Finely tuned models sacrifice explanatory depth

Feraz Azhar[*]

Department of Philosophy, University of Notre Dame, Notre Dame, IN 46556, USA

Abraham Loeb[†]

Black Hole Initiative, Harvard University, Cambridge, MA 02138, USA

(Dated: October 28, 2019)

It is commonly argued that an undesirable feature of a theoretical or phenomenological model is that salient observables are sensitive to values of parameters in the model. But in what sense is it undesirable to have such 'fine-tuning' of observables (and hence of the underlying model)? In this paper, we argue that the fine-tuning can be interpreted as a shortcoming of the explanatory capacity of the model: in particular it signals a lack of *explanatory depth*. In support of this argument, we develop a scheme—for models that arise broadly in the sciences—that quantitatively relates fine-tuning of observables described by these models to a lack of depth of explanations based on these models. A significant aspect of our scheme is that, broadly speaking, the inclusion of larger numbers of parameters in a model will decrease the depth of the corresponding explanation. To illustrate our scheme, we apply it in two different settings in which, within each setting, we compare the depth of two competing explanations. The first setting involves explanations for the Euclidean nature of spatial slices of the universe today: in particular, we compare an explanation provided by the big-bang model of the early 1970s (namely, a cosmological model that traces the evolution of the universe back to a singularity without encountering an inflationary period) with an explanation provided by a general model of cosmic inflation. The second setting has a more phenomenological character, where the goal is to infer from a limited sequence of data points, using maximum entropy techniques, the underlying probability distribution from which these data are drawn. In both of these settings we find that our analysis favors the model that intuitively provides the deeper explanation of the observable(s) of interest. We thus provide an account that unifies two 'theoretical virtues' of models used broadly in the sciences—namely, a lack of fine-tuning and explanatory depth—to show that, indeed, *finely tuned models sacrifice explanatory depth.*

Contents

[*]Email address: fazhar@nd.edu
[†]Email address: aloeb@cfa.harvard.edu

I. INTRODUCTION

One hope, in light of what is commonly claimed to be the fine-tuning of our existence—as described by our current favored physical theories—is that there is some alternate theory that renders our existence less finely tuned. Such a hope is not restricted to concerns about life however, for it arises for scientific theories more broadly: in particular, it arises whenever salient observables are deemed to be sensitive to values of free parameters in the theory. If one interprets part of the role that theories play as that of providing an *explanation* for measured values of observables, then a natural question arises: if measured values of observables are finely tuned, does this signal a problematic feature of the corresponding explanation of these measured values? In this paper we answer this question in the affirmative and argue that the feature of such an explanation that is compromised by finely tuned observables is that of its *depth*.

There are two types of theories in which we will be interested, that reflect a broad range of physical settings. The first type encompasses dynamical theories. Such theories can be characterized by (i) equations of motion for dynamical variables and (ii) free parameters, whose values are not fixed by the theory, but which need to be fixed in order to yield agreement with observations. Such free parameters typically come in two varieties: first, as initial conditions for the dynamical variables mentioned in (i) and secondly, as 'constants', that do not explicitly refer to the dynamical variables. In principle, the sensitivity of observables described by such a theory can be due to any combination of (i) and (ii). Our focus will be on sensitivity that arises due to (counterfactual) values of the free parameters as specified in (ii). Our formalism also applies to a second type of scenario encountered in the physical sciences. In particular, it applies to those scenarios in which established theoretical frameworks have yet to be developed but where there exist phenomenological characterizations of data. In such scenarios one can again identify salient observables, and our interest is in the sensitivity of these observables to (counterfactual) values of parameters in such phenomenological models. In what follows (and given what we have just described) we will not need to make much of the distinction between theories and models, and so for the sake of generality we shall refer primarily to "models". (A more precise description of the types of settings in which we are interested appears in Sec. II.)

Such models, we contend, play a crucial role in *explanations* for observables. Of course, what we mean by an explanation needs to be made precise: as is well known, the literature on scientific explanation is vast and controversial [see, for example, Hempel and Oppenheim (1948), Kitcher and Salmon (1989), and Woodward (2003)]. The concept of the *depth* of an explanation—one that has practical appeal—has received relatively less attention. [See, for example, Hitchcock and Woodward (2003) and Weslake (2010).] We commit, in this paper, to a particular schema for explanation that has precursors in the literature [see, for example, work by Woodward and Hitchcock (2003)], but our account of depth is novel. We show, in a way that has been largely ignored in the literature, that the depth of an explanation is related to the degree of fine-tuning of an observable. In a little more detail, the greater the range of counterfactual values of parameters [as described in (ii) above] that yield the same (or similar) values for some salient observable, the less finely tuned is that observable, and the deeper is the explanation of the measured value of the observable.[1] In this way, we provide an operational solution to a key aspect of model choice by identifying and uniting two important 'theoretical virtues' [see, for example, Kuhn (1977), McMullin (1982) and, more

---

[1] See also Ylikoski and Kuorikoski (2010) who note a relationship between explanatory power (or depth) and a lack of sensitivity understood in various ways.

recently, Schindler (2018)]: namely the virtue of a lack of fine-tuning of salient observables and the virtue of a deep explanation. [See also Keas (2018).]

Our plan for this paper is as follows. In Sec. II, we describe our schema for an explanation. We introduce a (quantitative) notion of fine-tuning of observables, adapting previous work of ours [in Azhar and Loeb (2018)], and describe how this notion relates to the depth of an explanation. We provide a measure for depth that depends on our description of fine-tuning that can, in principle, be used to compare the depth of explanations across a broad class of physical settings. An important aspect of our measure is that, broadly speaking, the inclusion of a larger number of parameters in some model will decrease the depth of the corresponding explanation. In Sec. III.A, we apply this measure to compare the depth of cosmological explanations for the Euclidean nature of spatial slices today, where we compare the big-bang model of the 1970s (namely, a cosmological model that traces the evolution of the universe back to a singularity without encountering an inflationary period) with a general model involving cosmic inflation (Guth, 1981). Our measure of depth favors the explanation provided by cosmic inflation. In Sec. III.B we turn our attention to more phenomenological characterizations of data, pursuing a basic problem that arises in situations in which one looks to infer an underlying probability distribution from a finite set of data. We show how our measure of depth, applied in such a way as to discriminate between two models, favors the (intuitively) correct model. We conclude in Sec. IV with a discussion of our results and a comparison with some previous work in the literature.

In short, we present work that conceptually unifies a notion of explanatory depth with a lack of fine-tuning, in a way that applies to a broad range of settings in the sciences. This conceptual unification is combined with a quantitative measure of depth, which one can use to discriminate between models. This measure accurately reflects the types of practical judgements that have been made in the past, as well as those that may be useful in other novel physical settings.

II. EXPLANATORY DEPTH FROM A LACK OF FINE-TUNING

In this section we describe the type of explanation for which we will later derive a measure of depth. This type of explanation will involve the following features.

(i) A set of parameters, $\boldsymbol{p} = \{p_1, p_2, \ldots, p_n\}$, which collectively take values in some (finite) subset, $\mathcal{P}$, of n-dimensional Euclidean space, $\mathbb{R}^n$ ($p_i \in \mathbb{R}$ for each $i = 1, 2, \ldots, n$). These parameters will form part of the explanans.

(ii) A set of observables, $\vec{O} = \{O_1, O_2, \ldots, O_m\}$, which collectively take values in $\mathbb{R}^m$ ($O_j \in \mathbb{R}$ for each $j = 1, 2, \ldots, m$). Specific values for this set of observables will correspond to the explananda. The observables can be thought of as maps acting on the parameters described in (i):

$$\vec{O} : \mathcal{P} \to \mathbb{R}^m$$
$$\boldsymbol{p} \mapsto \vec{O}(\boldsymbol{p}) \equiv \{O_1(\boldsymbol{p}), O_2(\boldsymbol{p}), \ldots, O_m(\boldsymbol{p})\}. \tag{1}$$

We remain open as to the precise manner in which such maps might be realized, but in accordance with the two types of models that we will primarily focus on in this paper (as mentioned in Sec. I), such a map might represent the output of a dynamical system, in which case the parameters would then naturally correspond to initial conditions for any dynamical variables, as well as constants that are left unspecified by the model that underlies the dynamical system. Alternatively, such a map could arise for a model that implements a

non-dynamical and phenomenological relationship between parameters and observables. We will explore models of both types in Sec. III. (In Appendix A we present a more general framework, in which the observables are related to the parameters in a manner that is explicitly probabilistic.)

An explanation on our account will then be an argument in which the parameters take certain values, say, $\boldsymbol{p}'$, such that the mapping in Eq. (1) yields some specific set of values for the observables, say, $\vec{O}_M$. We can represent this argument in the following summarized form (with "E" for "Explanation"):

$$E : \boldsymbol{p}' \wedge [\vec{O}(\boldsymbol{p}') = \vec{O}_M] \therefore \vec{O}_M. \tag{2}$$

[Note that there are similarities between this account of an explanation and that developed by Woodward and Hitchcock (2003)—we will elaborate on this connection in Sec. IV.]

