A CLASSICAL WAY FORWARD FOR THE REGULARITY AND NORMALIZATION PROBLEMS

ALEXANDER R. PRUSS

Abstract. Bayesian epistemology has struggled with the problem of regularity: how to deal with events that in classical probability have zero probability. While the cases most discussed in the literature, such as infinite sequences of coin tosses or continuous spinners, do not actually come up in scientific practice, there are cases that do come up in science. I shall argue that these cases can be resolved without leaving the realm of classical probability, by choosing a probability measure that preserves "enough" regularity. This approach also provides a resolution to the McGrew, McGrew and Vestrum normalization problem for the fine-tuning argument.

1. Introduction

The problem of regularity in Bayesian epistemology is that we would like our probability distributions to assign non-zero probability to every event that is possible. Perhaps the main reason for this condition is that that in science our evidence appears to be is given in terms of events that have non-zero probability, and a zero probability event stays at zero probability after Bayesian conditionalization on non-zero probability evidence E, thereby ruling out the possibility of ever learning that A is true. And if we try to overcome this by insisting that we have some zero probability evidence, we run into serious technical problems with conditionalizing on that evidence as the usual ratio formula $P(A \mid E) = P(A \cap E)/P(E)$ turns into zero-divided-by-zero, and solutions to this difficulty in turn tend to lead to the Borel-Kolmogorov paradox (Borel 1909) as well as problems with preservation of intuitive symmetries (e.g., Pruss 2015 and Pruss 2020). And when we have uncountably many mutually incompatible events, such as when a dart is thrown at a continuous target, it is mathematically impossible to assign a positive real number probability to all of them: most such events will have to get probability zero given finite additivity (Hájek 2003).

The problem of normalization comes up perhaps most vividly in the context of the McGrew, McGrew and Vestrup (2001) [MMV] coarse-tuning critique of fine-tuning arguments. Suppose some physical constant λ can in principle take on any real number value, but in order for life to occur that constant must have a value in some narrow interval, say $[1.977, 1.979]$. It

Key words and phrases. probability;symmetry;regularity;normalization;fine-tuning;set theory;definability;language.

is intuitive to think that on a single-universe no-designer hypothesis it is unlikely that the constant should have been observed to land in that narrow interval, while it seems rather more likely to do so on a life-friendly hypothesis such as of a large multiverse (taking into account the anthropic principle that only universes with life are observed) or of a benevolent designer. However, MMV cleverly observe that the ratio of the length of a "narrow" interval to the length of the whole real line is zero, and the same is true for *any* finite interval, be it narrow or wide. Moreover, the difference between a narrow and a wide interval is just a matter of affine transformation—sometimes, just a choice of unit system. Thus, MMV insist that if we think that the need for a constant to fall in the "narrow" interval like $[1.977, 1.979]$ for life to be possible is support for some intrinsically life-friendly hypothesis, we should also think that the need for a constant to fall in some "wide" but still finite interval like $[-10^{10^{100}}, 10^{10^{100}}]$ provides equal support for that hypothesis. In short, if fine-tuning is evidence for a life-friendly hypothesis, so is coarse-tuning, contrary to intuition.

While fine-tuning arguments are the *locus classicus* of the normalization problem, the problem can come up in other contexts of scientific reasoning. If a scientific theory predicts that some value λ is in a narrow range, and we observe it to be in that narrow range, then that can be significant evidence for the theory; on the other hand, if it predicts λ to be in a much a wider range, an observation of λ in that range would not be impressive.[1] And yet relative to the real line $(-\infty, \infty)$, all finite ranges a are infinitely narrow.

There is a natural connection between the regularity and normalization problems: both seem to deal with zero probability events. The MMV critique suggests that the prior probability of a constant lying in a finite interval of length L of real numbers should be $L/\infty = 0$, and hence in fine- (and coarse-) tuning reasoning we seem to be conditionalizing on zero probability events. Because of this, both problems have been approached via non-classical probabilities, such as hyperreal-valued ones (e.g., Benci, Horsten and Wenmackers 2018; Pruss 2005). Unfortunately, non-classical probabilities arguably suffer from at least three technical problems: non-conglomerability with multiple resulting paradoxes (e.g., Pruss 2012), loss of symmetries (e.g., Parker 2019; for a precise characterization of what symmetries are lost, see Pruss 2020) and underdetermination (Hájek 2003; Pruss 2018).

I shall argue, however, that there are unified in-principle solutions to both problems within the scope of classical countably-additive real-valued probability theory. This solution involves a defensible weakening of the regularity constraint.

At the same time, the in-principle solutions have a serious practical limitation: I do not know how to actually do calculations with them. Thus we should look at them as a model of how one *might* reason, rather than as a

[1] I am grateful to Robin Collins for this point.

model of how we in fact *do* reason. But at least we will know that regularity and normalization are not insurmountable obstacles.

The focus of this paper will be on assigning a probability to a single real-valued physical parameter. We start with a concrete example. This concrete example concerns an important episode from the history of physics, and in this respect should be more worrying to a Bayesian than the infinite sequences of coin tosses (Williamson 2007), infinitely precise darts thrown at targets, or uniform spinners with precisely defined outputs (e.g., Pruss 2013 and Parker 2019) that are staples of the Bayesian regularity literature but yet do not occur in real scientific practice except as idealized models. The approach of this paper is not intended to apply to such idealized cases.

2. The exponent in the law of gravitation and a restricted regularity

Newton's law L of gravitation states that there is a constant $G > 0$ such that masses m_1 and m_2 exert an attractive force on each other equal in magnitude to Gm_1m_2/r^2. I take it that during the first half of the 19th century, after Cavendish's laboratory measurement of gravitational attraction and before Le Verrier's observation of the precession of the perihelion of Mercury (and probably for a significant period of time prior and after that half century as well) it was reasonable for scientists to assign a credence greater than 1/2 to Newton's law of gravitation, and they typically did so. (My argument does not require a lower credence bound as big as 1/2: *any* non-zero credence will do, though for vividness I will use 1/2.)

But now consider the regularity problem. The primary evidence gathered for Newtonian gravitation from Kepler to Cavendish consisted of claims that had non-zero prior probability. The main astronomical evidence, for instance, involved careful measurements of the positions of solar system objects in the imaginary celestial sphere, all of which measurements had non-zero error bars. Prior to these careful measurements and with Newtonian theory bracketed, there was a non-zero probability that the measurements would fall into these ranges. There is no normalization problem here because the positions in the celestial sphere are given by angles—the right ascension and declination—which have finite ranges (from 0 to 24 hours in right ascension and $-180°$ to $180°$ in declination), and the error bars occupy a non-zero fraction of these ranges. And other evidence—such as perhaps that given by differences in brightness or Cavendish's observations of the movement of lead spheres—also involved non-zero error bars over a finite range of observable effects (e.g., movements of the lead spheres beyond a certain range would have destroyed the apparatus).

So, if the prior probability of L were zero, the posterior would have been zero as well. But there is reason to think the prior probability was zero. After all, for every real number p, there is a law L_λ that says that there is a constant $G > 0$ such that masses m_1 and m_2 exert an attractive force force

on each other equal in magnitude to Gm_1m_2/r^λ. Our Newtonian law L is just L_2 then. All of these uncountably many laws are incompatible (at least given the accepted background fact that there are at least two objects of non-zero mass). Thus, they cannot all have non-zero prior probability. But if we assign prior probability zero to all the laws L_λ, the posterior probability of L_2 on the evidence available in the first half of the nineteenth century will still be zero, and so the scientists assigning credence greater than $1/2$ were unreasonable—which is implausible.

