Philosophy of Cosmology

Chris Smeenk

1 Introduction

Cosmology has made enormous progress in the last several decades. It is no longer a neglected subfield of physics, as it was as recently as 1960; it is instead an active area of fundamental research that can boast of a Standard Model well-supported by observations. Prior to 1965 research in cosmology had a strikingly philosophical tone, with debates focusing explicitly on scientific method and the aims and scope of cosmology (see, e.g. Munitz 1962; North 1965; Kragh 1996). One might suspect that with the maturation of the field these questions have been settled, leaving little room for philosophers to contribute. Although the nature of the field has changed dramatically with an increase of observational knowledge and theoretical sophistication, there are still ongoing foundational debates regarding cosmology's proper aims and methods. Cosmology confronts a number of questions dear to the hearts of philosophers of science: the limits of scientific explanation, the nature of physical laws, and different types of underdetermination, for example. There is an opportunity for philosophers to make fruitful contributions to debates in cosmology and to consider the ramifications of new ideas in cosmology for other areas of philosophy and foundations of physics.

Due to the uniqueness of the universe and its inaccessibility, cosmology has often been characterized as "un-scientific" or inherently more speculative than other parts of physics. How can one formulate a scientific theory of the "universe as a whole"? Even those who reject skepticism regarding cosmology often assert instead that cosmology can only make progress by employing a distinctive methodology. These discussions, in my view, have by and large failed to identify the source and the extent of the evidential challenges faced by cosmologists. There are no convincing, general no-go arguments showing the impossibility of secure knowledge in cosmology; there are instead specific problems that arise in attempting to gain observational and theoretical access to the universe. In some cases, cosmologists have achieved knowledge as secure as that in other areas of physics — arguably, for example, in the account of big bang nucleosynthesis.

Cosmologists do, however, face a number of distinctive challenges. The finitude of the speed of light, a basic feature of relativistic cosmology, insures that global properties of the universe cannot be established directly by observations (§5). This is a straightforward limit on observational access to the universe, but there are other obstacles of a different kind. Cosmology relies on extrapolating local physical laws to hold universally. These extrapolations make it possible to infer, from observations of standard candles such as Type Ia supernovae,[1] the startling conclusion

[1] A "standard candle" is an object whose intrinsic luminosity can be determined; the observed apparent magnitude then provides an accurate measurement of the distance to the object.

that the universe includes a vast amount of dark matter and dark energy. Yet the inference relies on extrapolating general relativity, and the observations may reveal the need for a new gravitational theory rather than new types of matter. It is difficult to adjudicate this debate due to the lack of independent access to the phenomena (§3). The early universe (§6) is interesting because it is one of the few testing grounds for quantum gravity. Without a clear understanding of the initial state derived from such a theory, however, it is difficult to use observations to infer the dynamics governing the earliest stages of the universe's evolution. Finally, it is not clear how to take the selection effect of our presence as observers into account in assessing evidence for cosmological theories (§7).

These challenges derive from distinctive features of cosmology. One such feature is the interplay between global aspects of the universe and local dynamical laws. The Standard Model of cosmology is based on extrapolating local laws to the universe as a whole. Yet, there may be global-to-local constraints. The uniqueness of the universe implies that the normal ways of thinking about laws of physics and the contrast between laws and initial conditions do not apply straightforwardly (§4). In other areas of physics, the initial or boundary conditions themselves are typically used to explain other things rather than being the target of explanation. Many lines of research in contemporary cosmology aim to explain why the initial state of the Standard Model obtained, but the nature of this explanatory project is not entirely clear. And due to the uniqueness of the universe and the possibility of anthropic selection effects it is not clear what underwrites the assignment of probabilities.

What follows is not a survey of a thoroughly explored field in philosophy of physics. There are a variety of topics in this area that philosophers could fruitfully study, but as of yet the potential for philosophical work has not been fully realized.[2] The leading contributions have come primarily from cosmologists who have turned to philosophical considerations arising from their work. The literature has a number of detailed discussions of specific issues, but there are few attempts at a more systematic approach. As a result, this essay is an idiosyncratic tour of various topics and arguments rather than a survey of a well-charted intellectual landscape. It is also a limited tour, and leaves out a variety of important issues — most significantly, the impact of quantum mechanics on issues ranging from the origin of density perturbations in the early universe to the possible connections between Everettian and cosmological multiverses. But I hope that despite these limitations, this survey may nonetheless encourage other philosophers to actualize the potential for contributions to foundational debates within cosmology.

2 Overview of the Standard Model

Since the early 70s cosmology has been based on what Weinberg (1972) dubbed the "Standard Model." This model describes the universe's spacetime geometry, material constituents, and their dynamical evolution. The Standard Model is based on extending local physics — including general

[2]This is not to say that there is no literature on the topic, and much of it will be cited in below. For more systematic reviews of the literature by someone whose contributions have shaped the field, and which I draw on in the following, see Ellis (1999, 2007).

relativity, quantum physics, and statistical physics — to cosmological scales and to the universe as a whole. A satisfactory cosmological model should be sufficiently rich to allow one to fix basic observational relations, and to account for various striking features of the universe, such as the existence of structures like stars and galaxies, as consequences of the underlying dynamics.

The Standard Model describes the universe as starting from an extremely high temperature early state (the "big bang") and then expanding, cooling, and developing structures such as stars and galaxies. At the largest scales the universe's spacetime geometry is represented by the expanding universe models of general relativity. The early universe is assumed to begin with matter and radiation in local thermal equilibrium, with the stress-energy dominated by photons. As the universe expands, different types of particles "freeze out" of equilibrium, leaving an observable signature of earlier stages of evolution. Large-scale structures in the universe, such as galaxies and clusters of galaxies, arise later via gravitational clumping from initial "seeds." Here I will give a brief sketch of the Standard Model to provide the necessary background for the ensuing discussion.[3]

2.1 Expanding Universe Models

Einstein (1917) introduced a strikingly new conception of cosmology, as the study of exact solutions of general relativity that describe the spacetime geometry of the universe. One would expect gravity to be the dominant force in shaping the universe's structure at large scales, and it is natural to look for solutions of Einstein's field equations (EFE) compatible with astronomical observations. Einstein's own motivation for taking the first step in relativistic cosmology was to vindicate Mach's principle.[4] He also sought a solution that describes a static universe, that is, one whose spatial geometry is unchanging. He forced his theory to accommodate a static model by modifying his original field equations, with the addition of the infamous cosmological constant Λ. As a result Einstein missed one of the most profound implications of his new theory: general relativity quite naturally implies that the universe evolves dynamically with time. Four of Einstein's contemporaries discovered a class of simple evolving models, the Friedman-Lemaître-Robertson-Walker (FLRW) models, that have proven remarkably useful in representing the spacetime geometry of our universe.

These models follow from symmetry assumptions that dramatically simplify the task of solving EFE. They require that the spacetime geometry is both *homogeneous* and *isotropic*; this is also called imposing the "cosmological principle." Roughly speaking, homogeneity requires that at a given moment of cosmic time every spatial point "looks the same," and isotropy holds if there are no geometrically preferred spatial directions. These requirements imply that the models are topologically $\Sigma \times \mathfrak{R}$, visualizable as a "stack" of three-dimensional spatial surfaces $\Sigma(t)$ labeled by values of the cosmic time t. The worldlines of "fundamental observers," taken to be at rest with respect to matter, are orthogonal to these surfaces, and the cosmic time corresponds to the proper

[3]Of the several textbooks that cover this territory, see in particular Peebles (1993); Dodelson (2003); Weinberg (2008); see Longair (2006) for a masterful historical survey of the development of cosmology and astrophysics.

[4]At the time, Einstein formulated Mach's principle as the requirement that inertia derives from interactions with other bodies rather than from a fixed background spacetime. His model eliminated the need for anti-Machian boundary conditions by eliminating boundaries: it describes a universe with spatial sections of finite volume, without edges. See Smeenk (2012) for further discussion.

time measured by the fundamental observers. The spatial geometry of Σ is such that there is an isometry carrying any point $p \in \Sigma$ to any other point lying in the same surface (homogeneity), and at any point p the three spatial directions are isometric (isotropy).[5]

The cosmological principle tightly constrains the properties of the surfaces $\Sigma(t)$. These are three-dimensional spaces (Riemannian manifolds) of constant curvature, and all of the surfaces in a given solution have the same topology. If the surfaces are simply connected, there are only three possibilities for Σ: (1) spherical space, for the case of positive curvature; (2) Euclidean space, for zero curvature; and (3) hyperbolic space, for negative curvature.[6] Textbook treatments often neglect to mention, however, that replacing *global* isotropy and homogeneity with *local* analogs opens the door to a number of other possibilities. For example, there are models in which the surfaces Σ have finite volume but are multiply connected, consisting of, roughly speaking, cells pasted together.[7] Although isotropy and homogeneity hold locally at each point, above some length scale there would be geometrically preferred directions reflecting how the cells are connected. In these models it is in principle possible to see "around the universe" and observe multiple images of a single object, but there is at present no strong observational evidence of such effects.