As we mentioned above, our definition of the depth of an explanation will be related to a measure for fine-tuning. In a previous paper, we introduced two perspectives on fine-tuning—a local perspective and a global perspective—with corresponding measures for fine-tuning in each case (Azhar and Loeb, 2018). The perspective on fine-tuning that is relevant for our discussion is the global perspective. We now rehearse and adapt our account in that previous paper for our purposes here.

We will construct a measure for fine-tuning that allows one to compute the degree of fine-tuning of the observables with respect to each parameter *independently* (we focus on the ith parameter in what follows). (Our measure of depth will then combine these independent measures into a single one.) Consider then, at some point $\boldsymbol{p}'$ in parameter space $\mathcal{P}$, some direction, denoted by the unit vector $\hat{\boldsymbol{v}}_i^+$, where this vector points in the (positive) ith direction in parameter space. We need to find the minimum length of the vector, $\boldsymbol{v}_i^+$, that points in this direction, such that a significant change in the observables, computed at the point $\boldsymbol{p}' + \boldsymbol{v}_i^+$, occurs. One way in which we can quantify such a change is, for example, via an order-unity fractional change in the observables [as defined in Azhar and Loeb (2018)]. To be precise, we need to find $|\boldsymbol{v}_i^+|$ such that

$$\frac{|\vec{O}(\boldsymbol{p}' + \boldsymbol{v}_i^+) - \vec{O}(\boldsymbol{p}')|}{|\vec{O}(\boldsymbol{p}')|} \sim \mathcal{O}(1). \tag{3}$$

Another way in which we can consider a change in the observables to be significant is when any one of the observables that comprise $\vec{O}$ lies outside bounds established by experimental considerations. In our examples in Sec. III, we will indeed study such scenarios.

Having chosen a means to characterize changes in observables as significant, one then repeats the construction for the negative ith direction. Thus we find the length of the vector, $\boldsymbol{v}_i^-$, that points in this direction that leads to a significant change in the observables. The sum of the lengths of these two vectors will then correspond to the total range along the ith parameter direction over which observables do not change by a significant amount. Denote the size of this range by $|\boldsymbol{v}_i^\pm|$, where

$$|\boldsymbol{v}_i^\pm| \equiv |\boldsymbol{v}_i^+| + |\boldsymbol{v}_i^-|. \tag{4}$$

Furthermore, let $\Delta[\boldsymbol{p}', \hat{\boldsymbol{v}}_i^+]$ denote the size of the range of *allowed* parameter values (as encoded in one's definition of $\mathcal{P}$) in the positive ith direction starting at $\boldsymbol{p}'$, with $\Delta[\boldsymbol{p}', \hat{\boldsymbol{v}}_i^-]$ denoting the corresponding size in the negative ith direction. The sum of these two sizes, which we will denote

by $\Delta[\boldsymbol{p}', \hat{\boldsymbol{v}}_i^{\pm}]$, where

$$\Delta[\boldsymbol{p}', \hat{\boldsymbol{v}}_i^{\pm}] \equiv \Delta[\boldsymbol{p}', \hat{\boldsymbol{v}}_i^{+}] + \Delta[\boldsymbol{p}', \hat{\boldsymbol{v}}_i^{-}], \tag{5}$$

is the size of the allowed range considering both directions. The degree of global fine-tuning of the observables with respect to the ith parameter is then given by

$$\mathcal{G}_i(\vec{O}; \boldsymbol{p}') \equiv \log_{10} \left(\frac{\Delta[\boldsymbol{p}', \hat{\boldsymbol{v}}_i^{\pm}]}{|\boldsymbol{v}_i^{\pm}|} \right). \tag{6}$$

As in Azhar and Loeb (2018), we invoke the convention that if in some direction in parameter space, one reaches the edge of parameter space and no significant change in the observables has occurred, then the length of the vector that is encoded in the above formalism is the full distance that one can indeed travel in that direction.

This measure of (global) fine-tuning [in Eq. (6)] has the following features.

(a) $\mathcal{G}_i(\vec{O}; \boldsymbol{p}') \geq 0$ since, by construction, $|\boldsymbol{v}_i^{\pm}| \leq \Delta[\boldsymbol{p}', \hat{\boldsymbol{v}}_i^{\pm}]$.

(b) The minimum value of $\mathcal{G}_i(\vec{O}; \boldsymbol{p}')$ (viz. zero), occurs when $|\boldsymbol{v}_i^{\pm}| = \Delta[\boldsymbol{p}', \hat{\boldsymbol{v}}_i^{\pm}]$; the maximum value is unbounded. The minimum value corresponds to what we will call the *order-0 case*—in which the observables are minimally finely tuned. The level of fine-tuning of the observables increases as $\mathcal{G}_i(\vec{O}; \boldsymbol{p}')$ increases.

(c) The use of the logarithm is so that (for example) order unity differences in the value of $\mathcal{G}_i(\vec{O}; \boldsymbol{p}')$ can be thought of as being 'significantly' different.

Our measure of the depth of the explanation E, which we denote by $D_E(\vec{O}; \boldsymbol{p}')$, combines the degree of fine-tuning of the observables, with respect to each parameter direction, into a single number:

$$D_E(\vec{O}; \boldsymbol{p}') \equiv \frac{1}{\prod\limits_{i=1}^{n} \left[1 + \mathcal{G}_i(\vec{O}; \boldsymbol{p}') \right]}. \tag{7}$$

This measure of depth has the following features.

(i) It ranges between 0 and 1, viz. $0 < D_E(\vec{O}; \boldsymbol{p}') \leq 1$. The lower bound is approached as the degree of fine-tuning of the observables is large with respect to at least one of the parameters [that is, e.g., $\mathcal{G}_1(\vec{O}; \boldsymbol{p}') \to \infty \implies D_E(\vec{O}; \boldsymbol{p}') \to 0$]. The upper bound is saturated when the order-0 case is satisfied for *each* parameter [viz. $\forall i, \mathcal{G}_i(\vec{O}; \boldsymbol{p}') = 0 \iff D_E(\vec{O}; \boldsymbol{p}') = 1$].

(ii) Within the range described in (i), $D_E(\vec{O}; \boldsymbol{p}')$ is strictly monotonically decreasing with an increasing level of fine-tuning with respect to any parameter.

(iii) Ceteris paribus, adding another parameter (indexed by k, say) to the explanation, *always* reduces the depth of the explanation, unless the observables satisfy the order-0 case with respect to that parameter [viz. unless $\mathcal{G}_k(\vec{O}; \boldsymbol{p}') = 0$]. So, ceteris paribus, adding a parameter with respect to which the observables are insensitive does not change the depth of the resulting explanation.

(iv) The form of our measure of depth is very general. If we have two different models, say, with a common set of observables $\vec{O}_c$ (where, say, $\vec{O}_c$ is a proper subset of the observables described by each model), one could use our measure to compare the depth of the explanation provided by each model for (measured values of) just those observables. One could also use our measure to compare, for a single model that describes some vector of observables $\vec{O} = \{O_1, O_2, \dots\}$, the depth of the explanation provided for different observables in that vector: as in, for example, a comparison of the depth of the explanation for (the measured value of) O_1 vs. the depth of the explanation for O_2.

(v) Finally, note that the precise functional form that preserves the features listed above [viz. (i)–(iv)] is not unique: a monotonically increasing function f, say, could be inserted in the denominator in Eq. (7) and the features described above would continue to hold. So we could have defined, for example,

$$D_E(\vec{O}; \boldsymbol{p}') \equiv \left(\prod_{i=1}^{n} f\left[1 + \mathcal{G}_i(\vec{O}; \boldsymbol{p}') \right] \right)^{-1}, \tag{8}$$

and maintained the above features. Our choice of definition in Eq. (7) has been made for the sake of simplicity.

In sum, we have proposed a scheme to compute the depth of an explanation in terms of the degree of fine-tuning of the explananda, in which, for example, a deeper explanation is one in which the degree of fine-tuning is smaller. In the following section we will apply this scheme to two distinct examples to illustrate its scope.

III. APPLICATIONS OF OUR APPROACH

Our examples correspond to two (very) different settings in which explanations arise in the sciences. The first setting corresponds to one in which there is a theoretical framework for describing the system of interest—from which a dynamical system can be used to extract observables. The example we will pursue in this context is from cosmology—where we will compare two separate explanations for the Euclidean nature of spatial slices of our universe today. The second setting corresponds to one in which such theoretical frameworks do not (yet) exist but where phenomenological characterizations of observables have been developed. The example we will pursue in this context compares two explanations that characterize a stream of data using maximum entropy techniques. For both examples, we show that the explanation that one might expect to furnish a deeper explanation is indeed the one favored by our measure of depth.