The non-classical probability solutions to the regularity problem leave this problem unsolved. For instance, the hyperreal approach might allow us to assign an infinitesimal prior probability to L_2. But the evidence for Newtonian gravity consisted of claims whose probability was non-zero and non-infinitesimal: the error bars on right ascension and declination measurements were never infinitesimal. And when we conditionalize something with infinitesimal priors on an event with non-zero non-infinitesimal probability, we get something an infinitesimal posterior.[2] And so the hyperreal approach would at best allow the scientists to have infinitesimal credence, and not the credence greater than $1/2$ that they actually had. Similar points apply to other non-classical approaches, as long as they stay faithful to something like Bayesian conditionalization.

But there is in fact a simple and natural solution. While it is impossible that all the L_λ have non-zero real-valued priors, there is nothing mathematically barring the way for L_2 to have a non-zero real-valued prior. Granted, if we let L^* be the disjunction of all the L_λ, i.e., the law that there is some $G > 0$ and some real p such that the force is Gm_1m_2/r^λ, and if we insist that all possible values of p are equally likely given L^*, then by finite additivity we either have to assign zero probability to all the L_λ or to none of them. But in fact it is unlikely that any reasonable scientist had a uniform distribution of the exponent p conditionally on L^*. After all, we discern reasonable people's priors from their posteriors and their evidence, and these simply do not indicate a uniform distribution of p, given that clearly reasonable people did in fact have credence greater than $1/2$ that $p = 2$.

It would be *ad hoc*, of course, to say that $p = 2$ is the only exponent value that has non-zero conditional probability given L^*. We could, indeed, imagine the following scenario. In 1859, Le Verrier discovers that the motion of Mercury departs from Newton's law L_2 but in fact fits very nicely with the law $L_{2+1/2^{32}}$. He hypothesizes that the exponent p is in fact exactly equal to $2 + 1/2^{32}$. Increasingly precise measurements of the motions of solar and extra-solar objects fit with this hypothesis. Einstein comes up with the General Theory of Relativity as a logical curiosity, but it simply does not fit the data as well as $L_{2+1/2^{32}}$. In this scenario, it would be reasonable for

[2]Suppose $P(H)$ is an infinitesimal $\alpha > 0$ and $P(E) \geq r$ for some strictly positive real number r. Then $P(H \mid E) = P(H \cap E)/P(E) \leq \alpha \cdot (1/r)$, which is infinitesimal since an infinitesimal multiplied by any positive real yields an infinitesimal.

scientists to assign a credence greater than $1/2$ to $L_{2+1/2^{32}}$, and hence their priors for $L_{2+1/2^{32}}$ had to have been non-zero (and non-infinitesimal).

It may seem that the same thought experiment could be run for every value of p, leading to the conclusion that all the uncountably many possible exponent values had non-zero non-infinitesimal priors, which is impossible for a coherent probability distribution. But in fact the argument cannot be run for every value of p. It can only be run for every value of p that satisfies the condition that scientists could have hypothesized that value as the exponent. But not every value can be so hypothesized. Indeed it seems right to say that the set of numbers science is capable of thinking is at most countable. After all, all our mathematical expressions are finite in length and written in a finite alphabet, and there are at most countably many such expressions, and hence at most countably many "expressible" mathematical objects. One can then restrict regularity to hypotheses positing such objects. (There are some still serious technical problems with this proposal as it stands which will be discussed below.)

There is in principle no problem with assigning a non-zero classical real-valued countably-additive probability to every member of a countable set E of real numbers. For instance, we can enumerate the numbers as $\lambda_1, \lambda_2, \ldots$ and then assign probability 2^{-n} to λ_n. That's not quite what we want to do in the case of the exponent in the law of gravitation, however. For there is no countable set E of real numbers with the property that it is rational to be certain that if L^* holds, then the exponent is in E. Even if E is all the real numbers we can express, there should be a serious possibility that the exponent is something that we *cannot* express.

But what we can do is to fix a real $0 < \rho < 1$, and assign probability $2^{-n}\rho$ to the exponent being λ_n (for each n) and probability $1 - \rho$ to the exponent being in the complement of E. Moreover, if we so desire, we could even suppose some continuous distribution conditional on the complement of E. For instance, if ϕ is some probability density (say, a normal one), we could suppose that

$$P(p \in [a, b] - E) = (1 - \rho) \int_a^b \phi(x)\, dx.$$

Of course, there is still the problem that the choice of the credence to be assigned to each $\lambda_n \in E$ (conditionally on the exponent being in E) as well as the choices of the overall credence ρ of E and of the density ϕ all look objectionably arbitrary. More will be said on this in subsequent sections. However, for now it is worth noting that in worrying about this arbitrariness we are no longer in the territory of the regularity problem, but in the territory of the general problem of priors for Bayesianism. The subjective Bayesian can simply say that we have the priors we do, and that reverse engineering of scientists' priors yields something like the above model. The objective Bayesian can say that there are objective norms that constrain or even pick out specific choices of answers to these problems,

but that the above argument shows that there is no evident impossibility in satisfying such norms.

It thus seems, though we have yet to discuss some serious technical difficulties), we have a simple solution to the regularity problem for Bayesianism: only countably many hypotheses can in fact be expressed, and there is no mathematical impossibility in a non-zero classical probability to countably many hypotheses. Perhaps it would be nice if we could provide an algorithm to assign non-zero probabilities to hypotheses that we cannot in fact think, but our inability to do so does no harm to our epistemic lives.

One might object that the above example is artificial. We can reformulate Newton's Law of Gravitation in an intuitive way as $F = G'm_1m_2/S$, where S is the surface area of the sphere centered on one of the masses and reaching the other mass and G' is a redefined gravitational constant ($G' = 4\pi G$). This reformulation fits with the intuition about the gravitational influence getting uniformly spread out over space, and in the reformulated law there is no exponent given. However, of course, the exponent is still there. The new law is equivalent to $F = G'm_1m_2/S^1$. It is true that 1 is an intuitively very special exponent. But so was 2. And seeing things this way shows that in fact there are more constants in the Law of Gravitation than met the eye. For instance, we can formulate the original law as $F = G(m_1)^1(m_2)^1/r^2$. This shows that any mathematical law can be seen as part of a continuum of laws, and the problem of hypotheses claiming intuitively privileged values for constants needing to have non-zero non-infinitesimal probabilities is not an isolated one.

For another historical episode, there is a long history of hypotheses about the exact value of the fine-structure constant α, with it being currently believed that α is approximately 1/137.035999206 with an uncertainty of ± 11 in the last two digits of the denominator (Morel *et al.* 2020). There has been a long history of speculation about possible exact values of this constant (Kragh 2003). Eddington had a theory suggesting a value of 1/136. Then, for a while, 1/137 was a serious possibility. Thus, in the 1930s, Pauli expressed a belief that a future physics theory will yield "a numerical value of the fine-structure constant $\alpha = e^2/\hbar c = 1/137$" (quoted in Kragh 2003, 410). Other "nice" values have been proposed. And while it appears reasonable to be sceptical of such proposals on the basis of the current understanding of the physical meaning of the fine-structure constant (Carroll 2018), to assign zero or even infinitesimal credence to the 1/137 hypothesis in the 1930s would have been unreasonable. We could imagine a sequence of more and more precise measurements leading to our best value being 1/137.00000004 with an uncertainty of ± 8 in the last digit of the denominator, in which case it would be very reasonable to assign a not insignificant credence, perhaps even higher than 1/2, to the hypothesis that the exact value is 1/137.

