

# That Does Not Compute: David Lewis on Credence and Chance

GORDON BELOT

ABSTRACT. Like Lewis, many philosophers hold reductionist accounts of chance (on which claims about chance are to be understood as claims that certain patterns of events are instantiated) and maintain that rationality requires that credence should defer to chance (in the sense that under certain circumstances one's credence in an event must coincide with the chance of that event). It is a shortcoming of an account of chance if it implies that this norm of rationality is unsatisfiable by computable agents. This shortcoming is more common than one might have hoped.

## 1. INTRODUCTION

David Lewis was a reductionist about chance.<sup>1</sup> He thought that a full description of the pattern of instantiation of fundamental properties and relations at a possible world would not mention any facts about chances. But Lewis suggested that our actual cognitive abilities and limitations determine an ideal scientific theory for each possible world—the theory that best balances (by our standards) simplicity with informativeness about the pattern of events at that world.<sup>2</sup> On his *Best System Account* of chance the true claims about chance at a world are those that are implied by the ideal theory for that world.

Lewis also had an account of the relation between chance and rational credence: the *Principal Principle*.<sup>3</sup> Our primary interest here will be a consequence of the Principal Principle: a credence function is rationally permitted only if, for any event  $E$ , when conditionalized on the chance of  $E$  being  $x$ ,  $c$  assigns credence  $x$  to  $E$ .

For Lewis, the Principal Principle tells us something about what it means to be rational and at the same time encapsulates important truths about the concept of chance.<sup>4</sup> I suggest

Forthcoming in *Philosophy of Science*

Previous versions of the paper were presented at NYU, CMU, and the PSA. For helpful correspondence and suggestions, thanks very much to Dave Baker, Cian Dorr, John Earman, Dmitri Gallow, Kevin Kelly, Alex Meehan, Krzysztof Mierzewski, Richard Pettigrew, Laura Ruetsche, Tom Sterkenburg, Francesca Zaffora Blando, Snow Zhang, and, especially, Chris Porter.

<sup>1</sup>See Lewis (1986, 1994).

<sup>2</sup>Following Loewer (2004), I will speak of *informativeness* where Lewis speaks of *strength* and *fit*.

<sup>3</sup>See again Lewis (1986, 1994).

<sup>4</sup>In addition to Lewis (1986, 1994), see Letters 659 and 660 in Beebe and Fisher (2021).

the following as an example. The Principal Principle gives us possible grounds for criticizing reductionist accounts of chance: if we are thinking of theories of chance as something like summaries meant to be used by beings like us, there is something pathological about a theory of chance relative to which *no* priors count as rationally permitted.

I aim here to explain why it is harder than one might expect to find a satisfying package combining the Best System Account of chance and the Principal Principle.<sup>5</sup> One can show that for a certain *prima facie* attractive version of the Best System Account of chance, the only priors that satisfy the Principal Principle are non-computable. So fans of the Lewisian package must either find a more suitable version of the Best System Account, weaken the Principal Principle, or maintain that rationality requires us to perform tasks beyond the capability of any Turing machine.<sup>6</sup>

## 2. MATHEMATICAL FRAMEWORK

A *binary sequence* is a map  $S : \mathbb{N}^+ \rightarrow \{0, 1\}$ . We denote by  $2^\omega$  the set of all binary sequences. For any binary sequence  $S$  and natural number  $n$ , we write  $S \upharpoonright n$  for the binary string consisting of the first  $n$  bits of  $S$ . If  $\tau$  is a binary string, then we denote by  $\llbracket \tau \rrbracket$  the set of all binary sequences whose initial bits are given by  $\tau$ . We call such  $\llbracket \tau \rrbracket$  the *basic* subsets of  $2^\omega$ .

We will call a subset of  $2^\omega$  an *event* if can be constructed out of basic subsets by taking complements and unions a countable number of times in any order.<sup>7</sup> A *probability measure* on  $2^\omega$  is a map  $\nu$  that assigns numbers between 0 and 1 (inclusive) to events, that assigns 1 to  $2^\omega$ , and which is countably additive (i.e., if  $E_1, E_2, \dots$  are pairwise disjoint events, then  $\nu(\bigcup E_k) = \sum \nu(E_k)$ ). Any such map is completely determined by its behaviour on basic subsets.

A binary sequence  $S$  is *computable* if there is a Turing machine that on input of any  $n \in \mathbb{N}$  gives as output  $S \upharpoonright n$ . A real number  $x$  is *computable* if there exists a Turing machine that

<sup>5</sup>For (all too) full details, see Belot (2022).