A. The flatness problem in cosmology

The universe today is consistent with being spatially flat, that is, the geometry of space can be described as a three-dimensional Euclidean space. There are two different explanations of this observation that we will probe in this section. The first arises in the context of the big-bang model (BBM) of cosmology as it had been developed by the early 1970s—this model traces the evolution of the universe back to a singularity without invoking a period of cosmic inflation. The second arises, indeed, for a general model of cosmic inflation, in which the very early universe undergoes accelerated expansion for a short period of time [see Guth (1981)]. A key motivation for the introduction of cosmic inflation was to overcome perceived explanatory shortcomings in the BBM:

one of those being the finely tuned nature of the spatial flatness of the universe today according to the BBM. This comparison thus serves as a particularly appropriate setting in which to probe our account of explanatory depth.

In describing both of these accounts, we will assume, for the sake of simplicity that spatial sections of the universe are homogeneous and isotropic from the outset. Newton's gravitational constant will be denoted by G and we will set the speed of light to be unity, viz. $c \equiv 1$. We begin by first describing how the geometry of space evolves according to the BBM [see also Brawer (1996)], before connecting this description to the scheme for explanation introduced in Sec. II. We will then repeat this analysis for the case of cosmic inflation.

In our discussion of the BBM, we will consider an idealized setting in which the universe undergoes two sequential phases, where in each phase there is, in effect, a single (fluid) source that drives (and is driven by) the expansion of the universe. The first phase will be a radiation-dominated phase and this will be followed by a matter-dominated phase (where "matter" refers to pressureless "dust"). The radiation-dominated phase will last from some initial time t_i until t_T, which will denote the time of matter-radiation equality (an equality of mass/energy density). The matter-dominated phase will last from t_T until today, denoted by t_0.

The evolution of a homogeneous and isotropic space in any such phase can be described, in the context of general relativity, by Friedmann's equations. If, as is conventional, we let the physical (spatial) separation of points be determined by the scale factor, $a(t)$, then the first Friedmann equation describes how the evolution of the scale factor is related to the mass density of the universe, $\rho(t)$, as well as the curvature of spatial sections, represented by k ($k = 0$ corresponds to the situation where space is Euclidean). In particular let $H(t) \equiv \dot{a}(t)/a(t)$ denote the Hubble parameter, where overdots will denote derivatives with respect to time. Then the first Friedmann equation is given by

$$H^2(t) = \frac{8\pi G}{3}\rho(t) - \frac{k}{a^2(t)}. \tag{9}$$

We denote the critical density by $\rho_{\mathrm{crit}}(t) \equiv 3H^2(t)/(8\pi G)$ [viz. the density such that spatial sections, according to Eq. (9), are flat]. Then we can define the key variable that will be relevant in our discussion below, viz. the dimensionless density parameter $\Omega(t)$:

$$\Omega(t) \equiv \frac{\rho(t)}{\rho_{\mathrm{crit}}(t)}. \tag{10}$$

Note that $\Omega(t) = 1$ corresponds to flat spatial sections.[2] A useful quantity for our analysis below will be the deviation of $\Omega(t)$ from unity (viz. the deviation of spatial sections from flatness). A straightforward calculation gives, from Eqs. (9) and (10), that the fractional deviation from unity, of the dimensionless density parameter, is given by

$$\frac{\Omega(t) - 1}{\Omega(t)} = \frac{3k}{8\pi G\rho(t)a^2(t)}. \tag{11}$$

---

[2] Note also that in later connecting these introductory remarks to our scheme for explanation, the value of the dimensionless density parameter at some initial time will correspond to the "parameter" in the corresponding explanation (what we have earlier called $\boldsymbol{p}'$), whereas the value of the dimensionless density parameter today will correspond to the "observable" (what we have earlier called $\vec{O}$). In what follows—keeping with conventions in the cosmology literature—we will continue to refer to the function defined in Eq. (10) as a (dimensionless density) parameter.

The other equation we will require follows from the law of conservation of stress-energy and describes how the mass density of some (fluid) source evolves with time. This equation is given by

$$\dot{\rho}(t) + 3H(t)\left[\rho(t) + P(t)\right] = 0, \tag{12}$$

where $P(t)$ is the pressure of the fluid. In what follows we will assume that for any fluid source the pressure is related to the mass density by an equation of state in which $P(t) = w\rho(t)$, where $w = 1/3$ for radiation and $w = 0$ for dust. Assuming that we know, at some final time t_f, the values of $\rho(t_f)$ and $a(t_f)$, we can integrate Eq. (12) to show, for $t < t_f$, that

$$\rho(t) = \rho(t_f)\left[\frac{a(t_f)}{a(t)}\right]^{3(1+w)}. \tag{13}$$

We can now combine Eqs. (11) and (13) to show that for $t < t_f$,

$$\frac{\Omega(t) - 1}{\Omega(t)} = \left[\frac{a(t)}{a(t_f)}\right]^{1+3w} \frac{\Omega(t_f) - 1}{\Omega(t_f)}. \tag{14}$$

We now apply the above machinery to an analysis of our idealized version of the BBM, beginning from the present time, viz. t_0, and working our way back in time. Invoking Eq. (14) for the matter-dominated phase we have

$$\frac{\Omega(t_T) - 1}{\Omega(t_T)} = \frac{a(t_T)}{a(t_0)} \frac{\Omega(t_0) - 1}{\Omega(t_0)}. \tag{15}$$

For the matter-dominated phase it is straightforward to show, combining Eqs. (9) and (13), that

$$\frac{a(t)}{a(t_0)} = \left(\frac{t}{t_0}\right)^{2/3}, \tag{16}$$

so that Eq. (15) becomes

$$\frac{\Omega(t_T) - 1}{\Omega(t_T)} = \left(\frac{t_T}{t_0}\right)^{2/3} \frac{\Omega(t_0) - 1}{\Omega(t_0)}. \tag{17}$$

(Note that for the sake of simplicity, in deriving the evolution of the scale factor in Eq. (16) we have assumed that $k = 0$—such an assumption will suffice for our purposes in this section.) We can now connect the left-hand side of Eq. (17) to the earliest time we will consider in the context of the BBM, viz. t_i, by an application of Eq. (14) to the radiation-dominated phase. We find

$$\frac{\Omega(t_i) - 1}{\Omega(t_i)} = \left[\frac{a(t_i)}{a(t_T)}\right]^2 \frac{\Omega(t_T) - 1}{\Omega(t_T)}. \tag{18}$$

For the radiation-dominated phase, one can show, from Eqs. (9) and (13) (again assuming $k = 0$), that

$$\frac{a(t)}{a(t_T)} = \left(\frac{t}{t_T}\right)^{1/2}, \tag{19}$$

so that Eq. (18) becomes

$$\frac{\Omega(t_i) - 1}{\Omega(t_i)} = \frac{t_i}{t_T} \frac{\Omega(t_T) - 1}{\Omega(t_T)}. \tag{20}$$

Combining Eqs. (17) and (20) we thus obtain an expression for the fractional deviation from unity, of the density parameter, at the initial time t_i, as it relates to such a deviation today, viz.

$$\frac{\Omega(t_i) - 1}{\Omega(t_i)} = \frac{t_i}{t_T} \left(\frac{t_T}{t_0}\right)^{2/3} \frac{\Omega(t_0) - 1}{\Omega(t_0)}. \tag{21}$$

To determine t_T, viz. the time corresponding to matter-radiation equality, we apply Eq. (13) to each component separately, so that (using the subscripts "R" for radiation and "M" for matter)

$$\rho_R(t) = \rho_R(t_0) \left[\frac{a(t_0)}{a(t)}\right]^4, \tag{22}$$

$$\rho_M(t) = \rho_M(t_0) \left[\frac{a(t_0)}{a(t)}\right]^3. \tag{23}$$

Dividing these equations we find that

$$\frac{\rho_R(t)}{\rho_M(t)} = \frac{\rho_R(t_0)}{\rho_M(t_0)} \frac{a(t_0)}{a(t)}. \tag{24}$$

At t_T, the left-hand side of the above equation is unity (by definition of t_T), whereas $a(t)$ on the right-hand side can be treated as though we are in the matter-dominated phase [as for Eq. (16)]. Thus we find that

$$t_T = \left(\frac{\Omega_{R,0}}{\Omega_{M,0}}\right)^{3/2} t_0, \tag{25}$$

where we have used the convention that $\Omega_{i,0} \equiv \rho_i(t_0)/\rho_{crit}(t_0)$. Each term on the right-hand side of Eq. (25) can either be read off or estimated from recent cosmological data.[3] We find, using the above approximations, that $t_T \approx 3.1 \times 10^4$ years.