Imposing global isotropy and homogeneity reduces EFE — a set of 10 non-linear, coupled partial differential equations — to a pair of differential equations governing the scale factor $R(t)$ and $\rho(t)$, the energy density of matter. The scale factor measures the changing spatial distance between fundamental observers. The dynamics are then captured by the Friedmann equation:[8]

$$\left(\frac{\dot{R}}{R}\right)^2 = \frac{8\pi G\rho}{3} - \frac{k}{R^2} + \frac{\Lambda}{3}, \tag{1}$$

and (a special case of) the Raychaudhuri equation:

$$3\frac{\ddot{R}}{R} = -4\pi G\left(\rho + 3p\right) + \Lambda. \tag{2}$$

[5]An isometry is a transformation that preserves the spacetime geometry; more precisely, a diffeomorphism ϕ that leaves the spacetime metric invariant, i.e. $(\phi^* g)_{ab} = g_{ab}$.

[6]A topological space is *simply connected* if, roughly speaking, every closed loop can be smoothly contracted to a point. For example, the surface of a bagel is multiply connected, as there are two different types of loops that cannot be continuous deformed to a point. There is another possibility for a globally isotropic space with constant positive curvature that is multiply connected, namely projective space (with the same metric as spherical space but a different topology). These three possibilities are unique up to isometry. See, e.g., Wolf (2011), for a detailed discussion.

[7]See Ellis (1971) for a pioneering study of this kind of model, and Lachieze-Rey and Luminet (1995) for a more recent review.

[8]EFE are: $G_{ab} + \Lambda g_{ab} = 8\pi T_{ab}$, where G_{ab} is the Einstein tensor, T_{ab} is the stress-energy tensor, g_{ab} is the metric, and Λ is the cosmological constant. Equation (1) follows from the "time-time" component of EFE, and equation (2) is the difference between it and the "space-space" component. (All other components vanish due to the symmetries.) The Raychaudhuri equation is a fundamental equation that describes the evolution of a cluster of nearby worldlines, e.g. for the particles making up a small ball of dust, in response to curvature. It takes on the simple form given here due to the symmetries we have assumed: in the FLRW models the small ball of dust can change only its volume as a function of time, but in general there can be a volume-preserving distortion (shear) and torsion (rotation) of the ball as well.

$\dot{R}$ means differentiation with respect to the cosmic time t, G is Newton's gravitational constant, and Λ is the cosmological constant. The curvature of surfaces $\Sigma(t)$ of constant cosmic time is given by $\frac{k}{R^2(t)}$, where $k = \{-1, 0, 1\}$ for negative, flat, and positive curvature (respectively). The assumed symmetries force the matter to be described as a perfect fluid with energy density ρ and pressure p.[9] The energy density and pressure are given by the equation of state for different kinds of perfect fluids; for example, for "pressureless dust" $p = 0$, whereas for radiation $p = \rho/3$. Given a specification of the matter content, there exist unique solutions for the scale factor $R(t)$ and the energy density $\rho(t)$ for each type of matter included in the model.

Several features of the dynamics of these models are clear from inspection of these equations. Suppose we take "ordinary" matter to always have positive total stress-energy density, in the sense of requiring that $\rho + 3p > 0$. Then, from (2), it is clear that the effect of such ordinary matter is to decelerate cosmic expansion, $\ddot{R} < 0$ — reflecting the familiar fact that gravity is a force of attraction. But this is only so for ordinary matter. A positive cosmological constant (or matter with negative stress-energy) leads, conversely, to accelerating expansion, $\ddot{R} > 0$. Einstein satisfied his preference for a static model by choosing a value of Λ that precisely balances the effect of ordinary matter, such that $\ddot{R} = 0$. But his solution is unstable, in that a slight concentration (deficit) of ordinary matter triggers run-away contraction (expansion). It is difficult to avoid dynamically evolving cosmological models in general relativity.

Restricting consideration to ordinary matter and setting $\Lambda = 0$, the solutions fall into three types depending on the relative magnitude of two terms on the right hand side of eqn. 1, representing the effects of energy density and curvature. For the case of flat spatial geometry $k = 0$, the energy density takes exactly the value needed to counteract the initial velocity of expansion such that $\dot{R} \to 0$ as $t \to \infty$. This solution separates the two other classes: if the energy density is greater than critical, there is sufficient gravitational attraction to reverse the initial expansion, and the spatial slices Σ have spherical geometry ($k = +1$); if the energy density is less than critical, the sign of $\dot{R}$ never changes, expansion never stops, and the spatial slices have hyperbolic geometry ($k = -1$).[10] This simple picture does not hold if $\Lambda \neq 0$, as the behavior then depends on the relative magnitude of the cosmological constant term and ordinary matter.

The equations above lead to simple solutions for $R(t)$ for models including a single type of matter: for electromagnetic radiation, $R(t) \propto t^{1/2}$; for pressureless dust $R(t) \propto t^{2/3}$; and for a cosmological constant, $R(t) \propto e^t$.[11] Obviously, more realistic models include several types of matter. The energy density for different types of matter dilutes with expansion at different rates: pressureless dust – $\rho(t) \propto R^{-3}$; radiation – $\rho(t) \propto R^{-4}$; and a cosmological constant remains constant. As a result of these different dilution rates, a complicated model can be treated in terms of a sequence of

[9] The stress energy tensor for a perfect fluid is given by $T_{ab} = (\rho + p)\xi_a\xi_b + (p)g_{ab}$, where ξ_a is the tangent vector to the trajectories of the fluid elements.

[10] These are both classes of solutions, where members of the class have spatial sections with curvature of the same sign but different values of the spatial curvature at a given cosmic time.

[11] One can treat the cosmological constant as a distinctive type of matter, in effect moving it from the left to the right side of EFE and treating it as a component of the stress-energy tensor. It can be viewed instead as properly included on the left-hand side as part of the spacetime geometry. This issue of interpretation does not, however, make a difference with regard to the behavior of the solution.

simple models describing the effects of the dominant type of matter on cosmic evolution. At $t \approx 1$ second, the Standard Model describes the universe as filled with matter and radiation, where the latter initially has much higher energy density. Because the energy density of radiation dilutes more rapidly than that of matter, the initial radiation-dominated phase is followed by a matter-dominated phase that extends until the present. Current observations indicate the presence of "dark energy" (discussed in more detail below) with properties like a Λ term. Supposing these are correct, in the future the universe will eventually transition to a dark-energy-dominated phase of exponential expansion, given that the energy density of a Λ term does not dilute *at all* with expansion.

FLRW models with ordinary matter have a singularity at a finite time in the past. Extrapolating back in time, given that the universe is currently expanding, eqn. (2) implies that the expansion began at some finite time in the past. The current rate of expansion is given by the Hubble parameter, $H = \frac{\dot{R}}{R}$. Simply extrapolating this expansion rate backward, $R(t) \to 0$ at the Hubble time H^{-1}; from eqn. (2) the expansion rate must increase at earlier times, so $R(t) \to 0$ at a time less than the Hubble time before now. As this "big bang" is approached, the energy density and curvature increase without bound. This reflects the instability of evolution governed by EFE: as $R(t)$ decreases, the energy density and pressure both increase, and they both appear with the same sign on the right hand side of eqn. (2). It was initially hoped that the singularity could be avoided in more realistic models that are not perfectly homogeneous and isotropic, but Penrose, Hawking, and Geroch showed in the 1960s that singularities hold quite generically in models suitable for cosmology. It is essential for this line of argument that the model includes ordinary matter and no cosmological constant; since the Λ term appears in eqn. (2) with the opposite sign, one can avoid the initial singularity by including a cosmological constant (or matter with a negative stress-energy).

One of the most remarkable discoveries in twentieth century astronomy was Hubble's (1929) observation that the red-shifts of spectral lines in galaxies increase linearly with their distance.[12] Hubble took this to show that the universe is expanding uniformly, and this effect can be given a straightforward qualitative explanation in the FLRW models. The FLRW models predict a change in frequency of light from distant objects that depends directly on $R(t)$.[13] There is an approximately linear relationship between red-shift and distance at small scales for all the FLRW models, and departures from linearity at larger scales can be used to measure spatial curvature.

At the length scales of galaxies and clusters of galaxies, the universe is anything but homogeneous and isotropic, and the use of the FLRW models involves a (usually implicit) claim that above some length scale the average matter distribution is sufficiently uniform. By hypothesis the models

[12] Hubble's distance estimates have since been rejected, leading to a drastic decrease in the estimate of the current rate of expansion (the Hubble parameter, H_0). However, the linear redshift-distance relation has withstood scrutiny as the sample size has increased from 24 bright galaxies (in Hubble 1929) to hundreds of galaxies at distances 100 times greater than Hubble's, and as astrophysicists have developed other observational methods for testing the relation (see Peebles 1993, 82 - 93 for an overview).

[13] The problem is underspecified without some stipulation regarding the worldlines traversed by the observers emitting and receiving the signal. Assuming that both observers are fundamental observers, a photon with frequency ω emitted at a cosmic time t_1 will be measured to have a frequency $\omega' = \frac{R(t_1)}{R(t_2)}\omega$ at a later time t_2. (For an expanding universe, this leads to a red-shift of the light emitted.) Given a particular solution one can calculate the exact relationship between spectral shift and distance.

do not describe the formation and evolution of inhomogeneities that give rise to galaxies and other structures. Prior to 1965, the use of the models was typically justified on the grounds of mathematical utility or an argument in favor of the cosmological principle, with no expectation that the models were in more than qualitative agreement with observations — especially when extrapolated to early times. The situation changed dramatically with the discovery that the FLRW models provide an extremely accurate description of the early universe, as revealed by the uniformity of the cosmic background radiation (CBR, described below). The need to explain why the universe is so strikingly symmetric was a driving force for research in early universe cosmology (see §6 below).