<sup>6</sup>It is convenient here to focus on the Lewisian package because of its clarity and precision, but it is far from being the only account of chance, credence, and their relation that faces this sort of problem.

<sup>7</sup>For this (transfinite) construction, see, e.g., Folland (1999, §1.6). The basic subsets generate the product topology on  $2^\omega$ . The family of events is the corresponding  $\sigma$ -algebra of Borel sets.

on input  $n$  returns a rational number within  $1/n$  of  $x$ . A probability measure  $\nu$  on  $2^\omega$  is *computable* if there exists a Turing machine that on input of  $n$  and binary string  $\tau$  returns a rational number within  $1/n$  of  $\nu(\llbracket\tau\rrbracket)$ .

If  $S$  is a binary sequence then the *Dirac measure*  $\delta_S$  concentrated on  $S$  is the probability measure that assigns probability 1 to event  $E$  if  $S \in E$  and assigns probability 0 to  $E$  otherwise. A Dirac measure is computable if and only if the sequence it is concentrated on is.

If  $r$  is a number (strictly) between zero and one, then the *Bernoulli measure*  $\nu_r$  with parameter  $r$  is the probability measure that, for any binary string  $\tau$  containing  $\ell$  0's and  $m$  1's, assigns the basic subset  $\llbracket\tau\rrbracket$  probability  $r^\ell(1-r)^m$ . We call  $\nu_{.5}$  the *fair coin measure*. A Bernoulli measure is computable if and only if its parameter is.

### 3. WORLDS

We will be concerned with simple worlds whose histories can be encoded in binary sequences: at each such world time has the structure of the natural numbers; and the history of such a world just tells us, for each time, whether or not a single property is instantiated at that time. We work with these worlds in order to have a precise and tractable framework in which to interpret Lewis's ideas about chance and credence. The lessons we learn can be translated to richer settings (our 'possible worlds' could always be interpreted as subsystems of more interesting worlds).

A probability measure  $\nu$  on  $2^\omega$  can be thought of as an instruction manual for building a world. Nature is equipped with a set of coins with each possible bias. A probability measure  $\nu$  on  $2^\omega$  tells Nature which coin to flip a coin with bias  $\nu(\llbracket0\rrbracket)$  to determine whether the first bit is a 0. More generally,  $\nu$  tells nature to flip a coin of bias  $\nu(\llbracket\tau0\rrbracket \mid \llbracket\tau\rrbracket)$  to determine whether the next bit will be 0, given that  $\tau$  gives the history so far.<sup>8</sup> A Dirac measure  $\delta_S$  instructs Nature to construct a history by flipping coins of maximal bias in a way guaranteed

---

<sup>8</sup>So a probability measure on  $2^\omega$  gives the complete set of what Lewis (1986, 97) calls "history to chance conditionals"—and hence gives what he calls "the complete theory of chance" for a world.

to generate  $S$ . A Bernoulli measure  $\nu_r$  tells Nature to generate each bit by flipping a coin of bias  $r$ .

#### 4. CHANCE AND CREDESCENCE

Reductionism about chance is the thesis that chance-y facts supervene on non-chance-y facts. So in our context, a reductionist account of chance can be encoded in a map from  $2^\omega$  to the space of probability measures over  $2^\omega$  (this map will be merely partially defined if there are lawless worlds without chance facts). We call a probability measure  $\lambda$  a *chance law* of such an account if the relevant map assigns it to some world.

For present purposes, a Bayesian prior is a probability measure on  $2^\omega$  that assigns positive probability to each basic subset of  $2^\omega$ .<sup>9</sup>

It has seemed to many that rationality requires credence to defer to chance in a certain sense: in some situations, if you are rational, then your credence in an event must coincide with the chance of that event.

Knowing only that the chance of drawing a red ball from an urn is 0.95, everyone agrees, in accordance with the law of likelihood, that a guess of ‘red’ about some trial is much better supported than one of ‘not-red.’ But nearly everyone will go further, and agree that 0.95 is a good measure of the degree to which ‘red’ is supported by the limited data. (Hacking 1965, 136)

[T]he chancemaking pattern in the arrangement of qualities must be something that would, if known, correspondingly constrain rational credence. Whatever makes it true that the chance of decay is 50% must also, if known, make it rational to believe to degree 50% that decay will occur. (Lewis 1994, 478)

Fix a reductionist account of chance (such as the Best System Account). Corresponding to any chance law  $\lambda$  of the theory, there is the set of worlds  $\Lambda$  at which  $\lambda$  gives the chance facts. We call this set the *chance hypothesis* corresponding to that law. This is the chancemaking proposition: the proposition that the chance facts are given by  $\lambda$ .