Observe also that while our focus in this paper is on the credence assignment for a single real-valued parameter, the basic point generalizes to

arbitrary families of hypotheses: only countably many of the arbitrary family of hypotheses are actually expressible, and then we can have a catch-all hypothesis that some non-expressible hypothesis is correct.

3. Normalization

The above model in principle offers a solution to the normalization problem. The probability distribution for some real-valued constant λ is given as follows. We have a countably infinite set E of expressible values. There is some probability $0 < \rho < 1$ that λ lies in $E = \{\lambda_1, \lambda_2, \dots\}$, and for each $\lambda_n \in V$ there is a non-zero probability ρp_n that λ equals λ_n. And there is a continuous probability density function ϕ such that:

$$P(\lambda \in U - V) = (1 - \rho) \int_U \phi(x)\, dx$$

for every measurable subset U of $\mathbb{R}$. It is reasonable to suppose, further, that $\phi(x) > 0$ for almost all x (i.e., everywhere outside of some set of Lebesgue measure at most zero).

In general, then, we can say:

$$P(\lambda \in [a, b]) = \rho \sum_{n \in \{n : \lambda_n \in [a,b]\}} p_n + (1 - \rho) \int_a^b \phi(x)\, dx.$$

As long as $a < b$, $P(\lambda \in [a, b])$ will be bigger than zero, for two reasons. First, every rational number is expressible, and hence in E, and if $a < b$ are real, there are infinitely many rational numbers in the interval $[a, b]$. Second, if ϕ is almost everywhere positive, then $\int_a^b \phi(x)\, dx > 0$ for $a < b$. Similarly, if $[a, b]$ is a subset of a strictly larger interval $[c, d]$, then $P(\lambda \in [a, b]) < P(\lambda \in [c, d])$. Thus we do not have the MMV coarse-tuning problem that all finite intervals are equally unlikely.

Moreover, our probability distribution is a classical countably additive one. This implies that

$$\lim_{b \to \infty} P(\lambda \in [-b, b]) = 1$$

and

$$\lim_{\varepsilon \to 0+} P(\lambda \in [-b - \varepsilon, b + \varepsilon]) = P(\lambda = b).$$

If $b \in E$, then $P(\lambda = b) = \rho p_n$ for some n and if $b \notin E$, then $P(\lambda = b) = 0$. We would expect each ρp_n to be relatively small. Going back to the gravitational example, while there is something particularly elegant about the exponent 2 in Newton's Law, and it fits neatly with the idea of forces at a distance r being inversely proportional to the area of a sphere of radius r, this still is just one of infinitely many options, and so we should expect $P(L_2 \mid L^*) \ll 1$. Thus, in either case, the limit as we shrink the interval around b is much less than the limit as we grow the interval around b. Hence, not only is it false that small intervals are as likely as large ones, sufficiently small intervals are *much* smaller than sufficiently large ones.

McGrew, McGrew and Vestrup (2001) in their normalizability objection to fine tuning assumed that the probabilities for constants were governed by the Principle of Indifference. The gravitational exponent and fine-structure constant examples showed that this assumption is false of the epistemic priors of working scientists. Without this assumption, the way is clear to a solution.

4. A somewhat more elegant model

However, the above model suffers from a surfeit of arbitrariness: we need an assignment of non-zero probabilities to infinitely many expressible physical parameters, we need a value ρ to control how likely it is that the parameter falls in the set of expressibles, and finally we need a probability density function ϕ.

Here is a different approach, which is mathematically slightly more elegant. Let $\mathcal{F}$ be all Borel subsets of real numbers. This is the smallest collection of subsets of the reals that contains all intervals and is closed under complements and countable unions. Let $\mathcal{B}$ be the set of all probability measures on $\mathcal{F}$—i.e., the set of all Borel measures. Let $E_{\mathcal{B}}$ be the set of all expressible members of $\mathcal{B}$. This will be a countable set: $E_{\mathcal{B}} = \{\mu_1, \mu_2, \dots\}$. For each μ_n let there be a strictly positive real parameter q_n and suppose that $\sum_{n=1}^{\infty} q_n = 1$. Then we can make our credence that our parameter λ lies in a set A be:

$$P(\lambda \in A) = \sum_{n=1}^{\infty} q_n \mu_n(A).$$

This defines a countably-additive probability function.

This model has the property that for any possible expressible value λ_n for λ, the probability that $\lambda = \lambda_n$ is a positive real number. For if λ_n is expressible, then the measure that assigns probability 1 to the singleton $\{\lambda_n\}$ will also be expressible, and hence equal to some measure μ_{m_n} in $\mathcal{M}$. And then:

$$P(\lambda = \lambda_n) \geq q_{m_n} \mu_{m_n}(\{\lambda_n\}) = q_{m_n} > 0.$$

What makes this model more elegant is that it does not separate out hypotheses positing specific values λ_n from the hypothesis of a distribution over the other values. Instead of having three families of parameters, namely the values p_m, the single value ρ and the values of the density function $\phi(x)$ for $x \in \mathbb{R}$, there is now a single unified set of parameters in the model, the values q_m.

Moreover, we can give a somewhat natural picture of how to generate this model. We consider an infinite number of hypotheses about possible stochastic processes that could generate the constant λ, where these hypotheses correspond to different measures μ in $E_{\mathcal{B}}$. Each such hypothesis then gets a non-zero credence. Among the stochastic processes are trivial processes that select a single expressible value with probability 1.

5. Two limitations

There are, however, two technical limitations that both the original and the revised model face. A natural weakening of the regularity condition is to require that all *expressible* coherent hypotheses get non-zero credence. The two models above are guaranteed to satisfy this condition in the special case of the hypotheses H_{λ_n} that say that λ is precisely equal to λ_n for an expressible λ_n.

But our models fail to assign any credence to hypotheses according to which our physical constant lies in a set that is not Borel-measurable, and in particular do not assign a non-zero credence to such hypotheses. The problem of non-measurable sets is often considered to be a different problem from that of regularity, however, and hence it will be beyond the scope of this paper.[3]

And there is a second potential problem. Imagine that we have a Borel-measurable set U that is itself expressible but has no expressible members and let H be the hypothesis that our physical constant is in A. If U contained some expressible value λ_n, then our models would have assigned a non-zero credence to H, since they assign a non-zero credence to the hypothesis H_{λ_n} that entails H. So, the difficulty lies in the case where U is an expressible set but does not contain any expressible values. Such a set will have to be infinite, since if U were finite then we could precisely express one of its values, say using the description "The smallest value in U." Further, since every algebraic number is expressible, such a problematic set U cannot contain any algebraic numbers, and hence it cannot contain any interval of positive length. If expressibility is understood in the technical sense of set-theoretic definability (see Section 7, below), and ZF is consistent, then it is consistent with ZFC that there is such a set and it is also consistent that there isn't (Kanovei and Lyubetski 2017; Dorais and Hamkins 2019).