2.2 Thermal History

Alvy Singer's mother in *Annie Hall* is right: Brooklyn is not expanding. But this is not because the cosmic expansion is not real or has no physical effects. Rather, in the case of gravitationally bound systems such as the Earth or the solar system the effects of cosmic expansion are far, far too small to detect.[14] In many domains the cosmological expansion can be ignored. The dynamical effects of expansion are, however, the central theme in the Standard Model's account of the thermal history of the early universe.

Consider a given volume of the universe at an early time, filled with matter and radiation assumed to be initially in local thermal equilibrium.[15] The dynamical effects of the evolution of $R(t)$ are locally the same as slowly varying the volume of this region, imagining that the matter and radiation are enclosed in a box that expands (or contracts) adiabatically. For some stages of evolution the contents of the box interact on a sufficiently short timescale that equilibrium is maintained through the change of volume, which then approximates a quasi-static process. When the interaction timescale becomes greater than the expansion timescale, however, the volume changes too fast for the interaction to maintain equilibrium. This leads to a departure from equilibrium; particle species "freeze out" and decouple, and entropy increases. Without a series of departures from equilibrium, cosmology would be a boring subject — the system would remain in equilibrium with a state determined solely by the temperature, without a trace of things past. Departures from equilibrium are of central importance in understanding the universe's thermal history.[16]

[14] Quantitatively estimating the dynamical effects of the expansion on local systems is remarkably difficult. One approach is, schematically, to imbed a solution for a local system (such as a Schwarzschild solution) into an FLRW spacetime, taking care to impose appropriate junction conditions on the boundary. One can then calculate an upper bound on the effect of the cosmological expansion; the effect will presumably be smaller in a more realistic model, which includes a hierarchy of imbedded solutions representing structures at larger length scales such as the galaxy and the Local Group of galaxies. Because of the non-linearity of EFE it is surprisingly subtle to make the idea of a "quasi-isolated" system immersed in a background cosmological model precise, and to differentiate effects due to the expansion from those due to changes within the local system (such as growing inhomogeneity). See Carrera and Giulini (2010) for a recent systematic treatment of these issues.

[15] This assumption of local thermal equilibrium as an "initial state" at a given time presumes that the interaction time scales are much less than the expansion time scale at earlier times.

[16] The departures from equilibrium are described using the Boltzmann equation. The Boltzmann equation formulated in an FLRW spacetime includes an expansion term. As long as the collision term (for some collection of interacting particles) dominates over the expansion term then the interactions are sufficient to maintain equilibrium, but as the universe cools, the collision term becomes subdominant to the expansion term, and the particles decouple

Two particularly important cases are big bang nucleosynthesis and the decoupling of radiation from matter. The Standard Model describes the synthesis of light elements as occurring during a burst of nuclear interactions that transpire as the universe falls from a temperature of roughly 10^9 K, at $t \approx 3$ minutes, to 10^8 K, at about 20 minutes.[17] Prior to this interval, any deuterium formed by combining protons and neutrons is photodissociated before heavier nuclei can build up; whereas after this interval, the temperature is too low to overcome the Coulomb barriers between the colliding nuclei. But during this interval the deuterium nuclei exist long enough to serve as seeds for formation of heavier nuclei because they can capture other nucleons. Calculating the primordial abundances of light elements starts from an initial "soup" at $t \approx 1$ second, including neutrons, protons, electrons and photons in local thermal equilibrium.[18] Given experimentally measured values of the relevant reaction rates, one can calculate the change in relative abundances of these constituents, and the appearance of nuclei of the light elements. The result of these calculations is a prediction of light-element abundances that depends on physical features of the universe at this time, such as the total density of baryonic matter and the baryon to photon ratio. Observations of primordial element abundances can then be taken as constraining the cosmological model's parameters. Although there are still discrepancies (notably regarding Lithium 7) whose significance is unclear, the values of the parameters inferred from primordial abundances in conjunction with nucleosynthesis calculations are in rough agreement with values determined from other types of observations.

As the temperature drops below $\approx 4,000$K, "re-combination" occurs as the electrons become bound in stable atoms.[19] As a result, the rate of one of the reactions keeping the photons and matter in equilibrium (Compton scattering of photons off electrons) drops below the expansion rate. The photons decouple from the matter with a black-body spectrum. After decoupling, the photons cool adiabatically with the expansion, and the temperature drops as $T \propto 1/R$, but the black-body spectrum is unaffected. This "cosmic background radiation" (CBR) carries an enormous amount of information regarding the universe at the time of decoupling. It is difficult to provide a natural, alternative explanation for the black-body spectrum of this radiation.[20]

---

from the plasma and fall out of equilibrium. To find the number density at the end of this freezing out process, one typically has to solve a differential equation (or a coupled set of differential equations for multiple particle species) derived from the Boltzmann equation.

[17] See Olive et al. (2000) for a review of big bang nucleosynthesis.

[18] These are called "primordial" or "relic" abundances to emphasize that they are the abundances calculated to hold at $t \approx 20$ minutes. Inferring the values of these primordial abundances from observations requires an understanding of the impact of subsequent physical processes, and the details differ substantially for the various light elements.

[19] The term "re-combination" is misleading, as the electrons were not previously bound in stable atoms. See Weinberg (2008), §2.3 for a description of the intricate physics of recombination.

[20] The black-body nature of the spectrum was firmly established by the COBE (Cosmic Background Explorer) mission in 1992. The difficulty in finding an alternative stems from the fact that the present universe is almost entirely transparent to the CBR photons, and the matter that does absorb and emit radiation is not distributed uniformly. To produce a uniform sea of photons with a black-body spectrum, one would need to introduce an almost uniformly distributed type of matter that thermalizes radiation from other processes to produce the observed microwave background, yet is nearly transparent at other frequencies. Advocates of the quasi-steady state cosmology have argued that whiskers of iron ejected from supernovae could serve as just such a thermalizer of radiation in the far infrared. See, e.g., Li (2003) for a discussion of this proposal and persuasive objections to it.

The initial detection of the CBR and subsequent measurements of its properties played a crucial role in convincing physicists to trust the extrapolations of physics to these early times, and ever since its discovery, the CBR has been a target for increasingly sophisticated observational programs. These observations established that the CBR has a uniform temperature to within 1 part in 10^5, and the minute fluctuations in temperature provide empirical guidance for the development of early universe theories.

In closing, two aspects of the accounts of the thermal history deserve emphasis. First, the physics used in developing these ideas has independent empirical credentials. Although the very idea of early universe cosmology was regarded as speculative when calculations of this sort were first performed (Alpher, Bethe and Gamow 1948), the basic nuclear physics was not. Second, treating the constituents of the early universe as being in local thermal equilibrium before things get interesting is justified provided that the reaction rates are higher than the expansion rate at earlier times. This is an appealing feature since equilibrium has the effect of washing away dependence on earlier states of the universe. As a result processes such as nucleosynthesis are relatively insensitive to the state of the very early universe. The dynamical evolution through nucleosynthesis is based on well-understood nuclear physics, and equilibrium effaces the unknown physics at higher energies.

2.3 Structure Formation

By contrast with these successes, the Standard Model lacks a compelling account of how structures like galaxies formed. This reflects the difficulty of the subject, which requires integrating a broader array of physical ideas than those required for the study of nucleosynthesis or the FLRW models. It also requires more sophisticated mathematics and computer simulations to study dynamical evolution beyond simple linear perturbation theory.

Newtonian gravity enhances clumping of a nearly uniform distribution of matter, as matter is attracted more strongly to regions with above average density. Jeans (1902) studied the growth of fluctuations in Newtonian gravity, and found that fluctuations with a mode greater than a critical length exhibit instability, and their amplitude grows exponentially. The first study of a similar situation in general relativity (Lifshitz 1946) showed, by contrast, that expansion in the FLRW models counteracts this instability, leading to much slower growth of initial perturbations. Lifshitz (1946) concluded that the gravitational enhancement picture could not produce galaxies from plausible "seed" perturbations, and rejected it. Two decades later the argument was reversed: given the gravitational enhancement account of structure formation (no viable alternative accounts had been discovered), the seed perturbations had to be much larger than Lifshitz expected. Many cosmologists adopted a more phenomenological approach, using observational data to constrain the initial perturbation spectrum and other parameters of the model.

Contemporary accounts of structure formation treat observed large-scale structures as evolving by gravitational enhancement from initial seed perturbations. The goal is to account for observed properties of structures at a variety of scales — from features of galaxies to statistical properties of the large scale distribution of galaxies — by appeal to the dynamical evolution of the seed perturbations through different physical regimes. Harrison, Peebles, and Zel'dovich independently argued that the initial perturbations should be scale invariant, that is, lacking any characteristic length scale.[21] Assuming that these initial fluctuations are small (with a density contrast $\frac{\delta\rho}{\rho} << 1$), they can be treated as linear perturbations to a background cosmological model where the dynamical evolution of individual modes specified by general relativity. As the perturbations grow in amplitude and reach $\frac{\delta\rho}{\rho} \approx 1$, perturbation theory no longer applies and the perturbation mode "separates" from cosmological expansion and begins to collapse. In current models, structure grows hierarchically with smaller length scales going non-linear first. Models of evolution of structures at smaller length scales (e.g. the length scales of galaxies) as the perturbations go non-linear incorporate physics in addition to general relativity, such as gas dynamics, to describe the collapsing clump of baryonic matter.