---

<sup>9</sup>So we are forbidding rational agents to be certain *a priori* that any particular finite data set will never arise. Note that the set of binary strings is countable.

*Definition 1* (Chance-Credence Principle). Relative to a given reductionist account of chance, a prior  $\mu$  on  $2^\omega$  satisfies the *Chance-Credence Principle* if for any chance law  $\lambda$  with chance hypothesis  $\Lambda$  and for any event  $A$  we have  $\mu(A \mid \Lambda) = \lambda(A)$ .

The requirement that rational priors should satisfy the Chance-Credence Principle for a given reductionist theory of chance is a weak form of Lewis's Principal Principle.<sup>10</sup>

A prior  $\mu$  satisfies the Chance-Credence Principle for a given reductionist theory of chance if and only if: (i) the theory of chance admits only countably many chance laws  $\lambda_1, \lambda_2, \dots$  (with chance hypotheses  $\Lambda_1, \Lambda_2, \dots$ ); (ii) each chance law  $\lambda_k$  of the theory is *proper*, in the sense that  $\lambda_k(\Lambda_k) = 1$ ; and (iii)  $\mu$  can be written as a weighted sum of probability measures in which each  $\lambda_k$  appears with non-zero weight (and in which any other summands are concentrated on the set of lawless worlds).

## 5. THE BEST SYSTEM ACCOUNT

Lewis (1994, §4) tells us a bit about the map that encodes his favoured reductionist theory of chance. He works in the context of finite worlds that can be encoded in binary strings. In this context, under a straightforward frequentist approach a world  $\sigma$  would be assigned a Bernoulli measure  $\nu_r$  as its chance law if and only if  $r$  gives the relative frequency of 0's at  $\sigma$ . Lewis emphasizes that his Best System Account departs in two ways from this sort of frequentism.

- i) Some worlds will not be assigned a Bernoulli measure as their Best-System chance law: some worlds exhibit patterns that render it natural to think of Nature as generating bits by following instructions that call for a coin of bias  $r_1$  to be flipped to generate the first bit; a coin of bias  $r_2$  to be flipped to generate the second bit; and so on. As an extreme case, Lewis mentions a world where history alternates between 0 and 1, which is naturally thought of as generated by alternating between flipping coins with maximal bias in favour of 0 and 1—in other words, the chance facts there are encoded in a Dirac measure.

---

<sup>10</sup>On this point see Lewis (1986, 276 f.) and Pettigrew (2012, fn. 4).

ii) On the other hand, Lewis thinks that in some cases the chance law at a world should be a Bernoulli measure  $\nu_r$  even though  $r$  does not give the relative frequency of 0's at that world (e.g., in a case where the relative frequency of 0's at a world is given by a messy number  $r^*$  close to .5 then the fair coin measure may provide a better candidate than  $\nu_{r^*}$  to be the Best-System chance law of the world—although of course for long-lived worlds, the chance of getting a relative frequency of 0's differing even a little from .5 by tossing a fair coin will be minuscule).

How does this look when transposed to the setting of worlds whose histories are coded in infinite binary sequences? It is helpful to note two important features of the infinite context.

- 1) The idea behind the Best System Approach is that the chance facts at a world are given by the best scientific summary of that world—where *best* means something like *best by our ordinary scientific standards*. It is crucial to this picture, then, that the systems in competition be finitarily specifiable objects—otherwise, the question which of them is best according to ordinary scientific standards loses all sense. The specification of an arbitrary real number is an infinitary task—requiring, in general, the specification of infinitely many bits. So the specification of an arbitrary probability measure on  $2^\omega$  is infinitary twice over—requiring, in general, the assignment of a real number to each of the infinitely many binary strings. Now, as Turing (1936, 236) tells us: “The ‘computable’ numbers may be described briefly as the real numbers whose expressions as a decimal are calculable by finite means.” In the same way, the finitarily specifiable probability measures are the computable probability measures. So in the infinite context, only computable measures should be eligible candidates to be chance laws under the Best System Approach.
- 2) In this setting there is a natural way to make precise the notion of a *random-looking* sequence: the notion of a Martin-Löf random sequence.<sup>11</sup> If  $S$  is a binary sequence and  $\mu$  is a computable probability measure, then, roughly speaking,  $S$  is  $\mu$ -Martin-Löf random just in case  $S$  exhibits no finitarily specifiable patterns of behaviour that would be arbitrarily