Recent cosmological data allows us to find an upper bound for the size of the density parameter today, $\Omega(t_0)$. In particular, the *Planck* Collaboration's most recent results on cosmological parameters (*Planck* Collaboration, 2018) imply, for our purposes, that

$$1 - \Omega(t_0) = 0.001 \pm 0.002. \tag{26}$$

This implies that the magnitude of the fractional deviation from unity of the density parameter today [viz. the right-most term on the right-hand side of Eq. (21)] must be constrained in the following way so that the present-day value of the density variable agrees with cosmological observations:

$$\left|\frac{\Omega(t_0) - 1}{\Omega(t_0)}\right| < \frac{0.003}{0.997} \equiv \left|\frac{\Omega(t_0) - 1}{\Omega(t_0)}\right|_{max}. \tag{27}$$

---

[3] The computation of $\Omega_{R,0}$ is a little subtle. To estimate this latter quantity from recent cosmological data we assume that the energy density in radiation today, $\rho_{R,0}$, can be related to the temperature, T_0, of radiation today (in the cosmic microwave background). One finds that $\Omega_{R,0} \propto T_0^4/H_0^2$.

Thus from Eq. (21) we obtain a constraint on the *initial* density variable, viz.

$$\left| \frac{\Omega(t_i) - 1}{\Omega(t_i)} \right| < \frac{t_i}{t_T} \left(\frac{t_T}{t_0} \right)^{2/3} \left| \frac{\Omega(t_0) - 1}{\Omega(t_0)} \right|_{\max} \equiv \epsilon. \tag{28}$$

Note that assuming the time t_i refers to a time that is very early in the history of the universe, the upper bound, ϵ, expressed in Eq. (28) is small. One finds therefore that

$$\frac{1}{1 + \epsilon} < \Omega(t_i) < \frac{1}{1 - \epsilon}. \tag{29}$$

Thus the size of the entire range of values of $\Omega(t_i)$ that yield agreement with cosmological observations, which we will denote by $\Delta\Omega_i$, is given by

$$\Delta\Omega_i \equiv \frac{2\epsilon}{1 - \epsilon^2}. \tag{30}$$

We can now explicitly connect this analysis with the formalism described in Sec. II. In particular, the parameters involved in the BBM's explanation of the flatness of the universe today, viz. $\boldsymbol{p}' \rightarrow \Omega(t_i)$ and $\vec{O} \rightarrow \Omega(t_0)$. (Of course, the mapping between parameters and observables that enters into this explanation has a dynamical character.) Denote the maximum possible value of $\Omega(t_i)$ by $\Omega_{\max}$ (where we will later consider situations in which $\Omega_{\max} \geq 10$), the degree of global fine-tuning of the observable is given by [adapting Eq. (6)]

$$\mathcal{G}(\Omega(t_0); \Omega(t_i)) = \log_{10}\left(\frac{\Omega_{\max}}{\Delta\Omega_i} \right) = \log_{10}\left(\frac{\Omega_{\max}}{2\epsilon/(1 - \epsilon^2)} \right), \tag{31}$$

so that the depth of the explanation provided by the BBM, which we will denote by D_{BBM} is [adapting Eq. (7)]

$$D_{\mathrm{BBM}}(\Omega(t_0); \Omega(t_i)) = \left[1 + \log_{10}\left(\frac{\Omega_{\max}}{2\epsilon/(1 - \epsilon^2)} \right) \right]^{-1}. \tag{32}$$

For the case of cosmic inflation, we assume that there is a period of N e-folds of inflation that occur just before t_i, beginning at the time t_{Inf}, so that $a(t_i) = e^N a(t_{\mathrm{Inf}})$. Over this period of time, we will assume that the mass/energy density of the universe was constant. Thus, invoking Eq. (11) with $\rho(t_{\mathrm{Inf}}) = \rho(t_i)$, we find that

$$\frac{\Omega(t_{\mathrm{Inf}}) - 1}{\Omega(t_{\mathrm{Inf}})} = e^{2N} \frac{\Omega(t_i) - 1}{\Omega(t_i)}, \tag{33}$$

so that we have the following present-day observational bound on the density parameter at the beginning of inflation:

$$\left| \frac{\Omega(t_{\mathrm{Inf}}) - 1}{\Omega(t_{\mathrm{Inf}})} \right| = e^{2N} \left| \frac{\Omega(t_i) - 1}{\Omega(t_i)} \right| < e^{2N} \epsilon, \tag{34}$$

where the right-hand side of Eq. (34) uses Eq. (28). We see immediately that to obtain agreement with cosmological observations today, the fractional deviation from unity, of the density parameter at the beginning of inflation, can be significantly greater than the deviation at t_i. Indeed, more specifically, if there is *any* inflation ($N > 0$), then the initial allowable fractional deviation from unity of the density parameter, is greater for inflation than in the case of the BBM.

10

From Eq. (34), we can bound admissible values for $\Omega(t_{\text{Inf}})$ depending on the size of $e^{2N}\epsilon$ relative to unity:

$$\frac{1}{1+e^{2N}\epsilon} < \Omega(t_{\text{Inf}}) < \begin{cases} \Omega_{\text{max}}, & e^{2N}\epsilon \geq 1 \quad (35) \\ \min\left\{\frac{1}{1-e^{2N}\epsilon}, \Omega_{\text{max}}\right\} & e^{2N}\epsilon < 1. \quad (36) \end{cases}$$

When $e^{2N}\epsilon \geq 1$, viz. $N \geq (1/2)\ln(\epsilon^{-1})$, the total number of e-folds of inflation will be said to be "large" ($N \to N_{\text{L}}$). When $e^{2N}\epsilon < 1$, viz. $N < (1/2)\ln(\epsilon^{-1})$, the total number of e-folds of inflation will be said to be "small" ($N \to N_{\text{S}}$). A value of $N = (1/2)\ln(\epsilon^{-1})$ will correspond to the "critical number" of e-folds ($N \to N_{\text{critical}}$). As we shall see shortly (in a particular case), for $N > N_{\text{critical}}$, the depth of the explanation provided by inflation jumps to be close to unity (compared with $N < N_{\text{critical}}$), thus effectively solving the flatness problem.

We can now connect this inflationary analysis with the formalism described in Sec II. The relevant parameter involved in the explanation, provided by cosmic inflation, for the flatness of the universe today, is $\boldsymbol{p}' \to \Omega(t_{\text{Inf}})$ and the observable of interest is, as for the case of the BBM, $\vec{O} \to \Omega(t_0)$. (Again, the mapping between parameters and observables that enters into this explanation has a dynamical character.) The degree of global fine-tuning of the observable depends on whether the total number of e-folds is large or small (according to our convention in the previous paragraph). We find [adapting Eq. (6)] that

$$\mathcal{G}(\Omega(t_0); \Omega(t_{\text{Inf}})) = \begin{cases} \log_{10}\left(\dfrac{\Omega_{\text{max}}}{\Omega_{\text{max}} - \frac{1}{1+e^{2N}\epsilon}}\right) & e^{2N}\epsilon \geq 1 \quad (37) \\[4mm] \log_{10}\left(\dfrac{\Omega_{\text{max}}}{\min\left\{\frac{1}{1-e^{2N}\epsilon}, \Omega_{\text{max}}\right\} - \frac{1}{1+e^{2N}\epsilon}}\right) & e^{2N}\epsilon < 1. \quad (38) \end{cases}$$

The depth of the explanation provided by inflation, which we will denote by D_{INF} is thus [adapting Eq. (7)]

$$D_{\text{INF}}(\Omega(t_0); \Omega(t_{\text{Inf}})) = \begin{cases} \left[1 + \log_{10}\left(\dfrac{\Omega_{\text{max}}}{\Omega_{\text{max}} - \frac{1}{1+e^{2N}\epsilon}}\right)\right]^{-1} & e^{2N}\epsilon \geq 1 \quad (39a) \\[4mm] \left[1 + \log_{10}\left(\dfrac{\Omega_{\text{max}}}{\min\left\{\frac{1}{1-e^{2N}\epsilon}, \Omega_{\text{max}}\right\} - \frac{1}{1+e^{2N}\epsilon}}\right)\right]^{-1} & e^{2N}\epsilon < 1. \quad (39b) \end{cases}$$

In computing numerical estimates of the depth of the explanation for the flatness of the universe today, as described by the BBM and by cosmic inflation, we need to first estimate ϵ as it appears in Eq. (28). Using $t_0 = 13.8\text{Gyr}$ and a fiducial value of $t_i = 10^{-34}\text{s}$ (which is a rough estimate of the time at which inflation ends), we find that $\epsilon = \mathcal{O}(10^{-52})$. This yields $N_{\text{critical}} \approx 60$.