I acknowledge these limitations of the proposed model, but claim they are limitations that do not affect scientific practice. In practice, science tries to confirm two kinds of specific hypotheses about the values of constants: hypotheses about them being equal to a specific expressible number, say 2 or 1/137, and hypotheses about them lying in some interval of numbers of positive length, say $[1/137.035999217, 1/137.035999195]$. While no doubt some physicists have idly speculate about whether the exact value of a constant falls into some concrete weird set, such speculation does not seem to actually appear in real scientific research.

[3]Note, however, that often when one talks of non-measurable sets, one talks of Lebesgue non-measurable sets. I have chosen to work with measures on the σ-algebra of Borel sets rather than the larger σ-algebra Lebesgue sets in order to allow for the possibility of more measures in our set $E_{\mathcal{B}}$. The trade-off here is that these measures will measure fewer sets. One can also go for the other trade-off and replace the Borel sets with the Lebesgue measurable sets throughout the paper. It is a question for further investigation which approach is superior.

It would be nice to show there is a prior probability assignment that assigns non-zero credence to every expressible hypothesis. I do not know of a way of doing so. But what we have here appears to be a close second-best: the existence of a probability assignment that assigns non-zero credence to the kinds of hypotheses about constants that come up in science, and that is at least in principle amenable to the evaluation of fine-tuning arguments.

But we cannot declare victory yet. We need to discuss two further challenges. The first is to come up with a proposal for the weights q_n that is not objectionably *ad hoc* and the second is to deal with thorny logical issues around the concept of expressibility.

6. A random-description proposal for weights

In this section we give a model proposal for the weights p_n assigned to an enumeration of expressible reals or the weights q_n assigned to an enumeration of expressible Borel measures. Readers will note a resemblance between this proposal and Solomonoff priors (e.g., Solomonoff 1978) as well as Kolmogorov measures of complexity (e.g., Kolmogorov 1963). There are multiple ways to vary the details of the proposal, but rather than seeking full generality as to the types of proposal possible, I will pursue one specific example.

We can consider expressible numbers and Borel probability measures to be set-theoretic objects that can be finitely defined in some fixed mathematical language L with a finite set Σ of symbols, where an object a is definable in L just in case there is well-formed formula (wff) ϕ of L with one free variable such that a is the only object that satisfies ϕ. Let F be the countably infinite set of all the finite sequences of symbols in L: every wff will be a member of F, but F will also contain infinitely many non-wffs, and indeed infinitely many completely nonsensical sequences.

There is a natural probability measure P_F on F that models the following stochastic process. Let $\Sigma^* = \Sigma \cup \{\$\}$ where \$ is a symbol not in Σ. Now, uniformly randomly and independently choose a sequence of symbols from Σ^*, stopping when you pick \$, and letting the symbols prior to the \$ define the (possibly empty) member of F. (By the Strong Law of Large Numbers, with probability one you will eventually hit \$.) More explicitly,

$$P_F(\{s_1 s_2 \ldots s_n\}) = \frac{1}{(N+1)^{n+1}}$$

if $s_1 s_2 \ldots s_n$ is a sequence of symbols from Σ and N is the number of elements of Σ.

Let A be either the reals $\mathbb{R}$ or the set $\mathcal{B}$ of Borel probability measures on $\mathbb{R}$. Now, let M_A be the infinite subset of F consisting of those sequences that are grammatically correct formulas with one free variable that have a unique satisfier where that unique satisfier is a member of A. Then a member x of A is expressible in L just in case it satisfies some formula in M_A.

Now suppose $x_1, x_2, ...$ are the expressible members of A. Fix n. Let $d_1, d_2, \ldots$ be all the formulas in M_A that are satisfied by x_n. Then let:

$$r_n^A = \sum_i P_F(\{d_i\} \mid M_A).$$

We can now let $p_n = r_n^{\mathbb{R}}$ and $q_n = r_n^{\mathcal{B}}$.

Another way to describe what we are doing is that we are choosing a random member of A by repeatedly and independently generating random finite sequences of symbols from Σ, stopping as soon as we have generated a grammatically correct wff that has a unique satisfier and where that satisfier is an object in A, and letting the random member of A be that satisfier. The probability of x_n thus being generated is then defined as the weight r_n.

Intuitively, the weights here can be expected to strongly favor those mathematical objects that can be more briefly described, in light of the fact that P_F depends inversely exponentially on the length of the sequence. In the real number case, and using a standard mathematical language, numbers like 2 in the Newtonian example that can be more briefly described can be expect to get higher weight than numbers that are harder to express. In the Borel measure case, every measure that assigns probability one to an expressible singleton has non-zero weight, but when that expressible singleton can be described in a relatively simple way, as $\{2\}$ can, the weight can be expected to be rather higher than the weight of many other singletons. The measure thus is likely to fit with scientific preferences for more elegant theories.

In practice, it is of course very difficult to figure out the values of the weights given the above assignment. And, interestingly, details of our set theoretic constructions of entities such as real numbers will matter.

For instance, there are multiple set-theoretic ways of handling the relationship between the natural numbers and the reals. One common approach is to start with a construction of the natural numbers, often using the recursive definition $0 = \varnothing$ and $sn = n \cup \{n\}$, where s is the successor function, so that $1 = \{\varnothing\}$, $2 = \{\varnothing, \{\varnothing\}\}$, and so on. Then one constructs negative integers by using a set-theoretic way of marking a sign (e.g., if n is a positive integer, we can let $-n$ be the ordered pair $(\varnothing, n)$), rational numbers as equivalence classes of ordered pairs of integers with the second one non-zero (under the relation $(a, b) \sim (c, d)$ if and only if $ad = bc$), and real numbers as equivalence classes of Cauchy sequences of rational numbers. If one does this, then any description of any real number, however simple, will have to reprise an enormous amount of set theoretic detail. Even the real number zero will be an equivalence class of sequences of equivalence classes of pairs of objects.

The unconditional probability that a randomly chosen member of F will describe the real number zero will then be extremely small. (We can get a lower bound on that probability by writing out the details in the definitions, but that is a difficult task.) However, when we conditionalize on $M_{\mathbb{R}}$—i.e.,

conditionalize on the random description being the random description of *some* real number—the probability will be much higher, and the weight $r_n^{\mathbb{R}}$ corresponding to the real number zero can intuitively be expected to be non-negligible, because zero is significantly easier to describe than most other numbers.