---

<sup>11</sup>For details, see, e.g., Shen *et al.* (2017, Chapter 3). Here is the rough idea: a sequence is *not* Martin-Löf random relative to a given computable measure  $\mu$  just in case there is an effectively definable  $\mu$ -nullset of  $2^\omega$  to which it belongs.

surprising to an agent expecting to see a data stream generated by  $\mu$ .<sup>12</sup>

*Example 1.* If a computable Bernoulli measure  $\nu_r$  considers a binary sequence  $S$  to be Martin-Löf random then:  $S$  is non-computable; the relative frequency of 0's in  $S$  is  $r$ ; each finite string occurs infinitely often in  $S$  (and with the right relative frequency); as one looks at longer and longer initial segments, one finds that the relative frequency of 0's is greater than  $r$  infinitely often (and less than  $r$  infinitely often); and, further,  $S$  satisfies every other effectively specifiable criterion for a data stream to be sampled from  $\nu_r$ .

I claim that any plausible version of the Best-System Analysis should conform to the following constraints.

BS1. If  $S$  is assigned chance law  $\lambda$ , then  $S$  is Martin-Löf random relative to  $\lambda$ .

BS2. Each computable sequence is assigned a law of chance.

BS3. A sequence is assigned a computable Bernoulli measure  $\nu_r$  as its law of chance if and only if its is  $\nu_r$ -Martin-Löf random.

*Remark 1 (On BS1).* Let  $S$  be a sequence and  $\lambda$  a computable measure. If  $S$  is not  $\lambda$ -Martin-Löf random then a computable agent whose data consists of sufficiently long initial segments of  $S$  will eventually be in a position to reject *at any level of significance* the hypothesis that the data is being sampled from  $\lambda$ —which would appear to rule out the possibility that the Best System Account ratifies  $\lambda$  as the chance law for  $S$ . Note that each computable probability measure assigns measure one to the set of sequences that it considers Martin-Löf random.<sup>13</sup>

*Remark 2 (On BS2).* Let  $S$  be computable. Then  $\delta_S$  is *prima facie* a very good candidate to be a chance law for  $S$ :  $S$  is  $\delta_S$ -Martin-Löf random;  $\delta_S$ , being computable, is only finitely

<sup>12</sup>If  $\mu = 10^{-6}\delta_{\bar{0}} + (1 - 10^{-6})\delta_{\bar{1}}$  (where  $\bar{0}$  is the all-0's sequence and  $\bar{1}$  is the all-1's sequence), then an agent expecting to see data sampled from  $\mu$  will be very surprised to see a data set beginning with an initial 0—but no more surprised to see an all-0 data set of any larger size. An agent expecting to see data sampled from the fair coin measure would not be very surprised by an initial 0—but would be arbitrarily surprised by arbitrarily large all-0 data sets.

<sup>13</sup>Since there are only countably many Turing machines, there can be only countably many effective  $\mu$ -nullsets—and the union of countably many  $\mu$ -nullsets is a  $\mu$ -nullset.

complex; at the same time, it gives a maximally informative characterization of the infinite object  $S$ . It may be that some simpler measure achieves a better balance of simplicity and informativeness (maybe  $S$  begins with a long string that looks like it was generated by flipping a fair coin before settling down to a regular pattern).<sup>14</sup> So perhaps  $\delta_S$  is not the Best-System chance law of  $S$ —but given that  $\delta_S$  is available as an option, it is implausible that our ordinary standards would decree that  $S$  has no chance law whatsoever. Note that it then follows via BS1 that the chance law  $\lambda$  of a computable sequence  $S$  must satisfy  $\lambda(\{S\}) > 0$ .<sup>15</sup>