The current consensus regarding structure formation is called the ΛCDM model. The name indicates that the model includes a non-zero cosmological constant (Λ) and "cold" dark matter (CDM). (Cold dark matter is discussed in the next section.) The model has several free parameters that can be constrained by measurements of a wide variety of phenomena. The richness of these evidential constraints and their mutual compatibility provide some confidence that the ΛCDM model is at least partially correct. There are, however, ongoing debates regarding the status of the model. For example, arguably it does not capture various aspects of galaxy phenomenology. Although I do not have the space to review these debates here, it is clear that current accounts of structure formation face more unresolved challenges and problems than other aspects of the Standard Model.

3 Dark Matter and Dark Energy

The main support for the Standard Model comes from its successful accounts of big-bang nucleosynthesis, the redshift-distance relation, and the CBR. But pushing these lines of evidence further reveals that, if the Standard Model is basically correct, the vast majority of the matter and energy filling the universe cannot be ordinary matter. According to the "concordance model," normal matter contributes ≈ 4% of the total energy density, with ≈ 22% in the form of non-baryonic dark matter and another ≈ 74% in the form of dark energy.[22]

Dark matter was first proposed based on observations of galaxy clusters and galaxies.[23] Their dynamical behavior cannot be accounted for solely by luminous matter in conjunction with Newtonian gravity. More recently, it was discovered that the deuterium abundance, calculated from big bang nucleosynthesis, puts a tight bound on the total amount of baryonic matter. Combining this constraint from big bang nucleosynthesis with other estimates of cosmological parameters

[21] More precisely, the different perturbation modes have the same density contrast when their wavelength equals the Hubble radius, H^{-1}.

[22] Cosmologists use "concordance model" to refer to the Standard Model of cosmology with the specified contributions of different types of matter. The case in favor of a model with roughly these contributions to the overall energy density was made well before the discovery of cosmic acceleration (see, e.g., Ostriker and Steinhardt (1995); Krauss and Turner (1999)). Coles and Ellis (1997) give a useful summary of the opposing arguments (in favor of a model without a dark energy component) as of 1997, and see Frieman et al. (2008) for a more recent review.

[23] See Trimble (1987) for a discussion of the history of the subject and a systematic review of various lines of evidence for dark matter.

leads to the conclusion that there must be a substantial amount of non-baryonic dark matter. Accounts of structure formation via gravitational enhancement also seem to require non-baryonic cold dark matter. Adding "cold" dark matter to models of structure formation helps to reconcile the uniformity of the CBR with the subsequent formation of structure. The CBR indicates that any type of matter coupled to the radiation must have been very smooth, much too smooth to provide seeds for structure formation. Cold dark matter decouples from the baryonic matter and radiation early, leaving a minimal imprint on the CBR.[24] After recombination, however, the cold dark matter perturbations generate perturbations in the baryonic matter sufficiently large to seed structure formation.

The first hint of what is now called "dark energy" also came in studies of structure formation, which seemed to require a non-zero cosmological constant to fit observational constraints (the ΛCDM models). Subsequent observations of the redshift-distance relation, with supernovae (type Ia) used as a powerful new standard candle, led to the discovery in 1998 that the expansion of the universe is *accelerating*.[25] This further indicates the need for dark energy, namely a type of matter that contributes to eqn. (2) like a Λ term, such that $\ddot{R} > 0$.[26]

Most cosmologists treat these developments as akin to Le Verrier's discovery of Neptune. In both cases, unexpected results regarding the distribution of matter are inferred from observational discrepancies using the theory of gravity. Unlike the case of Le Verrier, however, this case involves the introduction of new *types* of matter rather than merely an additional planet. The two types of matter play very different roles in cosmology, despite the shared adjective. Dark energy affects cosmological expansion but is irrelevant on smaller scales, whereas dark matter dominates the dynamics of bound gravitational systems such as galaxies. There are important contrasts in the evidential cases in their favor and in their current statuses. Some cosmologists have called the concordance model "absurd" and "preposterous" because of the oddity of these new types of matter and their huge abundance relative to that of ordinary matter. There is also not yet an analog of Le Verrier's successful follow-up telescopic observations. Perhaps the appropriate historical analogy is instead the "zodiacal masses" introduced to account for Mercury's perihelion motion before GR. Why not modify the underlying gravitational theory rather than introduce one or both of these entirely new types of matter?

The ongoing debate between accepting dark matter and dark energy vs. pursuing alternative theories of gravity and cosmology turns on a number of issues familiar to philosophers of science. Does the evidence underdetermine the appropriate gravitational theory? At what stage should the need to introduce distinct types of matter with exotic properties cast doubt on the gravitational

[24] "Hot" vs. "cold" refers to the thermal velocities of relic particles for different types of dark matter. Hot dark matter decouples while still "relativistic," in the sense that the momentum is much greater than the rest mass, and relics at late times would still have large quasi-thermal velocities. Cold dark matter is "non-relativistic" when it decouples, meaning that the momentum is negligible compared to the rest mass, and relics have effectively zero thermal velocities.

[25] Type Ia supernovae do not have the same intrinsic luminosity, but the shape of the light curve (the luminosity as a function of time after the initial explosion) is correlated with intrinsic luminosity. See Kirshner (2009) for an overview of the use of supernovae in cosmology.

[26] These brief remarks are not exhaustive; there are further lines of evidence for dark matter and dark energy; see, e.g., Bertone et al. (2005) for a review of evidence for dark matter and Huterer (2010) on dark energy.

theory, and qualify as anomalies in Kuhn's sense? How successful are alternative theories compared to GR and the Standard Model, relative to different accounts of what constitutes empirical success? What follows is meant to be a primer identifying the issues that seem most relevant to a more systematic treatment of these questions.[27]

Confidence that GR adequately captures the relevant physics supports the mainstream position, accepting dark matter and dark energy. The application of GR at cosmological scales involves a tremendous extrapolation, but this kind of extrapolation of presumed laws has been incredibly effective throughout the history of physics. This particular extrapolation, furthermore, does not extend beyond the expected domain of applicability of GR. No one trusts GR at sufficiently high energies, extreme curvatures, and short length scales. Presumably it will be superseded by a theory of quantum gravity. Discovering that GR fails at low energies, low curvature, and large length scales — the regime relevant to this issue — would, however, be extremely surprising. In fact, avoiding dark matter entirely would require the even more remarkable concession that *Newtonian* gravity fails at low accelerations. In addition to the confidence in our understanding of gravity in this regime, GR has proven to be an extremely rigid theory that cannot be easily changed or adjusted.[28] At present, there is no compelling way to modify GR so as to avoid the need for dark matter and dark energy, while at the same time preserving GR's other empirical successes and basic theoretical principles. (Admittedly this may reflect little more than a failure of imagination; it was also not obvious how to change Newtonian gravity to avoid the need for zodiacal masses.)

The independence of the different lines of evidence indicating the need for dark matter and dark energy provides a second powerful argument in favor of the mainstream position. The sources of systematic error in estimates of dark matter from big bang nucleosynthesis and galaxy rotation curves (discussed below), for example, are quite different. Evidence for dark energy also comes from observations with very different systematics, although they all measure properties of dark energy through its impact on space-time geometry and structure formation. Several apparently independent parts of the Standard Model would need to be mistaken in order for all these different lines of reasoning to fail.

The case for dark energy depends essentially on the Standard Model, but there is a line of evidence in favor of dark matter based on galactic dynamics rather than cosmology. Estimates of the total mass for galaxies (and clusters of galaxies), inferred from observed motions in conjunction with gravitational theory, differ dramatically from mass estimates based on observed luminous matter.[29] To take the most famous example, the orbital velocities of stars and gas in spiral galaxies would be expected to drop with the radius as $r^{-1/2}$ outside the bright central region; observations indicate

[27]See Vanderburgh (2003, 2005) for a philosopher's take on these debates.

[28]See Sotiriou and Faraoni (2010) for a review of one approach to modifying GR, namely by adding higher-order curvature invariants to the Einstein-Hilbert action. These so-called "$f(R)$ theories" (the Ricci scalar R appearing in the action is replaced by a function $f(R)$) have been explored extensively within the last five years, but it has proven to be difficult to satisfy a number of seemingly reasonable constraints. Uzan (2010) gives a brief overview of other ways of modifying GR in light of the observed acceleration.

[29]The mass estimates differ both in total amount of mass present and its spatial distribution. Estimating the mass based on the amount of electromagnetic radiation received (photometric observations) requires a number of further assumptions regarding the nature of the objects emitting the radiation and the effects of intervening matter, such as scattering and absorption (extinction).

instead that the velocities asymptotically approach a constant value as the radius increases.[30] There are several other properties of galaxies and clusters of galaxies that lead to similar conclusions. The mere existence of spiral galaxies seems to call for a dark matter halo, given that the luminous matter alone is not a stable configuration under Newtonian gravity.[31] The case for dark matter based on these features of galaxies and clusters draws on Newtonian gravity rather than GR. Relativistic effects are typically ignored in studying galactic dynamics, given the practical impossibility of modeling a full galactic mass distribution in GR. But it seems plausible to assume that the results of Newtonian gravity for this regime can be recovered as limiting cases of a more exact relativistic treatment.[32]

There is another way of determining the mass distribution in galaxies and clusters that does depend on GR, but not the full Standard Model. Even before he had reached the final version of GR, Einstein realized that light-bending in a gravitational field would lead to the magnification and distortion of images of distant objects. This lensing effect can be used to estimate the total mass distribution of a foreground object based on the distorted images of a background object, which can then be contrasted with the visible matter in the foreground object.[33] Estimates of dark matter based on gravitational lensing are in rough agreement with those based on orbital velocities in spiral galaxies, yet they draw on different regimes of the underlying gravitational theory.