Another approach, however, is to modify the above constructions. Thus, instead of making every rational number an equivalence class of ordered pairs of integers, in the case of a number whose equivalence class includes $(n, 1)$ for some integer n, we replace the equivalence class by n itself, thereby making it be the case that some rational numbers literally are integers. And instead of making every real number be an equivalence class of Cauchy sequences of rationals, we replace those equivalence classes of Cauchy sequences of rationals that include a constant rational number sequence (e.g., $3/4, 3/4, 3/4, ...$) with the rational number in that constant sequence. If we proceed in this way, then among the reals, the integers and rationals will have relatively simple set-theoretic descriptions, and it will be much easier to obtain lower bounds for the unconditional probability that a member of F describes zero or two. For instance, zero will be $\varnothing$ and hence will be described by the formula

$$\forall x' \sim(x' \in x)$$

and two will be $\{\varnothing, \{\varnothing\}\}$ which can be described by the formula

$$\forall x'(x' \in x \leftrightarrow (\forall x'' \sim(x'' \in x') \vee \forall x''(x'' \in x' \leftrightarrow \forall x''' \sim(x''' \in x'')))).$$

Thus, the unconditional probability of picking a sequence of symbols describing zero is at least $(N + 1)^{-11}$ and of picking a sequence describing two is at least $(N + 1)^{-56}$ (there are 10 and 55 tokens in the two displayed expressions above, respectively), where N is the number of symbols in the language. If our language consists of two quantifiers, five truth-functional operators, one variable, one prime and two parentheses, then we can take $N = 11$, and our lower bounds will respectively be the small and very small numbers 12^{-11} and 12^{-56}. These are also lower bounds for the probability conditional on $M_{\mathbb{R}}$, but it is intuitively likely that after conditioning on $M_{\mathbb{R}}$, the probability will be much, much higher—and indeed it needs to be for our physics applications.

In practice, for further computations it would make sense to switch to a language where grouping is handled by Polish notation rather than parentheses, which would reduce the length of the two sample expressions by 2 and 10, respectively. Alternately, we might choose to work with some language further from set theory and closer to the mathematical languages that scientists actually work with, languages with abbreviations for many quantities commonly occurring in our mathematics, such as "2" and "π", as well for mathematically natural operations, such as multiplication, integration, and so on. The alphabet will be larger, but we will have much shorter expressions. For instance, we could then express 2 very simply with the formula $x = 2$.

It could even make some sense to combine the idea behind Solomonoff priors with the above approach. Thus, we could replace the measure P_F on strings of symbols from L by a measure induced by randomly generating computer programs in some fixed computer language, which programs generate strings of symbols from L. Thus, the probability of a number or Borel probability measure would then be the probability of a random computer program producing a description of that number or measure conditionally on the program producing a description of some number or measure.

7. Technical problems with expressibility

There is, however, another serious technical problem with the seemingly plausible proposals so far. One way to see the problem is to note that the argument that there are only countably many expressible reals—an argument known as the "Math Tea argument" in mathematical folklore—has serious gaps. To spell out the faulty but suggestive argument in greater precision, we might proceed as follows. Let expressibility be set-theoretic definability. A set[4] a is *definable* provided that there is a wff ϕ in the language of set theory (a first order language with one predicate $\in$) with one free variable x such that a satisfies ϕ when substituted for x, and nothing else satisfies ϕ when substituted for x.

Let W be the set of wffs in the language of set theory with at most one free variable expressed in some coding scheme like Gödel numbering. Let $M_{\mathbb{R}}$ be the subset of W corresponding to wffs satisfied by exactly one object and where that object is a real number. Let $\lambda_{\mathbb{R}} : M_{\mathbb{R}} \to \mathbb{R}$ be the function defined by letting $\lambda_{\mathbb{R}}(n)$ be the real number that satisfies the formula with code n. Then W is a countably infinite set, and $M_{\mathbb{R}}$ is a subset of it, so it is also countable. The expressible reals E are then the range of $\lambda_{\mathbb{R}}$, and the range of a function with a countable domain is countable.[5]

This argument as it stands commits two related technical errors. First, there is no reason given to think there exists a *set* of the codes corresponding to the wffs satisfied by exactly one object and where that object is a real number. If set theory is consistent, then by Tarski's Indefinability of Truth, the concept of truth cannot be defined in the language of set theory. And since truth can be defined in terms of satisfaction, neither can satisfaction be defined.[6] And without defining satisfaction in set theoretic language, we have no way to guarantee the existence of the set $M_{\mathbb{R}}$. Similarly, and actually

[4]In set theory, real numbers, probability measures and other such objects are all sets.

[5]If $f : A \to B$ with A countable, then A bijects with a subset of the natural numbers, so there is a well-ordering $\prec$ on A derived from arithmetic ordering on the naturals. For y in the range of f, let $g(y)$ be the $\prec$-least x such that $f(x) = y$. Then g is a one-to-one function from B to a countable set, so B is countable.

[6]Normally, we would define a wff as true provided it has no free variables and is satisfied. Above, however, we are only interested in the satisfaction of formulas with a free variable. But the satisfaction of formulas with a free variable can also define truth: a wff ϕ is true provided that it has no free variables and something satisfies $\phi \,\&(x = x)$.

more seriously, even if $M_{\mathbb{R}}$ were to exist, the existence of the function $\lambda_{\mathbb{R}}$ has also not been shown, since we "defined" that function using satisfaction.

A similar problem afflicts the alleged set of expressible Borel probability measures. Moreover, the proposal in Section 6 for the weights that real numbers and Borel probability measures are to be assigned requires a function like to $\lambda_{\mathbb{R}}$ between codes for wffs with unique satisfiers and these satisfiers.

Admittedly there is more than a whiff of paradox about the above critiques. For instance, there is no difficulty in defining the set W in Zermelo-Fraenkel (ZF) set theory for an appropriate coding scheme, like Gödel numbersing: wffs with a free variable are definable purely syntactically. And assuming realism about mathematics, it is then true that some but not all members of W are codes of wffs satisfied by exactly one object and where that object is a real number. Surely, then, there should be a subset of W consisting of precisely such codes. Admittedly, we can concede that not every collection of objects satisfying some predicate is a set—otherwise we fall into the Russell Paradox—but the idea that a "small" infinite set, like the countably infinite set W, should be "missing" a precisely described subcollection like $M_{\mathbb{R}}$ seems very strange.

Nonetheless, the critique is logically sound: no proof has been given for the existence of the mathematical objects involved in the constructions. Moreover, if ZF is consistent, then ZFC (ZF plus Choice) has a model $\mathfrak{M}$ in which *every* object is the unique satisfier of some wff with one free variable (Hamkins, Linetsky, and Reitz 2013) In particular, in $\mathfrak{M}$ the set E of expressible real numbers is *uncountable*, since it is equal to $\mathbb{R}$. Again, this seems impossible: Isn't E in one-to-one correspondence with a subset of the wffs, and isn't the set of all wffs countable? But there is no contradiction here. For the alleged one-to-one correspondence between E and a subset of the wffs is simply missing from the model $\mathfrak{M}$, whether or not it is a member of the true or intended model of set theory.

Here is one way to work around the above issue. Start with ZFC (one can also work with ZF if one prefers). For every wff ψ with exactly one free variable, add to ZFC the symbol S^ψ. Then add the following axiom schema:

$$\exists! x \psi(x) \to \forall x[\psi(x) \to \forall y[(\ulcorner \phi \urcorner, y) \in S^\psi \leftrightarrow (y \in x \,\&\, \phi(y))]],$$

where ϕ and ψ have exactly one free variable, $\phi(x)$ and $\psi(y)$ are respectively what one gets upon substituting x and y for the free variables, and $\ulcorner \phi \urcorner$ is the code for the wff ϕ. This says that if ψ uniquely describes some set x and ϕ is some formula, then S^ψ is a subset of x such that if m is the code of a formula with exactly one free variable and $y \in x$, then (m, y) is a member of S^ψ if and only if that formula is satisfied by y. In other words, S^ψ is a set that can define a satisfaction predicate for members of x. Let ZFC^S be the resulting theory. An easy compactness argument shows that ZFC^S is consistent if and only if ZFC is consistent. And once we have the symbol S^ψ, we can rigorously define our probability measures. See the Appendix for details.