*Remark 3 (On BS3).* Suppose that  $S$  is Martin-Löf random relative to a computable Bernoulli measure  $\nu_r$ . Then there is a sense in which  $\nu_r$  provides a maximally simple theory of  $S$ : if  $S$  is also Martin-Löf random relative to  $\mu$ , then as  $\mu$  is conditionalized on longer and longer initial segments of  $S$ , the probabilities that it gives for the next bit to be 0 must converge to  $r$ .<sup>16</sup> It is easy, however, to generate rivals to  $\nu_r$  that are in a sense more informative than it: if  $\tau$  is any initial segment of  $S$ , then  $S$  will be Martin-Löf random relative to the measure  $\nu_r^\tau$  that says that the initial bits of history are given by  $\tau$  and the rest generated by tossing a coin of bias  $r$ . So, intuitively, if  $S$  begins in some striking way—say with a hundred consecutive 0’s—but otherwise looks like it was generated by flipping a fair coin, then a measure of the form  $\nu_{.5}^\tau$  would offer a better balance of informativeness and simplicity than  $\nu_{.5}$  does. But I think that this intuition is an artifact of our habitual attention to small data sets. Maybe if the first three tosses of a coin come up Heads, we are a bit suspicious that the coin is not fair—but this suspicion is washed away when we see enough data to be reassured that each three-bit string occurs equally often (and in no special pattern). We should have the same reaction to a random-looking binary sequence that begins with one hundred 0’s: by the time you have seen the billionth block of a billion consecutive 0’s, that initial hundred will not look impressive anymore—because you will have seen that each one hundred-bit string occurs (essentially) equally often (and in no special pattern).

---

<sup>14</sup>Thanks to Cian Dorr for emphasizing this point to me.

<sup>15</sup>Since  $S$  is computable,  $\{S\}$  is an effectively definable  $\lambda$ -nullset if it is a  $\lambda$ -nullset.

<sup>16</sup>See Vovk (1987, Theorem 3).



## 6. TROUBLE

BS1–BS3 and the Chance–Credence Principle are mutually consistent. Consider the version of the Best-System Account in which each computable sequence is assigned the corresponding Dirac measure as its chance law, each non-computable sequences is assigned a computable Bernoulli measure as its chance law if and only if it is Martin-Löf random with respect to it; and all other sequences are lawless. Then any prior that can be written as a (nontrivial) weighted sum of each of the countably many computable Dirac measures and computable Bernoulli measures will satisfy the Chance-Credence Principle.

But no *computable* prior can satisfy the Chance–Credence Principle for any reductionist account of chance obeying BS1 and BS2. As noted above, BS1 and BS2 imply that each computable sequence  $S$  is assigned a chance law that assigns  $\{S\}$  positive probability. It follows that if a prior  $\mu$  satisfies the Chance–Credence Principle for an account of chance obeying BS1 and BS2, then it must be possible to write  $\mu$  as a weighted sum of probability measures in which each computable Dirac measure appears with positive weight. We will show that no such prior can be computable.<sup>17</sup> Recall that  $S \upharpoonright k$  is the binary string consisting of the first  $k$  bits of  $S$ . So  $\mu(\llbracket S \upharpoonright (n+1) \rrbracket \mid \llbracket S \upharpoonright n \rrbracket)$  is the probability that  $\mu$  gives to seeing the next bit of  $S$  after having been shown the first  $n$  bits of  $S$ . Let us say that prior  $\mu$  *learns* sequence  $S$  if after seeing sufficiently long initial segments of  $S$ ,  $\mu$  is always able to correctly predict the next bit with credence exceeding some fixed cut-off.<sup>18</sup> Consider any prior  $\mu$  that can be written in the form

$$\mu := \nu + c \cdot \delta_S,$$

where  $S$  is a binary sequence,  $c > 0$ , and  $\nu$  is a measure that considers  $\{S\}$  a nullset. Since  $\nu(\{S\}) = 0$ , we can make  $\nu(\llbracket S \upharpoonright n \rrbracket)$ —and hence also  $\nu(\llbracket S \upharpoonright (n+1) \rrbracket)$ —as small as we like by choosing  $n$  sufficiently large.<sup>19</sup> It follows that  $\mu$  learns  $S$ . So, in particular, if  $\mu$  satisfies the

<sup>17</sup>The following is a variant of the diagonalization argument of Putnam (1963).

<sup>18</sup>For what follows, we need to choose a computable cutoff  $C$  such that there is no string  $\tau$  with  $\mu(\llbracket \tau \rrbracket) = C$ . This is always possible (otherwise we could use  $\mu$  to show that the computable sequences are uniformly computable—which they are not).

<sup>19</sup>Because probability measures are continuous from above—see, e.g., Folland (1999, Theorem 1.8).