Figure 1 compares depths for the BBM and cosmic inflation as computed from Eqs. (32), (39a), and (39b). We note that the depth of the BBM (in blue) is always (i.e., for all chosen values of Ω_{max}) less than the depth as computed in the case of inflation (gold and green lines). The gold lines represent the depth as it appears in Eq. (39b), viz. in cases where the number of e-folds of inflation is small (i.e., less than the critical number of e-folds, $N_{\text{critical}} \approx 60$). The green lines represent the depth as it appears in Eq. (39a), viz. in cases where the number of e-folds of inflation
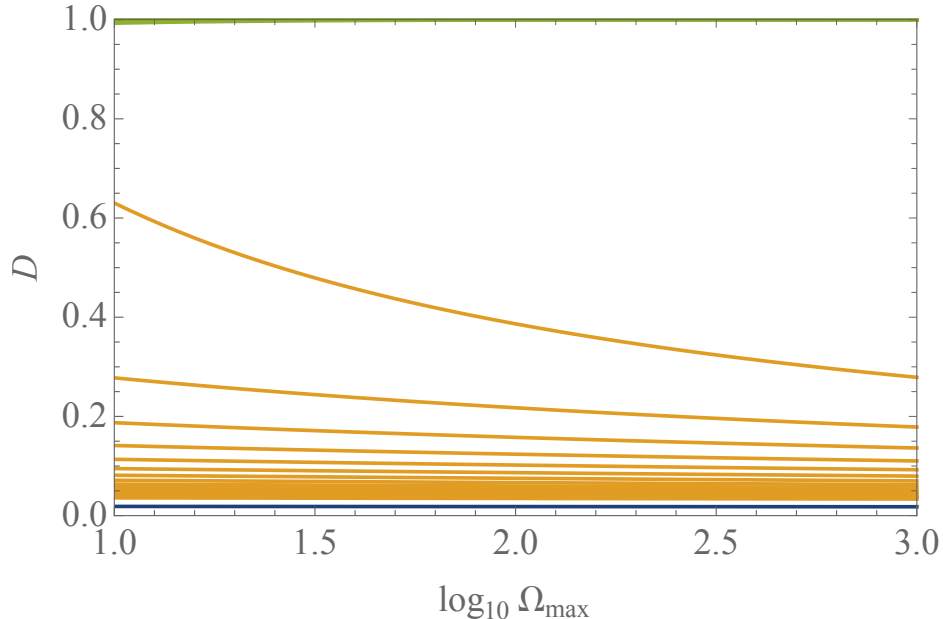
FIG. 1 The depth, D, of the explanation for the flatness of spatial slices of our universe today as a function of the maximum possible value of the initial density parameter. The depth can vary, in principle, over the range $(0, 1]$. The blue line corresponds to the depth as computed for the big-bang model (BBM) from Eq. (32). The gold lines correspond to the depth as computed for different inflationary scenarios in which each successive gold line starting from the bottom of the figure corresponds to an increasing number of assumed e-folds of inflation, N, from $N = 30$ through to $N = 60$ (in steps of two e-folds). The green lines display the depth as computed for inflationary scenarios in which N ranges from $N = 61$ through to $N = 91$ (in steps of 10 e-folds). We see that there is a distinct jump in the depth of the explanation when the number of e-folds is increased beyond a critical value (referred to as N_{critical} in the main text). For the parameters chosen in the main text, $N_{\mathrm{critical}} \approx 60$, in agreement with the generally accepted value for the total number of e-folds of inflation that is thought to solve various fine-tuning problems with the BBM (viz. problems that include the flatness problem). [See Azhar and Butterfield (2018) for further discussion of this point.]

is large (and greater than the critical number of e-folds).[4] Cosmic inflation thereby provides a deeper explanation for the flatness of our universe today compared with the BBM (as is commonly claimed).

B. Model selection in maximum-entropy modeling

We now turn to another example with a different character from the one described in Sec. III.A. In particular, this example probes the depth of explanations provided by phenomenological models of data. The basic problem is as follows: one observes data presented as a sequence of M samples, say, $x_1, x_2, \ldots, x_M$, where we know that each sample is drawn independently from the same underlying probability distribution $P(x)$. The question of interest is: what is the probability distribution? In the absence of any further information we have no way of determining a unique

---

[4] Note that increasing the number of e-folds of cosmic inflation well beyond the critical value ($N_{\mathrm{critical}} \approx 60$) does not (and indeed cannot) lead to a noticeable increase in the depth of the explanation (i.e., the green lines in Fig. 1 effectively lie on top of each other near the maximum value of unity). This is consistent with the fact that the observable universe today circumscribes a horizon corresponding to the (last) 60 e-folds of inflation.

distribution. But we can imagine constraining possibilities for such distributions by demanding that they satisfy constraints derived from the data. Of course the set of such possible distributions is biased by the chosen constraints, but there is a class of distributions—known as maximum entropy distributions—that provide a useful (and relatively principled) guide.

Maximum entropy distributions, in short, are those that maximize the classical, Shannon-information theoretic entropy, subject to specified constraints. More formally, when the (random) variable of interest is defined over some continuous domain $\mathcal{D}$, the maximum entropy distribution is constructed by maximizing, over all positive functions $P(x) \geq 0$, the differential entropy $S[P(x)]$, where

$$S[P(x)] \equiv -\int_{\mathcal{D}} dx \, P(x) \ln P(x), \tag{40}$$

subject to K constraints defined through some K functions $f_i(x)$:

$$\int_{\mathcal{D}} dx \, P(x) f_i(x) = \mu_i, \tag{41}$$

for $i = 1, 2, \ldots, K$. In addition, we must also enforce the usual normalization constraint:

$$\int_{\mathcal{D}} dx \, P(x) = 1. \tag{42}$$

A well-known theorem [see Cover and Thomas (1991, Ch. 11)] provides the answer to this optimization problem:

$$P(x; \alpha_0, \alpha_1 \ldots, \alpha_K) = e^{\alpha_0 + \sum_{i=1}^{K} \alpha_i f_i(x)}, \tag{43}$$

where $\alpha_0, \alpha_1, \ldots, \alpha_K$ need to be chosen so that Eqs. (41) and (42) are satisfied. Note that in what follows, we use semicolons to delimit values of the underlying random variable (i.e., x, above), from parameters on which the distribution depends (i.e., $\alpha_0 \ldots, \alpha_K$, above).

From an interpretative point of view, these distributions [as represented by Eq. (43)] are the distributions that maximize our uncertainty (understood in information theoretic terms) about the underlying data given the constraints. In this sense they are the 'least structured' distributions, consistent with the constraints. There has been renewed interest in their use, for example, in the biological sciences, especially in the case of underlying random variables that are discrete, and where the constraints of interest are of a particularly simple type.[5]

We will describe a toy example that captures the sense in which our measure of explanatory depth can play a role in choosing between maximum entropy models. Consider the scenario in which we draw M samples, $x_1, x_2, \ldots, x_M$, from a known continuous probability distribution $P(x)$. The distribution we choose is the exponential distribution and takes the form

$$P(x) = \mu \exp(-\mu x) \qquad x \geq 0, \mu > 0. \tag{44}$$

In what follows, to be concrete, we choose $\mu = 1$. We will look to fit two different maximum

---

[5] We refer here to maximum entropy distributions that are consistent with one- and two-point correlation functions derived from the data (we will indeed probe such scenarios in the main text—though where the underlying random variable is continuous). There are a variety of biological systems that have been explored in this way, including: neural systems, proteins, the immune system, and even aggregations of birds [see Tkačik and Bialek (2016) for a review].

entropy distributions to the data and will compare the depth of the explanations—about which we will say more shortly—provided by these distributions. The first distribution will be the maximum entropy distribution consistent with the empirical mean of the sample, denoted by $\langle x \rangle_\mathrm{e}$ ("e" for "empirical"). This distribution thus maximizes Eq. (40), for $\mathcal{D} \equiv [0, \infty)$, subject to a single constraint, as in Eq. (41) with $f_1(x) \equiv x$ and $\mu_1 \equiv \langle x \rangle_\mathrm{e}$. From Eq. (43), this distribution takes a form that we will write as

$$P_1(x; \lambda_0, \lambda_1) = \exp\left(-1 - \lambda_0 - \lambda_1 x\right). \tag{45}$$

Of course, this distribution has the same form as that from which the data are drawn [viz. Eq. (44)], and so we expect the depth of the explanation of the data, derived from this distribution, to be relatively high.

The second distribution we will consider will be a maximum entropy distribution consistent with the first two empirically derived moments, viz. $\langle x \rangle_\mathrm{e}$ and $\langle x^2 \rangle_\mathrm{e}$. This distribution thus maximizes Eq. (40), again for $\mathcal{D} \equiv [0, \infty)$, subject to two constraints that take the form of Eq. (41): one constraint has $f_1(x) \equiv x$ and $\mu_1 \equiv \langle x \rangle_\mathrm{e}$, whereas the second constraint has $f_2(x) \equiv x^2$ and $\mu_2 \equiv \langle x^2 \rangle_\mathrm{e}$. From Eq. (43), this maximum entropy distribution takes the form

$$P_2(x; \lambda_0, \lambda_1, \lambda_2) = \exp\left(-1 - \lambda_0 - \lambda_1 x - \lambda_2 x^2\right). \tag{46}$$

In each case, the empirically derived moments will also have empirically derived standard deviations associated with them so that each moment that we consider is properly represented as

$$\langle x^k \rangle_\mathrm{e} \pm \sigma_k, \tag{47}$$

where

$$\sigma_k \equiv \left[\frac{1}{M(M-1)} \sum_{i=1}^{M} \left(x_i^{\ k} - \langle x^k \rangle_\mathrm{e} \right)^2 \right]^{1/2}, \tag{48}$$

and where $x_i^{\ k}$ refers to the ith sample (out of the M samples) raised to the kth power.