So, as we saw, confining ourselves to ZFC, it seems we cannot decide whether our constructions make sense or not. However, it is more intuitive to think that they do: that we are not "missing" subcollections like $M_{\mathbb{R}}$ or mappings like $\lambda_{\mathbb{R}}$. We can just add the needed assumptions as additional axiom schemas for our mathematical theories. At this point we have two choices. First, we can say that these schemas are simply true. After all, it is intuitively plausible that at least for any set y definable by a formula ψ there is a satisfaction set S^{ψ}. Second, we can take a version of the Hamkins' set-theoretic multiverse approach (Hamkins 2012) and say that there are many universes of sets, and we will simply choose to do our mathematics in one in which the assumptions are true.

Either approach will let us define a probability distribution over those members of A (reals or Borel probability measures) that can be defined in the original language of ZFC (i.e., without the added symbols S^{ψ}). Now, in fact, our physics is first-order in the sense that its mathematical formulas make no reference to satisfaction or truth, but can all be expressed in ZFC. Thus, if all we want is regularity restricted to the kinds of hypotheses that come up in a physics like ours, this approach will do the job. But if we want, we can expand further, repeating the above construction to include objects definable in terms of ZFC^S, by using a further extended theory $(\text{ZFC}^S)^S$—we leave the logical details to the reader. This will allow us to extend our restricted regularity further.

And we can iterate this construction several more times if we so desire. Granted, some things that are intuitively expressible in ordinary English will always escape us. For English, unlike ZFC or any of our iterated extensions of it, contains its own satisfaction predicate. However, that satisfaction predicate is also a source of paradox: it can be used to express liar sentences. And we do not appear to lose anything of scientific interest by restricting our attention to what is definable in ZFC or at most in some finite iteration of our construction.

If we do not like the idea of extending set theory, there is a different solution to the technical problems: we contract our attention to a fixed level in the Lévy hierarchy. Say that the quantifiers $\forall x \in y(\cdots)$ and $\exists x \in y(\cdots)$, which are short for $\forall x(x \in y \rightarrow \cdots)$ and $\exists x(x \in y \,\&\, \cdots)$, are *restricted quantifiers*, and that a wff of set theory is *restricted* provided all its quantifiers are restricted. Let Σ_0 and Π_0 be the collection of all restricted wffs of set theory. Inductively, let Σ_{n+1} be the set of wffs of the form $\exists x\rho$ for a variable x and a wff ρ in Π_n and let Π_{n+1} be the set of wffs of the form $\forall x\rho$ for a variable x and a wff ρ in Σ_n. A Σ_n (respectively, Π_n) wff is then a sequence of n unrestricted quantifiers, the left-most being existential (universal), the next one being universal (existential), and continuing to alternate in type, all followed by a restricted wff.

Every wff of set theory can be proved to be equivalent to a Σ_n or Π_n wff for some finite n. Moreover, for an appropriate coding scheme, and for a fixed n, the satisfaction of a $\Sigma_n \cup \Pi_n$ wff, appropriately encoded,

can be defined in the language of set theory; indeed, by a $\Sigma_{n+1} \cup \Pi_{n+1}$ wff (Lévy 1965). In practice, all mathematical formulas that come up in physics are equivalent to a $\Sigma_n \cup \Pi_n$ wff for some moderate finite value of n. If we are willing to restrict our regularity assumption to n-expressible hypotheses, i.e., hypotheses expressible by wffs from $\Sigma_n \cup \Pi_n$, for some fixed finite n (say, $n = 1000$ to be really conservative), then we will still be able to deal with a countable infinity of n-expressible constants or Borel probability distributions, including all the ones expressible in a physics remotely like ours, but our set theoretic difficulties will disappear (for some more detail, see the Appendix).

The above solutions assume that expressibility is definability in the language of set theory. But one may wish to have a potentially broader notion of expressibility. One can specify numbers such as 5, $\sqrt{2}$ and $-(\log \pi)^e$ in the language of set theory. But one can also specify physical quantities in terms of other physical quantities. Thus, one might have hypothesized that the exponent in the law of gravitation is $p = 274\alpha$, where α is the fine-structure constant. And it might have been the case that α is not one of the set theoretically definable numbers, in which case p wouldn't have been either on this hypothesis. However, we can *think* about the number 274α because we can think about the number α, since we can gain reference to α through its physical significance.

Because of this, it might help to extend the above approaches to allow for what set theorists call definability with parameters. If B is a set, then we say can that a is *definable from the parameters in* B just in case there is a formula ϕ of set theory with free variables $x, y_1, ..., y_n$ such that there are elements $b_1, ..., b_n$ of B and a is the unique set that satisfies $\phi(y_1/b_1) \ldots (y_n/b_n)(x/a)$, where $\psi(z/c)$ says that c satisfies ψ when substituted for z. In other words, we are allowed to help ourselves to the objects from B when defining a, even if these objects themselves cannot be defined using pure set theory. In our example above, $p = 274\alpha$ is definable from the parameter in $\{\alpha\}$, or, more briefly, definable by α. The parameters, furthermore, need not be fundamental constants. For instance, in classical mechanics, if we could precisely specify which particles are constituent of two macroscopic objects, we could thereby gain reference to the distance (in some unit system) between the centers of mass of the two objects, and then could formulate hypotheses about the values of fundamental constants as expressed in terms of parameters that include that distance.

There is some finite collection B of mathematical objects such as real numbers, but perhaps not only real numbers[7], that describe features we have identified in the real world and to which we have gained reference by

[7]For instance, if quantum mechanics is the correct theory of the world, then we have gained reference to some cardinal κ as the cardinality of the Hilbert space of the global wavefunction. That cardinality is often thought to be $\aleph_0$, which is definable in set theory without parameters, but it might turn out to be some higher cardinal, in which case our expressive resources might increase when we use κ.

methods such as those suggested in the previous paragraph. It is natural to suggest that quantities definable with parameters from B should also count as expressible. And in the literal English sense of "expressible", they are expressible. However, for the purposes of the construction of our probability measures, such inclusion is not a good idea. For suppose that we are considering hypotheses about the value of α. But among the mathematical objects that our scientific practices have given us reference to is α itself: $\alpha \in B$. So it's trivial that α is definable from parameters in B. Moreover, while we can extend the proposal about weights described in Section 6 to consider wffs with parameters from B, the result is apt to be inscrutable, since the weights assigned to different reals or Borel measures will depend on how these reals or measures can be defined in terms of the parameters from B, the exact values of which parameters we do not know.

A better approach, then, is probably to restrict expressibility in the construction of the probability measures to the *a priori* expressibility given by set theoretic definability without parameters. And if we wish to consider hypotheses about the relationships between constants, we can do that in other ways. We can, for instance, define a secondary constant $\gamma = p/\alpha$, and consider hypotheses about the values of that constant, including the hypothesis that $\gamma = 274$.