Chance-Credence Principle for a Best System Account obeying BS1 and BS2, then  $\mu$  must learn each computable sequence. But given any computable prior  $\mu$ , we can construct a computable binary sequence  $S_\mu^\dagger$  that  $\mu$  does not learn:  $S_\mu^\dagger$  consists of blocks of 0's separated from each other by individual 1's; a 1 is called for whenever  $\mu$  has just seen at least one 0 and thinks that it is more likely than not that the next bit will be a 0. So no computable prior can satisfy the Chance-Credence Principle for a reductionist account of chance obeying BS1 and BS1.

A variant of this argument shows that no computable probability measure can be written as a weighted sum of probability measures in which each computable Bernoulli measure appears with non-zero weight—from which it follows that no computable prior can satisfy the Chance-Credence Principle for any reductionist account of chance obeying BS3.

## 7. OPTIONS

Roughly speaking, the problem we have run into is that on Lewis's Best System Account, the chance laws at a world are just good scientific summaries of the patterns of events at that world; and his Principal Principle implies that if you are rational, then to the extent that you are confident in a chance hypothesis it should guide your estimates of probability. Taken together, the package seems to imply that in order to be rational, you need to be a certain sort of universal learner: whatever the chance law at your world, if you are a good scientist and you see enough data, you should become confident in something like that law and so you should eventually mimic it in your estimates of probability. As Lewis (1986, 121) puts it: “if we start with a reasonable initial credence function and do enough feasible investigation, we may expect our credences to converge to the chances.” But this is a setting in which no computable universal learner can exist.

I do not think we should rest content with an account of chance and credence that tells us that rationality requires us to adopt a non-computable credence function—any more than we would be satisfied with an account of rationality that required rational agents to be able to solve the Halting Problem. So I think we ought to be interested in revising some part of the

framework used above, rejecting one or more of BS1–BS3, or weakening the Chance-Credence Principle. I end with a few observations about these options.

- 1) Would countenancing rational credal states represented by merely finitely additive probability measures help? No. For suppose that a prior  $M$  of this kind exists such that: (i) for any computable  $S$ ,  $M(\{S\}) > 0$ ; and (ii)  $M(\llbracket\tau\rrbracket)$  is defined for each binary string  $\tau$ . Then there must exist a countably additive  $\mu$  such that  $\mu(\llbracket\tau\rrbracket) = M(\llbracket\tau\rrbracket)$  for each string  $\tau$  and such that  $\mu(\{S\}) > 0$  for each computable sequence  $S$ .<sup>20</sup> Our diagonalization argument above then tells us that such an  $M$  cannot be computable even in the sense that its restriction to basic sets is computable.
- 2) Our Chance-Credence Principle requires that each law of chance  $\lambda$  be proper, in the sense that it assigns probability one to its own chance hypothesis  $\Lambda$ . We could relax this restriction to allow  $\lambda(\Lambda) \in (0, 1]$ . In that case it would be natural to replace the Chance-Credence Principle by the following.<sup>21</sup>

*Definition 2* (New Chance-Credence Principle). Relative to a given reductionist account of chance, a prior  $\mu$  on  $2^\omega$  satisfies the *New Chance-Credence Principle* if for any chance law  $\lambda$  with chance hypothesis  $\Lambda$ , and for any event  $A$  we have  $\mu(A \mid \Lambda) = \lambda(A \mid \Lambda)$ .

But this turns out to be of no help with the problem we have run into above concerning accounts of chance satisfying BS1 and BS2.

- 3) Some philosophers amend the Principal Principle to require priors to satisfy substantive conditions relative to chance laws only if they assign the corresponding chance hypotheses positive credence.<sup>22</sup> I reject such approaches on the grounds that the amended principle is too weak to embody the requirement of learnability of chances by rational priors that Lewis, rightly in my mind, built into the Principal Principle (consider, e.g., a prior concentrated on the lawless worlds of a reductionist account of chance).

<sup>20</sup>See Propositions 3.2.7 and 3.3.1 of Bhaskara Rao and Bhaskara Rao (1983).

<sup>21</sup>This stands to the New Principle of Hall (1994) and Lewis (1994) as the Chance-Credence Principle stands to the Principal Principle of Lewis (1986). For another direction in which one might generalize the Principal Principle in the face of improper laws of chance, see Roberts (2001) and Ismael (2008). For critical discussion, see Briggs (2009).

<sup>22</sup>For this approach, see, e.g., Pettigrew (2012).