Critics of the mainstream position argue that introducing dark matter and dark energy with properties chosen precisely to resolve the mass discrepancy is *ad hoc*. Whatever the strength of this criticism, the mainstream position does convert an observational discrepancy in cosmology into a problem in fundamental physics, namely that of providing a believable physics for dark matter and dark energy.

In this regard the prospects for dark matter seem more promising. Theorists have turned to extensions of the Standard Model of particle physics in the search for dark matter candidates, in the form of weakly interacting massive particles. Although the resulting proposals for new types of

[30]This behavior is usually described using the rotation curve, a plot of orbital velocity as a function of the distance from the galactic center. The "expected" behavior (dropping as $r^{-1/2}$ after an initial maximum) follows from Newtonian gravity with the assumption that all the mass is concentrated in the central region, like the luminous matter. The discrepancy cannot be evaded by adding dark matter with the same distribution as the luminous matter; in order to produce the observed rotation curves, the dark matter has to be distributed as a halo around the galaxy.

[31]In a seminal paper, Ostriker and Peebles (1973) argued in favor of a dark matter halo based on an N-body simulation, extending earlier results regarding the stability of rotating systems in Newtonian gravity to galaxies. These earlier results established a criterion for the stability of rotating systems: if the rotational energy in the system is above a critical value, compared to the kinetic energy in random motions, then the system is unstable. The instability arises, roughly speaking, because the formation of an elongated bar shape leads to a larger moment of inertia and a lower rotational energy. Considering the luminous matter alone, spiral galaxies appear to satisfy this criterion for instability; Ostriker and Peebles (1973) argued that the addition of a large, spheroidal dark matter halo would stabilize the luminous matter.

[32]This assumption has been challenged; see Cooperstock and Tieu (2007) for a review of their controversial proposal that a relativistic effect important in galactic dynamics, yet absent from the Newtonian limit, eliminates the need for dark matter.

[33]Gravitational lensing occurs when light from a background object such as a quasar is deflected due to the spacetime curvature produced, according to GR, by a foreground object, leading to multiple images of a single object. The detailed pattern of these multiple images and their relative luminosity can be used to constrain the distribution of mass in the foreground object.

particles are speculative, there is no shortage of candidates that are theoretically natural (according to the conventional wisdom) and as yet compatible with observations. There also do not appear to be any fundamental principles that rule out the possibility of appropriate dark matter candidates.

With respect to dark energy, by contrast, the discovery of accelerating expansion has exacerbated what many regard as a crisis in fundamental physics.[34] Dark energy can either take the form of a true Λ term or some field whose stress-energy tensor effectively mimics Λ. As such it violates an energy condition associated with "ordinary" matter, although few theorists now take this condition as inviolable.[35] A more fundamental problem arises in comparing the observed value of dark energy with a calculation of the vacuum energy density in QFT. The vacuum energy of a quantum field diverges. It is given by integrating the zero-point contributions to the total energy, $\frac{1}{2}\hbar\omega(k)$ per oscillation mode, familiar from the quantum harmonic oscillator, over momentum (k). Evaluating this quartically divergent quantity by introducing a physical cutoff at the Planck scale, the result is 120 orders of magnitude larger than the observed value of the cosmological constant.[36] This is sometimes called the "old" cosmological constant problem: why isn't there a cancellation mechanism that leads to $\Lambda = 0$? Post-1998, the "new" problem concerns understanding why the cosmological constant is quite small (relative to the vacuum energy density calculated in QFT) but *not* exactly zero, as indicated by the accelerating expansion.

Both problems rest on the crucial assumption that the vacuum energy density in QFT couples to gravity as an effective cosmological constant. Granting this assumption, the calculation of vacuum energy density qualifies as one of the worst theoretical predictions ever made. What turns this dramatic failure into a crisis is the difficulty of controling the vacuum energy density, by, say, introducing a new symmetry. Recently, however, an anthropic response to the problem has drawn increasing support. On this approach, the value of Λ is assumed to vary across different regions of the universe, and the observed value is "explained" as an anthropic selection effect (we will return

[34] See, in particular, Weinberg (1989) for an influential review of the cosmological constant problem prior to the discovery of dark energy, and, e.g., Polchinski (2006) for a more recent discussion.

[35] Energy conditions place restrictions on the stress-energy tensor appearing in EFE. They are useful in proving theorems for a range of different types of matter with some common properties, such as "having positive energy density" or "having energy-momentum flow on or within the light cone". In this case the strong energy condition is violated; for the case of an ideal fluid discussed above, the strong energy condition holds iff $\rho + 3p \geq 0$. Cf., for example, Chapter 9 of Wald (1984) for definitions of other energy conditions.

[36] In more detail, the relevant integral is

$$\rho_\nu = \int_0^\ell \frac{d^3k}{(2\pi)^3}\frac{\sqrt{k^2+m^2}}{2} \approx \frac{\ell^4}{16\pi^2}$$

For a Planck scale cut-off, $\ell_p \approx 1.6 \times 10^{-35}m$, the resulting vacuum energy density is given by $\rho_\nu \approx 2 \times 10^{110} erg/cm^3$, compared to observational constraints on the cosmological constant — $\rho_\Lambda \approx 2 \times 10^{-10} erg/cm^3$. Choosing a much lower cut-off scale, such as the electroweak scale $\ell_{ew} \approx 10^{-18}m$, is not enough to eliminate the huge discrepancy (still 55 orders of magnitude). Reformulated in terms of the effective field theory approach, the cosmological constant violates the technical condition of "naturalness." Defining an effective theory for a given domain requires integrating out higher energy modes, leading to a rescaling of the constants appearing in the theory. This rescaling would be expected to drive the value of terms like the cosmological constant up to the scale of the cut-off; a smaller value, such as what is observed, apparently requires an exquisitely fine-tuned choice of the bare value to compensate for this scaling behavior, given that there are no symmetry principles or other mechanisms to preserve a low value.

to this approach in §7 below).

Whether abandoning the assumption that the vacuum energy is "real" and gravitates is a viable response to the crisis depends on two issues. First, what does the empirical success of QFT imply regarding the reality of vacuum energy? The treatment of the scaling behavior of the vacuum energy density above indicates that vacuum energy in QFT is not fully understood given current theoretical ideas. This is not particularly threatening in calculations that do not involve gravity, since one can typically ignore the vacuum fluctuations and calculate quantities that depend only on relative rather than absolute values of the total energy. This convenient feature also suggests, however, that the vacuum energy may be an artifact of the formalism that can be stripped away while preserving QFT's empirical content. Second, how should the standard treatment of the vacuum energy from flat-space QFT be extended to the context of the curved spacetimes of GR? The symmetries of flat spacetime so crucial to the technical framework of QFT no longer obtain, and there is not even a clear way of identifying a unique vacuum state in a generic curved spacetime. Reformulating the treament of the scaling behavior of the vacuum energy density is thus a difficult problem. It is closely tied to the challenge of combining QFT and GR in a theory of quantum gravity. In QFT on curved spacetimes (one attempt at combining QFT and GR) different renormalization techniques are used that eliminate the vacuum energy. The question is whether this approach simply ignores the problem by *fiat* or reflects an appropriate generalization of renormalization techniques to curved spacetimes. These two issues are instances of familiar questions for philosophers — what parts of a theory are actually supported by its empirical success, and what parts should be preserved or abandoned in combining it with another theory? Philosophers have offered critical evaluations of the conventional wisdom in physics regarding the cosmological constant problem, and there are opportunities for further work.[37]

Returning to the main line of argument, the prospects for an analog of Le Verrier's telescopic observations differ for dark matter and dark energy. There are several experimental groups currently searching for dark matter candidates, using a wide range of different detector designs and searching through different parts of the parameter space (see, e.g., Sumner 2002). Successful detection by one of these experiments would provide evidence for dark matter that does not depend directly on gravitational theory. The properties of dark energy, by way of contrast, insure that any attempt at a non-cosmological detection would be futile. The energy density introduced to account for accelerated expansion is so low, and uniform, that any local experimental study of its properties is practically impossible given current technology.