8. Uniformity and independence

As noted, much of the Bayesian regularity literature deals with cases such as infinite sequences of coin flips, spinners and dart throws. In these cases there is a very natural classical probability distribution that we can think of either as coming from an independence assumption, as in coin flip cases, or from a Principle of Indifference in continuous cases. This classical probability measure that assigns zero probability to various expressible outcomes, such as all the coins landing heads or the spinner stopping at zero degrees, and hence the present approach will not work to help save regularity as long as that probability measure is insisted on.

However, two things can be said about these cases. First, these cases are idealizations that simply do not appear to come up in the course of serious scientific theorizing. Moreover, the zero probability outcomes are not something we actually conditionalize on, because they are not something we are capable of observing. We do not have any facility for observing infinite sequences of coin flips or for measuring precise outcomes of continuous processes. The outcomes we actually observe—such as a long finite sequence of coin flips or an outcome of a continuous process with non-zero error bars—have non-zero classical probabilities. So as long as our interest is in saving scientific practice, we need not worry about these kinds of cases.

Second, if we think of the probabilities in theses cases as genuine epistemic probabilities, it is plausible that we really should not have the epistemic probability distributions that create the regularity problems. Suppose

we come across an actual infinite coin flip sequence. If we have observed a million coin-tosses in the sequence and found that they were all heads, then it seems quite reasonable to suppose that, contrary to what we had previously thought, every result in the sequence is actually heads. This inference can only work if our prior probability for all the tosses in the infinite sequence being heads is strictly bigger than zero, and indeed is not merely infinitesimal. For our evidence, namely a million coin tosses landing heads, is something that has non-zero non-infinitesimal probability, so that if the prior were zero or infinitesimal, the posterior would remain zero or infinitesimal. Perhaps, instead, our priors should have been a probability distribution analogous to the ones in this paper, such as $(1-\rho)P_1+\rho P_2$, where ρ is some small positive real constant, P_1 is the distribution resulting from treating the coin tosses as fair and independent, and P_2 is some distribution that assigns non-zero weight to expressible outcomes, such as all-heads, all-tails, heads-on-even-and-tails-on-odd, and so on.

9. Subjective and objective Bayesianism

On subjective Bayesianism, the above models show that it is *possible* to save enough regularity for typical scientific applications and for fine-tuning arguments. And subjective Bayesians can indeed adopt the proposed random-description weights, though they might instead wish to adopt some other system of weights.

Objective Bayesianism divides into a version on which there is a unique correct set of priors, and a version on which there are objective constraints, beyond consistency and regularity, on the priors, but these constraints do not uniquely determine one correct set of priors. On the unique priors version, we have some difficulties for our solution. First, the approaches where we restrict to some high but ultimately arbitrary level $\Sigma_n \cup \Pi_n$ of the Lévy hierarchy or to a specified number of iterations of the construction that went from ZFC to ZFC^S are going to be too arbitrary. Second, the random-description weights depend on the precise choice of language and notation. A version of set theory with a different set of fundamental predicates than $\in$ or even a notational change such as that between parentheses, dot notation and Polish notation is apt to affect the weights, and so it is very difficult to be confident that one has got the unique right version.

But if the objective Bayesianism simply holds that there are objective constraints on priors, and one is willing to allow for weakened versions of regularity among these constraints, then our proposals may well be compatible with that.

10. Conclusions

If we restrict regularity to hypotheses that are in some precisely specified sense "expressible", and there are only countably many such hypotheses, the mathematical problems associated with the regularity requirement for

Bayesianism shrink considerably. We have offered two proposals for probability measures in the special case of a single real-valued constant, one privileging expressible numbers and the other privileging expressible probability measures. These proposals also offer a novel solution to the coarse-tuning objection to fine-tuning arguments. These proposals do not solve all regularity problems, but they do appear to solve the ones that actually occur in scientific practice.

There are some thorny set-theoretic issues involved. These can be remedied by extensions to the ZFC axioms or by further restrictions of expressibility. The latter approach is much more innocent from the mathematical point of view, and fits with all the physics theories that are being taken seriously, but is also less elegant due to the arbitrary choice of level in the Lévy hierarchy.

Finally, just as in case of Kolmogorov complexity and Solomonoff priors, doing actual probability calculations with random description probabilities is difficult. We may be stuck with intuitive approximations for a while. That shouldn't bother the subjective Bayesian, but the objective Bayesian will presumably be less happy. However, at least we can now have something mathematically coherent that we are approximating with our intuitions. And typical Bayesian applications to the real world are, after all, all handwavy approximations: we do not, after all, have exact priors for all, and perhaps any, of the scientific hypotheses we meet up with. Bayesianism can best be thought of as a *model* of reasoning. But now we have models that incorporate enough regularity to reconstruct episodes from the history of science and in principle distinguish fine- from coarse-tuning. That is progress.

Appendix: Some set-theoretic machinery

Say a formula ϕ in a language L with one free variable defines an object a in a model $\mathfrak{M}$ (which we shall assume includes an interpretation of L) provided that o satisfies ϕ in $\mathfrak{M}$ when substituted for the free variable and nothing other than o does.

Let L be the language of set theory (i.e., a first order language with a single binary predicate $\in$) formulated in such a way that the alphabet is finite (so, to have infinitely many variables, we will have to resort to allow for variables made out of multiple symbols, for instance using primes). Let L^S be the language of L together with a new symbol S^ψ for every formula ψ with exactly one free variable. Work with a coding system $\ulcorner\phi\urcorner$ for a formula ϕ where the codes are natural numbers, and where there are formulas of ZFC that can decide whether a code is the code of a formula that is a wff with exactly one free variable, and whether some other natural number is the length of the formula with the given code. Let ZFC^S be the theory whose axioms are all the axioms of ZFC together with the Satisfaction Schema:

$$\exists! x\psi(x) \rightarrow \forall x[\psi(x) \rightarrow \forall y[(\ulcorner\phi\urcorner, y) \in S^\psi \leftrightarrow (y \in x \,\&\, \phi(y))]],$$

for all formulas ψ and ψ with one free variable, and where $\rho(z)$ is ρ with all instances of its free variable replaced by z.

Suppose ZFC is consistent (if not, all bets are off) and hence by the completeness theorem has a model $\mathfrak{M}$. I claim that ZFC^S is a consistent theory. To see this, let F be a finite subset of the axioms of ZFC^S. Then F consists of a finite number of axioms of ZFC and a finite number of instances of the Satisfaction Schema, $S_1, ..., S_n$, where S_i is the sentence:

$$\exists! x \psi_i(x) \rightarrow \forall x[\psi_i(x) \rightarrow \forall y[(\ulcorner \phi_i \urcorner, y) \in S^{\psi_i} \leftrightarrow (y \in x \,\&\, \phi_i(y))]].$$

Let I be the subset of $\{1, ..., n\}$ consisting of indices i such that $\exists! x \psi_i(x)$ is true in $\mathfrak{M}$. We will let $\mathfrak{M}'$ be a model of L^S that has the same objects and relations as $\mathfrak{M}$, and whose interpretation of L^S agrees with the interpretation of L given by $\mathfrak{M}$, but which additionally interprets the symbols S^ψ as follows. First, if ψ is not one of the ψ_i for $i \in I$, then S^ψ can have any interpretation we like in $\mathfrak{M}$. Now, fix $i \in I$ and suppose $\psi = \psi_i$. Let $j_1, ..., j_m$ be those indices in I such that ψ_{j_k} is the same formula as ψ_i for $1 \leq k \leq m$. Let $n_k = \ulcorner \phi_k \urcorner$. Then because $\mathfrak{M}$ satisfies the ZF axioms, $\mathfrak{M}$ contains a unique object s_i satisfying (when substituted for x) the formula:

$$\forall z \left[z \in x \leftrightarrow \exists n \exists y \left(z = (n, y) \,\&\, \bigvee_{k=1}^{m} (n = n_{j_k} \,\&\, \phi_{j_k}(y)) \right) \right].$$

Now interpret S^{ψ_i} as s_i in $\mathfrak{M}'$. The instances of the Satisfaction Schema with indices $j_1, ..., j_m$ will be satisfied under this interpretation. Hence, we now have an interpretation of L^S under which all the axioms in F are satisfied. Hence every finite subset of the axioms of ZFC^S is consistent, and so ZFC^S is consistent.