- 4) We could replace BS2 with the assumption that the set of computable sequences to be assigned laws of chance is a set that is learnable in sense considered above (and proceed analogously with BS3). There is a sense in which the sets of sequences so-learnable are those that can be computed quickly.<sup>23</sup> So in making this move, there is a sense in which the computable sequences that we deem lawless are ones that exceed a certain complexity cut-off—which might seem consonant with the spirit of the Best-System Approach. Maybe. But there is always going to be a worry about arbitrariness here: if a set of computable binary sequences is learnable in the relevant sense, then so is the result of enlarging that set by adding any finite set of computable binary sequences to it; and if two sets of computable sequences are learnable in this sense, then so is their union.
- 5) There are many notions of learning that one might substitute for the one implicit in the Principal Principle—and thereby construct weaker alternatives to the Chance-Credence Principle that may comport better with computability. One possibility would be to require that as a rational prior is conditionalized on longer and longer initial segments of a (typical) sequence associated with a given chance law, its probability estimates for the next bit converge to those given by that chance law (conditional on the same data).<sup>24</sup> An interesting feature of this requirement is that it is consistent with BS3: the so-called *indifference* (or *Bayes-Laplace*) prior meets the requisite condition with respect to each computable Bernoulli measure. But problems remain with BS1 and BS2: a family of Dirac measures is learnable in this weaker sense if and only if it is learnable in the sense relevant to the Chance-Credence Principle.

The moral is: work remains to be done for anyone who is attracted to Lewis's accounts of chance and credence and who is also inclined to take computability to be a constraint on rational priors.

---

<sup>23</sup>See Blum and Blum (1975, 127).

<sup>24</sup>For this notion see Kalai and Lehrer (1994).

## REFERENCES

- Beebee, Helen and A. R. J. Fisher, eds. 2021. *Philosophical Letters of David Lewis*, vol. 1. Oxford: Oxford University Press.
- Belot, Gordon. 2022. “Unprincipled.” Unpublished.
- Bhaskara Rao, K. P. S. and M. Bhaskara Rao. 1983. *Theory of Charges: A Study of Finitely Additive Measures*. London: Academic Press.
- Blum, Lenore and Manuel Blum. 1975 “Toward a Mathematical Theory of Inductive Inference.” *Information and Control* 28:125–55.
- Briggs, R. A. 2009. “The Anatomy of the Big Bad Bug.” *Noûs* 43:428–49.
- Folland, Gerald. 1999. *Real Analysis: Modern Techniques and Their Applications* (second edition). New York: Wiley.
- Hacking, Ian. 1965. *Logic of Statistical Inference*. Cambridge: Cambridge University Press.
- Hall, Ned. 1994. “Correcting the Guide to Objective Chance.” *Mind* 103:505–17.
- Ismael, Jenann. 2008. “Raid! Dissolving the Big, Bad Bug.” *Noûs* 42:292–307.
- Kalai, Ehud and Ehud Lehrer. 1994. “Weak and Strong Merging of Opinions.” *Journal of Mathematical Economics* 23:73–86.
- Lewis, David. 1986. “A Subjectivist’s Guide to Objective Chance.” In *Philosophical Papers*, vol. II, 83–132. Oxford: Oxford University Press.
- Lewis, David. 1994. “Humean Supervenience Debugged.” *Mind* 103:473–90.
- Loewer, Barry. 2004. “David Lewis’s Humean Theory of Objective Chance.” *Philosophy of Science* 71:1115–25.
- Pettigrew, Richard. 2012. “Accuracy, Chance, and the Principal Principle.” *Philosophical Review* 121:241–75.
- Putnam, Hilary. 1963. *Probability and Confirmation*. Washington, DC: United States Information Agency.

- Roberts, John. 2001. "Undermining Undermined: Why Humean Supervenience Never Needed to be Debugged (Even If It's a Necessary Truth)." *Philosophy of Science* 68:S98–S108.
- Shen, Alexander, Vladimir Uspensky, and Nikolay Vereshchagin. 2017. *Kolmogorov Complexity and Algorithmic Randomness*. Providence, RI: American Mathematical Society.
- Turing, Alan. 1936. "On Computable Numbers, with an Application to the *Entscheidungsproblem*." *Proceedings of the London Mathematical Society* 42:230–65.
- Vladimir Vovk. 1987. "On a Criterion for Randomness." *Soviet Mathematics Doklady* 35:656–60.