophical Treatise of Universal Induction." *Entropy* 13 (6):1076–136. doi: 10.3390/e13061076.
Savage, Leonard J. 1972. *The Foundations of Statistics*, 2nd rev. ed. Hoboken, NJ: Wiley.
Schervish, Mark J. 1985. "Self-Calibrating Priors Do Not Exist: Comment." *Journal of the American Statistical Association* 80 (390):341–42. doi: 10.2307/2287893.
Schmidhuber, Jürgen. 2002. "The Speed Prior: A New Simplicity Measure Yielding Near-Optimal Computable Predictions." In *Computational Learning Theory: Proceedings of COLT 2002*, edited by Jyrki Kivinen and Robert H. Sloan, 216–28. doi: 10.1007/3-540-45435-7_15.
Solomonoff, Ray J. 1964. "A Formal Theory of Inductive Inference. Part I." *Information and Control* 7 (1): 1–22. doi: 10.1016/S0019-9958(64)90223-2.
Solomonoff, Ray J. 1997. "The Discovery of Algorithmic Probability." *Journal of Computer and System Sciences* 55 (1):73–88. doi: 10.1006/jcss.1997.1500.
Solomonoff, Ray J. 2009. "Algorithmic Probability: Theory and Applications." In *Information Theory and Statistical Learning*, edited by Frank Emmert-Streib and Matthias Dehmer, 1–23. New York: Springer.
Staffel, Julia. 2019. *Unsettled Thoughts: A Theory of Degrees of Rationality*. Oxford: Oxford University Press.
Sterkenburg, Tom F. 2016. "Solomonoff Prediction and Occam's Razor." *Philosophy of Science* 83 (4):459–79. doi: 10.1086/687257.
Sterkenburg, Tom F. 2018. "Universal Prediction: A Philosophical Investigation." PhD diss., Rijksuniversiteit Groningen.
Sterkenburg, Tom F. 2019. "Putnam's Diagonal Argument and the Impossibility of a Universal Learning Machine." *Erkenntnis* 84 (3):633–56. doi: 10.1007/s10670-018-9975-x.

Vallinder, Aron. 2012. "Solomonoff Induction: A Solution to the Problem of the Priors?" MA thesis, Lund University.

Veness, Joel, Kee Siong Ng, Marcus Hutter, William Uther, and David Silver. 2011. "A Monte-Carlo AIXI Approximation." *Journal of Artificial Intelligence Research* 40:95–142. doi: 10.1613/jair.3125.

Wood, Ian, Peter Sunehag, and Marcus Hutter. 2013. "(Non-)Equivalence of Universal Priors." In *Algorithmic Probability and Friends. Bayesian Prediction and Artificial Intelligence*, edited by David L. Dowe, 417–25. New York: Springer. doi: 10.1007/978-3-642-44958-1_33.32.

Cite this article: Neth, Sven. 2023. "A Dilemma for Solomonoff Prediction." *Philosophy of Science* 90 (2): 288–306. https://doi.org/10.1017/psa.2022.72