In what sense, now, do maximum entropy distributions furnish an explanation of the data? The explanandum in which we will be interested will not correspond to the exact set of samples obtained but corresponds to the statistics of the sequence of samples as characterized by the *first two moments*. The rationale for this is that the precise set of samples are *representative* of the underlying distribution and what is salient about those samples is aggregate properties as captured by statistics that are deemed relevant. The first two moments represent two particularly simple features of the statistics of the data.

In the case of the first maximum entropy model, represented by $P_1(x; \lambda_0, \lambda_1)$, the parameters involved in the explanation are given by $\boldsymbol{p}' \rightarrow (\lambda_0, \lambda_1)$ and the observables are given by $\vec{O} \rightarrow \left(\langle x \rangle_\mathrm{e}, \langle x^2 \rangle_\mathrm{e}\right)$. The explanation is non-dynamical and phenomenological in character. One can represent the explanation, denoted by E_1, in the following way:

$$E_1 : (\lambda_0, \lambda_1) \wedge \left[\int_0^\infty dx\ P_1(x; \lambda_0, \lambda_1) = 1, \int_0^\infty dx\ P_1(x; \lambda_0, \lambda_1)x = \langle x \rangle_\mathrm{e} \right] \therefore \left(\langle x \rangle_\mathrm{e}, \langle x^2 \rangle_\mathrm{e} \pm \sigma_2\right). \tag{49}$$

Note that in this first case, we relax the precision with which one needs to account for the second moment $\langle x^2 \rangle_\mathrm{e}$, as $P_1(x; \lambda_0, \lambda_1)$ is not defined in such a way as to demand agreement with the precise value of the second moment [unlike in the case of $P_2(x; \lambda_0, \lambda_1, \lambda_2)$]. Thus in what follows, we allow

for the explananda in this first case to correspond, effectively, to $\vec{O} \to \vec{O}_1 \equiv \left(\langle x \rangle_{\mathrm{e}}, \langle x^2 \rangle_{\mathrm{e}} \pm \sigma_2 \right)$.

Similarly, in the case of the second maximum entropy model, represented by $P_2(x; \lambda_0, \lambda_1, \lambda_2)$, the parameters are given by $\boldsymbol{p}' \to (\lambda_0, \lambda_1, \lambda_2)$ and the observables are given by $\vec{O} \to \left(\langle x \rangle_{\mathrm{e}}, \langle x^2 \rangle_{\mathrm{e}} \right)$. One can represent the explanation, denoted by E_2, in the following way:

$$E_2 : (\lambda_0, \lambda_1, \lambda_2) \wedge \left[\int_0^\infty dx \, P_2(x; \lambda_0, \lambda_1, \lambda_2) = 1, \int_0^\infty dx \, P_2(x; \lambda_0, \lambda_1, \lambda_2)x = \langle x \rangle_{\mathrm{e}}, \right.$$
$$\left. \int_0^\infty dx \, P_2(x; \lambda_0, \lambda_1, \lambda_2)x^2 = \langle x^2 \rangle_{\mathrm{e}} \right] \therefore (\langle x \rangle_{\mathrm{e}}, \langle x^2 \rangle_{\mathrm{e}}). \tag{50}$$

Our goal then is to compare the depth of the explanations provided by E_1 and E_2, which we will denote by D_{E_1} and D_{E_2}, respectively.

Let us begin by analyzing E_1. In this case, one can solve for the parameter values λ_0 and λ_1 analytically. In particular, we'll rewrite the distribution in Eq. (45) in the following way:

$$P_1(x; Z_1, \lambda_1) \equiv \frac{1}{Z_1} \exp\left(-\lambda_1 x \right), \tag{51}$$

where $Z_1 \equiv \exp\left(1 + \lambda_0 \right)$. Enforcing the normalization constraint [viz. the first equation in square brackets in Eq. (49)] gives $Z_1 \to Z_{1:\mathrm{sol}}(\lambda_1) \equiv 1/\lambda_1$; then enforcing the remaining constraint [viz. the second equation in square brackets in Eq. (49)] yields $\lambda_1 \to \lambda_{1:\mathrm{sol}} \equiv 1/\langle x \rangle_{\mathrm{e}}$. Thus we find

$$P_1(x; Z_{1:\mathrm{sol}}(\lambda_{1:\mathrm{sol}}), \lambda_{1:\mathrm{sol}}) = \frac{1}{\langle x \rangle_{\mathrm{e}}} \exp\left(-\frac{x}{\langle x \rangle_{\mathrm{e}}} \right). \tag{52}$$

Now, we interpret this distribution as explaining the data if (and only if) it is also the case that the second moment of this distribution satisfies the corresponding experimental constraint, namely, that $\int_0^\infty dx \, P_1(x; Z_{1:\mathrm{sol}}(\lambda_{1:\mathrm{sol}}), \lambda_{1:\mathrm{sol}})x^2 \in \langle x^2 \rangle_{\mathrm{e}} \pm \sigma_2$. Assuming this condition is satisfied (which it indeed will be in the numerical work that follows), we can compute the depth of the resulting explanation by computing the range of λ_1-values over which the empirical constraints continue to hold. We can determine this range through the following steps.

(i) Fix δ to be some small positive real-valued number (small enough so that the following steps do not terminate at the first iteration in k).

(ii) Fix $k = 1$ and compute each of the following quantities:

$$\lambda_1^* \equiv \lambda_{1:\mathrm{sol}} + k\delta,$$
$$\langle x \rangle_{P_1^*} \equiv \int_0^\infty dx \, P_1(x; Z_{1:\mathrm{sol}}(\lambda_1^*), \lambda_1^*) \, x,$$
$$\langle x^2 \rangle_{P_1^*} \equiv \int_0^\infty dx \, P_1(x; Z_{1:\mathrm{sol}}(\lambda_1^*), \lambda_1^*) \, x^2.$$

(iii) If $\langle x \rangle_{P_1^*} \notin \langle x \rangle_{\mathrm{e}} \pm \sigma_1$ and/or $\langle x^2 \rangle_{P_1^*} \notin \langle x^2 \rangle_{\mathrm{e}} \pm \sigma_2$ then the range of admissible values of λ_1 (greater than $\lambda_{1:\mathrm{sol}}$) is delineated by λ_1^* as determined from the last value of k for which the constraints were satisfied.

If the constraints are satisfied at this step, increment k by 1 and repeat steps (ii) and (iii) until at least one of the constraints is indeed violated.

15

(iv) Repeat steps (i) to (iii) starting at $k = -1$ and decrementing k with each iteration to determine the range of admissible values of λ_1 less than $\lambda_{1:\text{sol}}$.

This yields a range, $\delta\lambda_1$, around the value $\lambda_{1:\text{sol}}$, over which the empirical constraints hold. The depth of the explanation is then given by

$$D_{E_1} = \left[1 + \log_{10}\left(\frac{1}{\delta\lambda_1}\right)\right]^{-1}, \tag{53}$$

where a fiducial value of unity for the possible range of λ_1 has been chosen to facilitate comparison.

Let us now move on to the case of E_2. We rewrite the distribution in Eq. (46) in the following way:

$$P_2(x; Z_2, \lambda_1, \lambda_2) = \frac{1}{Z_2}\exp\left(-\lambda_1 x - \lambda_2 x^2\right), \tag{54}$$

where $Z_2 \equiv \exp\left(1 + \lambda_0\right)$. Enforcing the normalization constraint [viz. the first equation in square brackets in Eq. (50)] allows us to consider Z_2 expressly as a function of λ_1 and λ_2, viz. $Z_2 \to Z_{2:\text{sol}}(\lambda_1, \lambda_2)$. In solving for the remaining free parameters in Eq. (54) (viz. in computing $\lambda_{1:\text{sol}}$ and $\lambda_{2:\text{sol}}$) note the theorem by Dowson and Wragg (1973, Theorem 3, p. 691) that states, in our context, that if $\langle x^2 \rangle_\text{e} > 2\langle x \rangle_\text{e}^2$, there is no distribution that maximizes the entropy. So in cases where $\langle x^2 \rangle_\text{e} \leq 2\langle x \rangle_\text{e}^2$ we implement a simple numerical procedure to establish the desired maximum entropy distribution, which we denote by

$$P_2(x; Z_{2:\text{sol}}(\lambda_{1:\text{sol}}, \lambda_{2:\text{sol}}), \lambda_{1:\text{sol}}, \lambda_{2:\text{sol}}) = \frac{1}{Z_{2:\text{sol}}(\lambda_{1:\text{sol}}, \lambda_{2:\text{sol}})}\exp\left(-\lambda_{1:\text{sol}}x - \lambda_{2:\text{sol}}x^2\right). \tag{55}$$