There are different routes open for those hoping to avoid dark energy and dark matter. Dark energy is detected by the observed departures from the spacetime geometry that one would expect in a matter-dominated FLRW model. Taking this departure to indicate the presence of an unexpected contribution to the universe's overall matter and energy content thus depends on assuming that the FLRW models hold. There are then two paths open to those exploring alternatives to dark energy. The first is to change the underlying gravitational theory, and to base cosmology on an alternative to GR that does not support this inference. A second would be to retain GR but reject the FLRW

[37]See, in particular, Rugh and Zinkernagel (2002) for a thorough critical evaluation of the cosmological constant problem, as well as Earman (2001) and Saunders (2002).

models. For example, models which describe the observable universe as having a lower density than surrounding regions can account for the accelerated expansion without dark energy. Cosmologists have often assumed that we are not in a "special" location in the universe. This claim is often called the "Copernican Principle," to which we will return in §5 below. This principle obviously fails in these models, as our observable patch would be located in an unusual region — a large void.[38] It has also been proposed that the accelerated expansion may be accounted for by GR effects that come into view in the study of inhomogeneous models without dark energy. Buchert (2008) reviews the idea that the back-reaction of inhomogeneities on the background spacetime leads to an effective acceleration. These proposals both face the challenge of accounting for the various observations that are regarded, in the concordance model, as manifestations of dark energy.

On the other hand, dark matter can only be avoided by modifying gravity — *including Newtonian gravity* — as applied to galaxies. Milgrom (1983) argued that a modification of Newtonian dynamics (called MOND) successfully captures several aspects of galaxy phenomenology. According to Milgrom's proposal, below an acceleration threshold ($a_0 \approx 10^{-10}\ m/s^2$) Newton's second law should be modified to $F = m\frac{a^2}{a_0}$. This modification accounts for observed galaxy rotation curves without dark matter. But it also accounts for a wide variety of other properties of galaxies, many of which Milgrom successfully predicted based on MOND (see, e.g., Sanders and McGaugh (2002); Bekenstein (2010) for reviews). Despite these successes, MOND has not won widespread support. Even advocates of MOND admit that at first blush it looks like an extremely odd modification of Newtonian gravity. Yet it fares remarkably well in accounting for various features of galaxies — too well, according to its advocates, to be dismissed as a simple curve fit. MOND does not fare as well for clusters of galaxies and may have problems in accounting for structure formation. In addition to these potential empirical problems, it is quite difficult to embed MOND within a compelling alternative to GR.

In sum, it is reasonable to hope that the situation with regard to dark matter and dark energy will be clarified in the coming years by various lines of empirical investigation that are currently underway. The apparent underdetermination of different alternatives may prove transient, with empirical work eventually forcing a consensus. Whether or not this occurs, there is also a possibility for contributions to the debate from philosophers concerned with underdetermination and evidential reasoning. The considerations above indicate that even in a case where competing theories are (arguably) compatible with all the evidence that is currently available, scientists certainly do not assign equal credence to the truth of the competitors. Philosophers could contribute to this debate by helping to articulate a richer notion of empirical support that sheds light on these judgments (cf. the closing chapter of Harper 2011).

4 Uniqueness of the Universe

The uniqueness of the universe is the main contrast between cosmology and other areas of physics. The alleged methodological challenge posed by uniqueness was one of the main motivations for the steady state theory. The claim that a generalization of the cosmological principle, the "perfect

[38] See Ellis (2011) for an overview of the use of inhomogeneous models as an alternative to dark energy.

cosmological principle," is a precondition for scientific cosmology, is no longer accepted.[39] It is, however, often asserted that cosmology cannot discover new laws of physics as a direct consequence of the uniqueness of its object of study.[40] Munitz (1962) gives a concise formulation of this common argument:

> With respect to these familiar laws [of physics] ... we should also mark it as a prerequisite of the very meaning and use of such laws that we be able to refer to an actual or at least possible *plurality* of instances to which the law applies. For unless there were a plurality of instances there would be neither interest nor sense in speaking of a law at all. If we knew that there were only one actual or possible instance of some phenomenon it would hardly make sense to speak of finding a law for this unique occurrence *qua* unique. This last situation however is precisely what we encounter in cosmology. For the fact that there is at least but not more than one universe to be investigated makes the search for *laws* in cosmology inappropriate. (Munitz 1962, 37)

Ellis (2007) reaches a similar conclusion:

> The concept of 'Laws of Physics' that apply to only one object is questionable. *We cannot scientifically establish 'laws of the universe' that might apply to the class of all such objects, for we cannot test any such proposed law except in terms of being consistent with one object (the observed universe).* (Ellis 2007, 1217, emphasis in the original)

His argument for this claim emphasizes that we cannot perform experiments on the universe by creating particular initial conditions. In many observational sciences (such as astronomy) the systems under study also cannot be manipulated, but it is still possible to do without experiments by studying an ensemble of instances of a given type of system. However, this is also impossible in cosmology.

If these arguments are correct, then cosmology should be treated as a merely descriptive or historical science that cannot discover novel physical laws. Both arguments rest on problematic assumptions regarding laws of nature and scientific method. Here I will sketch an alternative account that allows for the possibility of testing cosmological laws despite the uniqueness of the universe.

Before turning to that task, I should mention a different source of skepticism regarding the possibility of scientific cosmology based on distinctive laws. Kant argued that attempts at scientific cosmology inevitably lead to antinomies because no object corresponds to the idea of the "universe". Relativistic cosmology circumvents this argument insofar as cosmological models have global properties that are well-defined, albeit empirically inaccessible. (This is discussed further in §5.) Yet contemporary worries resonant with Kant concerning how to arrive at the appropriate

[39]The pronouncements of the steady state theorists drew a number of philosophers into debates regarding cosmology in the 60s. See Kragh (1996) for a historical account of the steady state theory, and the rejection of it in favor of the big bang theory by the scientific community, and Balashov (2002) for a discussion of their views regarding laws.

[40]See Pauri (1991); Scheibe (1991); Torretti (2000) for discussions of the implications of uniqueness and the status of laws in cosmology.

concepts for cosmological theorizing. Smolin (2000) criticizes relativistic cosmology for admitting such global properties, and proposes instead that: "Every quantity in a cosmological theory that is formally an observable should in fact be measurable by some observer inside the universe."[41] A different question arises, for example, in extrapolating concepts to domains such as the early universe. Rugh and Zinkernagel (2009) argue that there is no physical footing for spacetime concepts in the very early universe due to the lack of physical processes that can be used to determine spacetime scales.

Munitz's formulation makes his assumptions about the relationship between laws and phenomena clear: the phenomena are instances of the law, just as $\mathsf{F}a \wedge \mathsf{G}a$ would be an instance of the "law" $\forall x(\mathsf{F}x \rightarrow \mathsf{G}x)$. Even if we grant this conception of laws, Munitz's argument would only apply to a specific kind of cosmological law. If we take EFE as an example of a "cosmological law," then it has multiple instantiations in the straightforward sense that every subregion of a solution of EFE is also a solution.[42] The same holds for other local dynamical laws applicable in cosmology, such as those of QFT. A single universe has world enough for multiple instantiations of the local dynamics. This is true as well of laws whose effects may have, coincidentally, only been important within some finite subregion of the universe. For example, consider a theory, such as inflation (see §6 below), whose implications are only manifest in the early universe. The laws of this theory would be "instantiated" again if we were ever able to reach sufficiently high energy levels in an experimental setting. Although the theory may *in practice* only have testable implications "once," it has further counterfactual implications. Munitz's argument would apply, however, to cosmological laws that are formulated directly in terms of global properties, as opposed to local dynamical laws extrapolated to apply to the universe as a whole. Subregions of the universe would not count as instantiations of a "global law" in the same sense that they are instantiations of the local dynamical laws. Penrose's Weyl curvature hypothesis (proposed in Penrose 1979) is an example of such a law.[43] This law is formulated as a constraint on initial conditions and it does differ strikingly in character from local dynamical laws.

Phenomena are not, however, "instantiations" of laws of nature in Munitz's straightforward logical sense. Treating them as such attributes to the laws empirical content properly attributed only to equations derived from the laws with the help of supplementary conditions.[44] A simple example should help to make this contrast clear. Newton's three laws of motion must be combined with other assumptions regarding the relevant forces and distribution of matter to derive a set of equations of motion, describing, say, the motion of Mars in response to the Sun's gravitational field. It is this derived equation describing Mars's motion that is compared to the phenomena and used to calculate the positions of Mars given some initial conditions. The motion of Mars is not an

[41]This is the first of two principles Smolin advocates as necessary to resolve the problem of time, and he further argues that they bring cosmological theorizing more in line with scientific practice.

[42]That is, for any open set O of the spacetime manifold M, if $\langle M, g_{ab}, T_{ab}\rangle$ is a solution of EFE then so is $\langle O, g_{ab}|_O, T_{ab}|_O\rangle$ taken as a spacetime in its own right.

[43]The Weyl tensor represents, roughly speaking, the gravitational degrees of freedom in GR with the degrees of freedom for the source terms removed. Penrose's hypothesis holds that this tensor vanishes in the limit as one approaches the initial singularity.

[44]This mistake also underlies much of the discussion of "ceteris paribus" laws, and here I draw on the line of argument due to Smith (2002); Earman and Roberts (1999).

"instance" of Newton's laws; rather, the motion of Mars is well approximated by a solution to an equation derived from Newton's laws along with a number of other assumptions.

Ellis's argument does not explicitly rest on a conception of phenomena as instatiations of laws. But he and Munitz both overlook a crucial aspect of testing laws. Continuing with the same example, there is no expectation that at a given stage of inquiry one has completely captured the motion of Mars with a particular derived equation, even as further physical effects (such as the effects of other planets) are included. The success of Newton's theory (in this case) consists in the ability to give more and more refined descriptions of the motion of Mars, all based on the three laws of motion and the law of gravity. This assessment does not depend primarily on "multiplicity of instances," experimental manipulation, or observation of other members of an ensemble. Instead, the modal force of laws is reflected in their role in developing a richer account of the motions. Due to this role they can be subject to ongoing tests.