The real numbers $\mathbb{R}$ as well as the set $\mathcal{B}$ of all Borel probability measures on $\mathbb{R}$ can both be defined in ZFC. Let A be either $\mathbb{R}$ or $\mathcal{B}$. Let ψ_A be the defining formula of A. Work in ZFC^S. Let $F(n)$ be a formula of set theory that is true just in case n is the code of a formula which is a wff with exactly one free variable. Let:

$$M_A = \{n \in \mathbb{N} : F(n) \,\&\, \exists! x[(n, x) \in S^{\psi_A}]\}$$

be the set of codes of formulas that have exactly one satisfier and where that satisfier is in A. Let

$$\lambda_A = \{(n, x) \in S^{\psi_A} : n \in M_A\}.$$

Then λ_A is a function from M_A to A such that $\lambda_A(n) = x$ just in case n the code for a formula uniquely satisfied by $x \in A$. As usual, for $U \subseteq A$, let

$$\lambda_A^{-1}[U] = \{n \in M_A : \lambda_A(n) \in U\}$$

be the pre-image of the set U.

We can now rigorously define our random-description measure on A. Let $\ell(n)$ be the length of the formula with code n (we assume the codes are assigned in such a way that the length of a formula can be read off from it

using the tools in ZFC—this is true for Gödel numbering). Let N be the number of symbols in L. Define:

$$\alpha_n = \frac{1}{(N+1)^{\ell(n)+1}}.$$

This is the probability of randomly generating a sequence of symbols with code n. Now, for $U \subseteq A$, we can let

$$P_A(U) = \frac{\sum_{n \in \lambda_A^{-1}[U]} \alpha_n}{\sum_{n \in M_A} \alpha_n}.$$

This is the conditional probability that a randomly chosen sequence of symbols generates a definition satisfied by a member of U conditionally on its generating a definition satisfied by some (necessarily definable) member of A.

We can let E_A be the range of λ_A: this is the set of expressible members of A, and it is a countable set since the domain of λ_A is a subset M_A of the natural numbers. If $x_1, x_2, ...$ is an enumeration of $E_{\mathbb{R}}$, then we can now define the weight p_n of x_n to be $P_{\mathbb{R}}(\{x_n\})$. Similarly, if $\mu_1, \mu_2, ...$ is an enumeration of $E_{\mathcal{B}}$, we can define the weight q_n of μ_n to be $P_{\mathcal{B}}(\{\mu_n\})$.

For any fixed level N of the Lévy hierarchy, and an appropriate coding scheme like Gödel numbering, one can express a satisfaction predicate S_N for that level in ZFC such that $S_N(n, a)$ just in case a satisfies the level N formula with code n. One can, of course, also express a predicate F_N that says that a formula has exactly one free variable and is at level N. If A is $\mathbb{R}$ or $\mathcal{B}$, let

$$M_{A,N} = \{n \in \mathbb{N} : F_N(n) \,\&\, \exists! x S_N(n, x)\}.$$

Define

$$\lambda_{A,N} = \{(n, x) : S_N(n, x) \,\&\, n \in M_{A,N}\}.$$

We can now define our probability measures as above, but using $M_{A,N}$ and $\lambda_{N,A}$ in place of M_A and λ_A.

References

[1] Benci, V., Horsten, L., and Wenmackers, S. 2018. "Infinitesimal probabilities", *British Journal for the Philosophy of Science* 69:509–552.

[2] Borel, E. 1909. *Eléments de la théorie des probabilités*. Paris: Gauthier-Villars.

[3] Carroll, Sean. 2018. "Atiyah and the fine-structure constant", `https://www.preposterousuniverse.com/blog/2018/09/25/atiyah-and-the-fine-structure-constant/`.

[4] Dorais, F. G.; and Hamkins, J. D. 2019. "When does every definable nonempty set have a definable element?," *Mathematical Logic Quarterly* 65:407-411.

[5] Hájek, Alan. 2003. "What conditional probability could not be", *Synthese* 137:273–323.

[6] Hamkins, Joel D. 2012. "The set-theoretic multiverse", *Review of Symbolic Logic* 5:416–449.

[7] Hamkins, J. D.; Linetsky, D.; and Reitz, J. 2013. "Pointwise definable models of set theory", *Journal of Symbolic Logic* 78: 139–156.

[8] Kanovei, V. G. and Lyubetsky, V. A. 2017. "A countable definable set containing no definable elements", *Mathematical Notes* 102:338–349.
[9] Kolmogorov, A. N. 1963. "On tables of random numbers", *Sankhyā Series A* 25:369–376.
[10] Kragh, Helge. 2003. "Magic number: A partial history of the fine-structure constant", *Archive for History of Exact sciences* 57:395–431.
[11] Lévy, Azriel. 1965. *A Hierarchy of Formulas in Set Theory*. Providence, RI: American Mathematical Society.
[12] McGrew, Timothy; McGrew, Lydia; and Vestrup, Eric. 2001. "Probabilities and the fine-tuning argument: A sceptical view", *Mind* 110:1027–1038.
[13] Morel, L.; Yao, Z.; Cladé, P.; and Guellati-Khélifa, S. 2020. "Determination of the fine-structure constant with an accuracy of 81 parts per trillion", *Nature* 588:61–65.
[14] Parker, Matthew W. 2019. "Symmetry arguments against regular probability: A reply to recent objections", *European Journal for Philosophy of Science* 9.
[15] Pruss, Alexander R. 2005. "Fine- and coarse-tuning, normalizability, and probabilistic reasoning", *Philosophia Christi* 7:405–423.
[16] Pruss, Alexander R. 2012. "Infinite lotteries, perfectly thin darts and infinitesimals", *Thought* 1:81–89.
[17] Pruss, Alexander R. 2015. "Popper functions, uniform distributions and infinite sequences of heads", *Journal of Philosophical Logic* 44:259–271.
[18] Pruss, Alexander R. 2018. "Underdetermination of infinitesimal probabilities", *Synthese*. Online first at `https://doi.org/10.1007/s11229-018-02064-x`.
[19] Pruss, Alexander R. 2020. "Non-classical probabilities invariant under symmetries". `https://arxiv.org/pdf/2010.07366.pdf`
[20] Solomonoff, R. J. 1978. "Complexity-based induction systems: Comparisons and convergence theorems", *IEEE Transactions on Information Theory* 24:422–432.
[21] Williamson, Timothy. 2007. "How probable is an infinite sequence of heads?" *Analysis* 67:173–180.