With such a distribution in hand, one can now compute the depth of the resulting explanation of the first two empirically derived moments by separately computing the range of λ_1-values over which the empirical constraints continue to hold (i.e., while holding λ_2 fixed), followed by the range of λ_2-values over which the empirical constraints continue to hold (i.e., holding λ_1 fixed). The procedure we follow mimics, in a straightforward way, the procedure we outlined for E_1, and yields ranges over which the constraints hold, viz. $\delta\lambda_1$ around the value $\lambda_{1:\text{sol}}$, and $\delta\lambda_2$ around the value $\lambda_{2:\text{sol}}$. The depth of the explanation is then given by

$$D_{E_2} = \left[1 + \log_{10}\left(\frac{1}{\delta\lambda_1}\right)\right]^{-1}\left[1 + \log_{10}\left(\frac{1}{\delta\lambda_2}\right)\right]^{-1}, \tag{56}$$

where, again, a fiducial value of unity for the possible range of λ_i (with $i = 1, 2$) has been chosen to facilitate comparison.

The results of our trials are summarized in Table I. We note that for any number of samples M, E_1 provides a deeper explanation than does E_2, and in many cases E_2 does not furnish an explanation at all. Thus the explanation that references the distribution from which the samples are actually drawn is comparatively better—and increasing the number of parameters by a single parameter (as for E_2) *decreases* the depth of the explanation (or, indeed, invalidates it entirely).

IV. DISCUSSION

We have formalized the intuition that an increased level of (global) fine-tuning of observables, and thus (by definition) of fine-tuning of the model that describes those observables, is associated

Number of samples M	Random seed 1		Random seed 2	
	D_{E_1}	D_{E_2}	D_{E_1}	D_{E_2}
10^2	0.578	$+$	0.538	0.312
10^3	0.450	$+$	0.439	0.166
10^4	0.362	0.111	0.347	0.111
10^5	0.312	$+$	0.312	$+$
10^6	0.251	$+$	0.270	$+$
10^7	0.222	$+$	0.228	$+$

TABLE I   Depth of two explanations (D_{E_1} and D_{E_2}) in our maximum entropy toy example. We probe scenarios where the number of samples ranges over five orders of magnitude. For a fixed number of samples, we sample from the exponential distribution in Eq. (44) (in one of two ways, using different random seeds to generate the sequence) and then look to explain statistics of the samples. In particular, we look to explain the first and second moments via one of two explanations, E_1 or E_2, as described in Eqs. (49) and (50), respectively. Results quoted as '+' refer to cases where a maximum entropy distribution (and hence an explanation) does not exist (see the discussion in the main text). In each case, we see that E_1 provides a deeper explanation than E_2, or else E_2 does not furnish an explanation at all. The decrease in the depth of the explanation as the number of samples increases arises because of the manner in which we have constrained errors in the statistics of samples [see Eq. (48)]: larger samples have smaller standard deviations, and thus a smaller target range for the observables (which the underlying parameters help to explain).

with some deficiency in the capacity of the explanation in which such a model partakes. In particular, we have identified this deficiency with a lack of depth of the explanation. In doing so, we have effected the following: (i) a conceptual unification of two theoretical virtues that, we contend, practically guide model choice, viz. the virtue of a lack of fine-tuning and the virtue of a deep explanation and (ii) a quantitative scheme that allows one to compare the depth of explanations with a view to aiding model choice. Our formalism is general in that it can be applied to a broad range of scenarios across the sciences, where established theoretical frameworks exist and also in phenomenological settings where established theoretical frameworks do not yet exist.

Salient technical features of our quantitative scheme are discussed in some detail in Sec. II, but let us highlight some conceptual issues that are captured by our scheme. Ceteris paribus, if the value of some explanandum is significantly affected by the addition of a single extra parameter to the explanans, the depth of the explanation decreases. It is only the case, according to our measure, that the depth of an explanation does not decrease if the value of the explanandum is not significantly affected by values that the extra parameter can take. This is a desirable feature of our scheme for it says, in effect, that the depth of the explanation is not affected by the introduction of parameters in the explanans, with respect to which the explanandum is, in effect, independent.

We suggest that there is a comparative role that can be played by our measure, in that two different explanations for common explananda can be compared. And when one can parse the observables of a model into, e.g., distinct sets of observables, viz. distinct explananda, the depth of the explanations of these explananda can also be compared. In this way, our work highlights a general functional role for explanatory depth, namely, it can be used in an operational setting to discriminate between models or to evaluate how well a single model explains different observables.[6] A further operational role for our measure is as a catalyst for the search for new models. For

[6] On this front, the examples we have selected in this paper refer to modern scientific models, but we expect our approach to work well in other cases from the history of science: for example, in the transition from a geocentric world view, as described by Ptolemy, to subsequent non-geocentric models.

example, if one was to find an upper bound on the depth of an explanation that was significantly less than unity, this would motivate the question of whether there is room for improvement using a different model.

An interesting feature of our account of explanatory depth is its relationship to work by Woodward and Hitchcock (2003) (on explanation) and Hitchcock and Woodward (2003) (on explanatory depth). In short, their approach complements part of our construction while differing in other respects and it will be instructive to develop this comparison. Roughly, they define an explanation through a set of explanans variables, $\{X_1, X_2, \ldots, X_n\}$, that enter into an explanatory generalization, g, which relates the explanandum variable, Y, to the explanans variables, where $Y = g(X_1, X_2, \ldots, X_n)$. In particular, an explanation for the explanandum variable taking the specific value y is an argument in which the explanans variables take some set of values, say $\{x_1, x_2, \ldots, x_n\}$, such that when they enter into the explanatory generalization, one obtains the value y, viz. $y = g(x_1, x_2, \ldots, x_n)$. A defining feature of explanatory generalizations is that they are "invariant under testing interventions" (Woodward and Hitchcock, 2003, p. 17). The invariance refers to a condition in which the explanatory generalization remains intact for counterfactual values of the explanans variables. The intervention is a process that brings about such a change in the values of the explanans variables (without directly affecting the explanandum variable), such as one that might be implemented by an experimenter (but an intervention does not have to invoke human agency). A testing intervention is an intervention that *does not* lead to the same value of the explanandum variable (y in the example above). The depth of an explanation is tied to the "range of invariance" (Hitchcock and Woodward, 2003, p. 182) of the relevant explanatory generalization. A comparison of the depth of two explanations could advert to an assessment of, for example, the relative range of values of explanans variables over which their explanatory generalizations hold (or some comparison related to the topology of that range) or the accuracy of the two explanatory generalizations over their respective ranges of invariance [see Hitchcock and Woodward (2003) for a more complete account].

There is a clear sense in which there is a correspondence between our approach and the one just described. One can interpret our parameters, $\boldsymbol{p}$, as corresponding to the explanans variables while our observables, $\vec{O}$, correspond to (a higher-dimensional version of) the explanandum variable. A difference between these approaches arises in that we focus on particular types of relationships between parameters and observables (viz. what one may interpret as particular types of explanatory generalizations) and, moreover, in developing a notion of explanatory depth, we focus on a specific feature of such relationships. In particular, we focus on relationships that can be described as being more 'law-like', such as those that one commonly finds for dynamical systems (see Sec. III.A) or in certain phenomenological settings (see Sec. III.B). This restriction on the scope of the types of relationships we probe, opens up a novel window on the notion of depth that is not captured, we contend, by their subsequent work on depth [in Hitchcock and Woodward (2003)]. That is, in our account of depth, for any fixed number of parameters (and for a single observable, say), what matters are the ranges of values of these parameters that do not yield significant changes in the value of the observable. In other words, the effective *invariance of the value of the explanandum variable itself*, for counterfactual values of the explanans variables, is a marker of a deep explanation. This sense of depth, we contend, closely follows the types of intuitions that are captured in theoretical developments in physics and more generally across the sciences.