The standard arguments that it is not possible to discover laws in cosmology assume that the universe is not only unique, but in effect "given" to us entirely, all at once — leaving cosmologists with nothing further to discover, and no refinements to make and test. A novel law in cosmology could be supported by its success in providing successively more refined descriptions of some aspect of the universe's history, just as Newtonian mechanics is supported (in part) by its success in underwriting research related to the solar system. This line of argument, if successful, shows that cosmological laws are testable in much the same sense as Newton's laws. This suggests that "laws of the universe" should be just as amenable to an empiricist treatment of the laws of nature as are other laws of physics.[45]

None of this is to say that there are no distinctive obstacles to assessing cosmological laws. But we need to disentangle obstacles that arise due to specific features of our universe from those that follow from the uniqueness of the object of study. Consider (contrary to the Standard Model) a universe that reached some finite maximum temperature as $t \to 0$, and suppose (perhaps more absurdly) that physicists in this universe had sufficient funds to build accelerators to probe physics at this energy scale. Many of the challenges faced in early universe cosmology in our universe would not arise for cosmologists in this other possible universe. They would have independent lines of evidence (from accelerator experiments and observations of the early universe) to aid in reconstructing the history of the early universe, rather than basing their case in favor of novel physics solely on its role in the reconstruction. This suggests that obstacles facing cosmology have to do primarily with theoretical and observational accessibility, which may be exacerbated by uniqueness, rather than with uniqueness of the universe *per se*.

5 Global Structure

The Standard Model takes the universe to be well-approximated by an FLRW model at sufficiently large scales. To what extent can observations determine the spacetime geometry of the universe directly? The question can be posed more precisely in terms of the region visible to an observer at

[45] There may be other philosophical requirements on an account of laws of nature that do draw a distinction between laws of physics and laws of the universe.

a location in spacetime p — the *causal past*, $J^-(p)$, of that point. This set includes all points from which signals traveling at or below the speed of light can reach p.[46] What can observations confined to $J^-(p)$ reveal about: (1) the spacetime geometry of $J^-(p)$ itself, and (2) the rest of spacetime outside of $J^-(p)$? Here we will consider these questions on the assumption that GR and our other physical theories apply universally, setting aside debates (such as those in §3) about whether these are the correct theories. How much do these theories allow us to infer, granting their validity?

Spacetime geometry is reflected in the motion of astronomical objects and in effects on the radiation they emit, such as cosmological red-shift. To what extent would the spacetime geometry be fixed by observations of an "ideal data set," consisting of comprehensive observations of a collection of standard objects, with known intrinsic size, shape, mass, and luminosity, distributed throughout the universe? Of course astronomers cannot avail themselves of such a data set. Converting the actual data recorded by observatories into a map of the universe, filled with different kinds of astronomical objects with specified locations and states of motion, is an enormously difficult task. The difficulty of completing this task poses one kind of epistemic limitation to cosmology. Exploring this limitation would require delving into the detailed astrophysics used to draw conclusions regarding the nature, location, and motion of distant objects. This kind of limitation contrasts with one arising from a different source, namely that we have an observational window on $J^-(p)$ rather than the entire spacetime. Even if we had access to an ideal data set, what we can observe is not sufficient to answer questions regarding global spacetime geometry unless we accept further principles underwriting local-to-global inferences.

The modest goal of pinning down the geometry of $J^-(p)$ observationally can be realized by observers with the ideal data set mentioned above (see Ellis 1980, Ellis et al. 1985). The relevant evidence comes from two sources: the radiation emitted by distant objects reaching us along our null cone, and evidence, such as geophysical data, gathered from "along our world line," so to speak. Ellis et al. (1985) prove that the ideal data set is necessary and sufficient, in conjunction with EFE and a few other assumptions, to determine the spacetime geometry of $J^-(p)$. Considering the ideal data set helps to clarify the contrast between what we can in principle determine locally, namely the spacetime geometry of $J^-(p)$, and what we can determine globally.

For points p, q with non-intersecting causal pasts, we would not expect the physical state on $J^-(p)$ to fix that of $J^-(q)$.[47] Does the spacetime geometry of $J^-(p)$, or of a collection of such sets, nonetheless constrain the large-scale or global properties of spacetime? Global properties of spacetime vary in general relativity, because unlike earlier theories such as Newtonian mechanics, spacetime is treated as dynamical rather than as a fixed background. EFE impose a local constraint on the spacetime geometry, but this is compatible with a wide variety of global properties.[48]

[46]In Minkowski spacetime, this set is the past lobe of the light cone at p, including interior points and the point p itself. A point p *causally precedes* q ($p < q$), if there is a future-directed curve from p to q with tangent vectors that are timelike or null at every point. The sets $J^\pm(p)$ are defined in terms of this relation: $J^-(p) = \{q : q < p\}$, $J^+(p) = \{q : p < q\}$, the *causal past* and *future* of the point p, and the definition generalizes immediately to spacetime regions.

[47]The Gauss-Codacci constraint equations do impose some restrictions on spacelike separated regions, although these would not make it possible to determine the state of one region from the other; see Ellis and Sciama (1972).

[48]A *local* property of a spacetime is one that is shared by locally isometric spacetimes, whereas global properties are not. (Two spacetimes are locally isometric iff for any point p in the first spacetime, there is an open neighborhood

Various global properties have been defined as part of stating and proving theorems such as the singularity theorems, including "causality conditions" which specify the extent to which a spacetime deviates from the causal structure of Minkowski spacetime (see Geroch and Horowitz 1979 for a clear introduction). For example, a *globally hyperbolic* spacetime possesses a Cauchy surface, a null or spacelike surface Σ intersected exactly once by every inextendible timelike curve. In a spacetime with a Cauchy surface, EFE admit a well-posed initial value formulation: specifying appropriate initial data on a Cauchy surface Σ determines a unique solution to the field equations (up to diffeomorphism). This is properly understood as a global property of the entire spacetime. Although submanifolds of a given spacetime may be compatible or incompatible with global hyperbolicity, this property cannot be treated as a property ascribed to local regions and then "added up" to deliver a global property.

What does $J^-(p)$ reveal about the rest of spacetime? Suppose we do not impose any strong global assumptions such as isotropy and homogeneity. Fully specifying the physical state in the region $J^-(p)$ places few constraints on the global properties of spacetime. This is clear if we consider what is shared by all the spacetimes into which $J^-(p)$ can be isometrically embedded, where we allow p to be any point in a given spacetime.[49] (That is, we shift from considering the causal past of a single observer to the causal past of *all* possible observers in the spacetime.) Call this the set of spacetimes "observationally indistinguishable" (OI) from a given spacetime. Except for the exceptional case where there is a p' such that, like Borges's Aleph, $J^-(p')$ includes the *entire* spacetime, there is a technique (due to Malament 1977; Manchak 2009) for constructing OI counterparts that do not share all the global properties of the original spacetime. The property of having a Cauchy surface, for example, will not be shared by all the members of a set of OI spacetimes.[50] More generally, the only properties that will be held in common in all members of the set of OI spacetimes are those that can be conclusively established by a single observer somewhere in the spacetime.[51]

The scope of underdetermination can be reduced by imposing constraints that eliminate potential OI spacetimes. Consider, for example, restricting consideration to spacetimes that are spatially homogeneous. The isometries on Σ (implied by homogeneity), which carry any point on Σ into any other, block the construction of an indistinguishable counterpart with different global properties.[52]

of the point such that it can be mapped to an isometric open neighborhood of the second spacetime (and vice versa).)

[49]The underdetermination problem still arises if we consider the past of future-inextendible curves; see Glymour (1977); Malament (1977) for discussion.

[50]Malament (1977) reviews several different definitions of observational indistinguishability and gives a series of constructions of OI-spacetimes lacking specific global properties. Note that Malament defines OI in terms of the *chronological* rather than *causal* sets, which include the interior of the light cone but not the cone itself. (The definition follows the one given in footnote (46), dropping the phrase "or null.") Manchak (2009) proves that Malament's technique for constructing such spacetimes fails only in the exceptional case noted in the text. Cf. Norton (2011), who argues that the inductive generalizations from $J^-(p)$ to other regions of spacetime lack clear justification.

[51]As Malament emphasizes, this includes the *failure* of the causality conditions to hold.

[52]Pick a point in $p \in M$ such that p lies in Σ and its image $\phi(p) \in M'$ under the isometric imbedding map ϕ. If homogeneity holds, then M' must include an isometric "copy" Σ' of the *entire Cauchy surface* Σ along with its entire causal past. Take ξ to be an isometry of the spatial metric defined on Σ, and ξ' an isometry on Σ'. Since $\phi \circ \xi(p) = \xi' \circ \phi(p)$, and any point $q \in \Sigma$ can be reached via ξ, it follows that Σ is isometric to Σ'. Mapping points

Homogeneity is just one example of a global property that could be imposed. Whatever property is imposed to eliminate underdetermination, it must be *global* to be effective given that the technique for constructing indistinguishable counterparts preserves local properties.