Now, it appears from our definition of the measure of depth [in Eq. (7)], that a marker of a particularly deep explanation is one in which the observables are largely *independent* of the parameters, and so one might wonder why such parameters are to be included in the explanatory scheme. On the other hand, it appears that the inclusion of parameters that are manifestly irrelevant to the system of interest (for example, the inclusion of parameters that track the prices of stocks,

in a description of the evolution of spatial slices of the universe) does not affect the depth of the explanation; and so one might wonder why the inclusion of such parameters doesn't somehow decrease the depth of the explanation. Such objections ignore the context in which our account of explanatory depth has been developed. The parameters that we include play a salient role in the models in which they arise. For a dynamical system, the parameters are either initial conditions for dynamical variables that are necessary to describe the (physical) system of interest or else they are constants that are needed to make (physical) sense of the model, either internally or with respect to other models/theories with which the model is supposed to cohere. For the types of phenomenological models we have explored above, each parameter can be thought of as tracking one of the empirical constraints that the system is deemed to satisfy. Parameters that are not included are those not thought to contribute to our understanding of the system of interest as described by the model. We admit there are choices to be made in leaving out such parameters, but such choices are part of the practice of model building.[7]

More broadly, our account of depth and the account of fine-tuning that it rests on are perhaps best understood as accounts that chart a middle ground between subjective and objective features of model building in the sciences. The objectivity rests primarily on the quantitative nature of the measures we have described; but, again, there is a strong contextual element to our account. One must choose a parameterization within the context of a model and also specify a range over which parameters may vary. Choices must also be made in determining which observables are relevant. So, returning to the example with which we began this paper, in describing the fine-tuning of life, the question arises: the fine-tuning of precisely *what* is supposed to be surprising? [See Maudlin (2017) for an interesting account of this issue.] However, generally, we take this contextual element to be positive feature, as our scheme reflects the types of estimates that enter into the practice and development of science itself.

Appendix A: Observables from probabilistic maps

The schema for depth we have developed in Sec. II can naturally be extended to the case where observables are probabilistically related to parameters (in a way that is distinct from our phenomenological account in Sec. III.B). In particular, in the context of some physical model, $\mathcal{M}$, one can construct a probability density function for the observables given the parameters: $P(\vec{O}|\boldsymbol{p}, \mathcal{M})$. We assume also that we have access to a prior over parameters, viz. $P(\boldsymbol{p}|\mathcal{M})$. Instead of now restricting, to be finite, the domain over which $\boldsymbol{p}$ can take values, we assume, for the sake of simplicity, that $\boldsymbol{p} \in \mathbb{R}^n$ but that the prior, $P(\boldsymbol{p}|\mathcal{M})$, is nonzero only over a finite region. Similarly we assume that in principle, $\vec{O} \in \mathbb{R}^m$, but $P(\vec{O}|\boldsymbol{p}, \mathcal{M})$ is nonzero over some finite region.

---

[7] One goal that such model building (and scientific theorizing more broadly) seems attuned to is the pursuit of parsimonious descriptions of previously collected data, with a view to accurate future predictions. This was a major accomplishment of quantum chromodynamics, which explained a large amount of phenomenological data on nuclear physics; and quantum electrodynamics, which achieved similar successes for electromagnetic phenomena. This goal also underlies the effort to find a theory that unifies all the forces of nature.

In this case therefore, an explanation of the vector of observables taking the value $\vec{O}_M$ (more specifically, taking a value in some small m-dimensional box, $\Delta_{\vec{O}_M}$, centered on $\vec{O}_M$) will comprise an argument in which the probability of the observables taking these values is larger than some fiducial value μ.[8] We can represent the resulting argument that comprises the explanation in the following summarized form:

$$\bar{E} : \boldsymbol{p}' \wedge [P(\vec{O}_M|\boldsymbol{p}', \mathcal{M})\Delta_{\vec{O}_M} > \mu] \therefore \vec{O}_M. \tag{A1}$$

The depth of this explanation can be computed via the following steps.

(i) For the ith parameter direction about the point $\boldsymbol{p}'$, we construct the range, δ_i, of parameter values over which observables do not change significantly. That is, the range over which the probability that the vector of observables takes values in the same small m-dimensional box described above, centered on $\vec{O}_M$, remains high: $P(\vec{O}_M|\boldsymbol{p}, \mathcal{M})\Delta_{\vec{O}_M} > \mu$.[9]

(ii) Next we find the probability of parameters lying in this range as gleaned from the appropriate marginal distribution:

$$P_i(\delta_i) \equiv \left[\prod_{k \neq i} \int_{\mathbb{R}} dp_k\right] \int_{\delta_i} dp_i\, P(\boldsymbol{p}|\mathcal{M}). \tag{A2}$$

(iii) Finally we define a corresponding measure of global fine-tuning, $\bar{\mathcal{G}}_i(\vec{O}; \boldsymbol{p}')$, extending the treatment in Sec. II and in Azhar and Loeb (2018):

$$\bar{\mathcal{G}}_i(\vec{O}; \boldsymbol{p}') \equiv \log_{10}\left(\frac{1}{P_i(\delta_i)}\right). \tag{A3}$$

This measure is manifestly non-negative, with the minimum value (viz. zero) occurring when the probability of the observables lying in $\Delta_{\vec{O}_M}$, centered on $\vec{O}_M$, is greater than the fiducial value μ, independent of the value of the ith parameter.

Our measure of the depth of the explanation in Eq. (A1), which we denote by $\bar{D}_{\bar{E}}(\vec{O}; \boldsymbol{p}')$ is then defined by the following:

$$\bar{D}_{\bar{E}}(\vec{O}; \boldsymbol{p}') \equiv \frac{1}{\displaystyle\prod_{i=1}^{n}\left[1 + \bar{\mathcal{G}}_i(\vec{O}; \boldsymbol{p}')\right]}. \tag{A4}$$

This measure has analogous features to the measure described in Sec. II [see points (i)–(v) under Eq. (7)] with appropriate reassignments, e.g., $\mathcal{G} \to \bar{\mathcal{G}}$ and $D_E \to \bar{D}_{\bar{E}}$.

---

[8] Of course, we need to make precise how one determines this value. The variable μ is a stand-in for a 'high value' and precisely what this value is (or should be) is a determination best made in the specific context in which an explanation is being constructed. We will not need to specify such a value to describe our scheme.

[9] Here we have assumed a small shift in the point in parameter space determined by a shift solely in the ith direction. We have represented the new point thus obtained by $\boldsymbol{p}$ instead of $\boldsymbol{p}'$. By convention, limits of the range δ_i are determined by the *first* point in parameter space such that the probability of observables lying in $\Delta_{\vec{O}_M}$ is less than or equal to μ.

References

Azhar, F. and Butterfield, J. (2018). "Scientific Realism and Primordial Cosmology". In: *The Routledge Handbook on Scientific Realism*, ed. J. Saatsi. London: Routledge, pp. 304–320. PhilSci-Archive: 12192.

Azhar, F. and Loeb, A. (2018). Gauging fine-tuning. *Physical Review* D **98**: 103018. PhilSci-Archive: 15036.

Brawer, R. (1996). *Inflationary Cosmology and the Horizon and Flatness Problems: The Mutual Constitution of Explanation and Questions*. M.Sc. Dissertation, Massachusetts Institute of Technology, MA, USA.

Cover, T. M. and Thomas, J. A. (1991). *Elements of Information Theory*. New York: Wiley.

Dowson, D. C. and Wragg, A. (1973). Maximum-entropy distributions having prescribed first and second moments. *IEEE Transactions on Information Theory* **19**: 689.

Guth, A. H. (1981). Inflationary universe: A possible solution to the horizon and flatness problems. *Physical Review* D **23**: 347.

Hempel, C. G. and Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science* **15**: 135.

Hitchcock, C. and Woodward, J. (2003). Explanatory generalizations, Part II: Plumbing explanatory depth. *Noûs* **37**: 181.

Keas, M. N. (2018). Systematizing the theoretical virtues. *Synthese* **195**: 2761.

Kitcher, P. and Salmon, W. C. (eds.) (1989). *Scientific Explanation*. Minnesota studies in the philosophy of science, Volume 13. Minneapolis: University of Minnesota Press.

Kuhn, T. S. (1977). "Objectivity, Value Judgment, and Theory Choice". In: *The Essential Tension: Selected Studies in Scientific Tradition and Change*. Chicago: University of Chicago Press, pp. 320–339.

Maudlin T. (2017). Fine-tuned for what? Talk at the *International Conference on the Physics of Fine-Tuning*, Accessed: October 10, 2019.

McMullin, E. (1982). "Values in Science". In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association, Vol. 1982*, Volume Two: Symposia and Invited Papers, Chicago: University of Chicago Press, pp. 3–28.

Planck Collaboration (2018). *Planck* 2018 results. VI. Cosmological parameters. arXiv e-prints:1807.06209 [astro-ph.CO].

Schindler, S. (2018). *Theoretical Virtues in Science: Uncovering Reality through Theory*. Cambridge: Cambridge University Press.

Tkačik, G. and Bialek, W. (2016). Information processing in living systems. *Annual Review of Condensed Matter Physics* **7**: 89. arXiv e-prints:1412.8752 [q-bio.QM].

Weslake, B. (2010). Explanatory depth. *Philosophy of Science* **77**: 273.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

Woodward, J. and Hitchcock, C. (2003). Explanatory generalizations, Part I: A counterfactual account. *Noûs* **37**: 1.

Ylikoski, P. and Kuorikoski, J. (2010). Dissecting explanatory power. *Philosophical Studies* **148**: 201.