This line of argument clarifies the "cosmological principle." The cosmological principle is the strongest of many possible "uniformity principles" or global stipulations that allow local-to-global inferences. If we require only that the $J^{-}(p)$ sets for all observers can be embedded in a cosmological model, then the global properties of spacetime are radically underdetermined. Introducing different constraints on the construction of the indistinguishable counterparts mitigates the degree of underdetermination. The cosmological principle is the strongest of these constraints — strong enough to eliminate the underdetermination: *every* observer can take their limited view on the universe as accurately reflecting its global properties.

However, this merely pushes the original question back one step: what grounds do we have for imposing such a global constraint on spacetime?[53] It is unappealing to simply assert that the cosmological principle holds *a priori*, or to treat it as a pre-condition for cosmological theorizing. But one may hope to justify the principle by appealing to a weaker general principle in conjunction with theorems relating homogeneity and isotropy. Global isotropy around every point implies global homogeneity, and it is natural to seek a similar theorem with a weaker antecedent formulated in terms of observable quantities. The Ehlers-Geren-Sachs theorem (Ehlers et al. 1968) shows that if all fundamental observers in an expanding model find that freely propagating background radiation is exactly isotropic, then their spacetime is an FLRW model.[54] If our causal past is "typical," observations along our worldline will constrain what other observers should see. This assumption is often called the "Copernican principle," which requires that no point p is distinguished from other points q by any spacetime symmetries. This principle rules out models such as Ellis et al. (1978)'s example of a "cylindrical" counterpart to the observed universe (see Fig. 1).[55] (This example illustrates the tension between the Copernican Principle and anthropic reasoning (see §7 below). Ellis et al. (1978) point out that in their model one would only expect to find observers near the axis of symmetry of the model, as that is the only region hospitable to life.) Combining the observed near isotropy of the CBR, the EGS theorem, and the Copernican Principle yields an argument in favor of the approximate validity of the FLRW models.

Alternatively, one could dispense with the Copernican Principle and its ilk by showing that an early phase of the universe's evolution leads to an approximately FLRW universe. This was the aim

along an inextendible timelike curve from M into M' eventually leads to an isometric copy of our original spacetime, assuming that both spacetimes are inextendible.

[53] See Beisbart (2009) for a thorough discussion of different attempts to justify the cosmological principle.

[54] Recent work has clarified the extent to which this result depends on the various exact claims made in the antecedent. The fundamental observers do not need to measure *exact* isotropy for a version of the theorem to go through: Stoeger et al. (1995) have further shown that *almost* isotropic CMBR measurements imply that the spacetime is an *almost* FLRW model, in a sense that can be made precise; see Clarkson and Maartens (2010) for a review.

[55] Their model replaces *temporal* evolution in the Standard Model with *spatial* variation, with spherical symmetry around a preferred axis. They construct the model to recapture the observational results of the Standard Model for observers situated near the axis of symmetry. Such a preferred location is exactly what the Copernican Principle rules out.

of Misner's "chaotic cosmology" program launched shortly after the discovery of the CBR, an aim taken up with greater acclaim by inflationary cosmology (see §6 below). If this approach succeeds, then homogeneity and isotropy over some length scale would be a consequence of underlying physics, effectively replacing *a priori* principles regarding the uniformity of nature with factual claims about the universe's evolution. The warrant for an inductive inference regarding distant regions of the universe would then depend on the justification for this account. Note, however, that the account may not justify the conclusion that the universe is *globally* almost-FLRW. In the case of inflation, for example, homogeneity and isotropy hold in the interior of an inflationary bubble (which could be much larger than $J^{-}(p)$), but the universe at much larger scales has dramatic non-uniformities (bubble walls, colliding bubbles, regions between the bubbles, and so on).

The Copernican Principle has come under increased scrutiny recently due to its role in the case for dark energy. Departures from an FLRW geometry could simply indicate the failure of the models rather than the presence of a new kind of matter. Recently there have been two suggestions for ways to test the Copernican Principle on scales comparable to the observable universe. First, the Sunyaev-Zel'dovich effect[56] can be used to indirectly measure the isotropy of the CBR as observed from distant points. Any anisotropies in the CBR as seen at a distant point q will be reflected in a temperature difference in the scattered radiation; the distortion in the observed black-body spectrum in principle reveals the failure of isotropy from distant points not on our worldline (Caldwell and Stebbins 2008). This allows one to prove that the local universe is almost-FLRW based on an EGS theorem and observations of the CBR *without* invoking the Copernican Principle (Clifton et al. 2011). A second test of the Copernican principle is based on a consistency relation between several observables that holds in the FLRW models (Uzan et al. 2008).

These discussions focus on whether $J^{-}(p)$ can be well approximated by an FLRW model. This question is closely tied to assessing the case for dark energy, and in determining the parameters of the Standard Model. What are the further implications if the universe is almost-FLRW on much larger scales, or if the cosmological principle holds *globally* throughout all of spacetime? More generally, what are the empirical stakes of determining the global properties of spacetime? Some global spacetime properties are plausibly treated as pre-conditions for the possibility of formulating local dynamical laws.[57] And the global properties are obvious candidates for fundamental features of spacetime from a realist's point of view. Proofs of the singularity theorems require assumptions regarding global causal structure. Further, the origin and eventual fate of the universe are quite different in a globally almost-FLRW model and in an observationally indistinguishable counterpart to it. Yet despite all of this, there is a clear contrast between claiming that the observable universe is almost-FLRW and the extension of that to a global claim regarding all of spacetime. The former plays a fundamental role in evidential reasoning in contemporary cosmology, whereas the latter is disconnected from empirical research by its very nature. Thus the status of the cosmological

[56]The Sunyaev-Zel'dovich effect refers to the distortion of the spectrum of CBR photons that results from scattering by hot gases in galaxy clusters. Due to the scattering by the hot gas the CBR spectrum will have an excess of high energy photons and a deficit of low energy photons; measurements of this distortion can in principle be used to measure the temperature and mass of the gas in the cluster.

[57]For example, topological properties such as temporal orientability, which allows for a globally consistent choice of the the direction of time, seem to be presupposed in formulating local dynamical laws.

principle seems to differ significantly in practice from that of other principles supporting inductive generalizations — it does not lead, as in Newton's case of taking gravity to be truly *universal*, to a wide variety of further claims that can serve as the basis for a subsequent research program.

6 Early Universe Cosmology

Extrapolating the Standard Model backwards in time leads to a singularity within a finite time, and as $t \to 0$ the temperature and energy scales increase without bound. Even if the singularity itself is somehow avoided, the early universe is expected to have reached energy scales far higher than anything produced at Fermilab or CERN. The early universe is thus a fruitful testing ground for high-energy physics, and since the early 80s there has been an explosion of research in this area. Yet it is not clear whether observations of the early universe can play anything like the role that accelerator experiments did in guiding an earlier phase of research in particle physics. Other aspects of the Standard Model are based on extrapolating well-established physics, but the physics applied to the early universe often cannot be tested by other means. Instead the case in favor of new physical ideas is often based on their role in a plausible reconstruction of the universe's history. Here I will assess a common style of argument adopted in this literature, namely that a theory of early universe cosmology should be accepted because it renders the observed history of the universe *probable* rather than merely *possible*.

There is general agreement that the (cosmological) Standard Model should be supplemented with an account of physical processes in the very early universe. The early universe falls within the domains of applicability of both quantum field theory (QFT) and general relativity, yet the two theories have yet to be combined successfully. The framework of the Standard Model is thus not expected to apply to the very early universe. Although research in quantum gravity is often motivated by calls for "theoretical unification" and the like, it can also be motivated by the more prosaic demand for a consistent theory applicable to phenomena such as the early universe and black holes (cf. Callender and Huggett 2001). This "overlapping domains" argument does not imply anything in detail regarding what an early universe theory should look like, or how it would augment or contribute to the Standard Model.

The overlapping domains argument should not be confused with the common claim that general relativity is incomplete because it "breaks down" as $t \to 0$, and fails to provide a description of what happens at (or before) the singularity.[58] It is hard to see how general relativity can be convicted of incompleteness on its own terms. (Here I am following the line of argument in Earman (1995); Curiel (1999).) If general relativity proved to be the correct final theory, then there is nothing more to be said regarding singularities; the laws of general relativity apply throughout the entire spacetime, and there is no obvious incompleteness. On the other hand there are good reasons to doubt that general relativity is the correct final theory, and further reasons to expect that the

[58] Here I am adopting the usual way of describing the objection, although this language can be quite misleading as it implicitly assumes that the singularity can be "localized" in some sense. There are convincing arguments in favor of taking singular as an adjective describing spacetime as a whole; see Curiel (1999); Geroch et al. (1982); Earman (1995).

successor to general relativity will have novel implications for singularities. But then the argument for incompleteness is based on grounds other than the mere existence of singularities.

Cosmologists often give a very different reason for supplementing the Standard Model: it is *explanatorily deficient*, because it requires an "improbable" initial state. Guth (1981) gave an influential presentation of two aspects of the Standard Model as problematic:

> The standard model of hot big-bang cosmology requires initial conditions which are problematic in two ways: (1) The early universe is assumed to be highly homogeneous, in spite of the fact that separated regions were causally disconnected (horizon problem) and (2) the initial value of the Hubble constant must be fine tuned to extraordinary accuracy ... (flatness problem). (p. 347)

Horizons in cosmology measure the maximum distance light travels within a given time period; the horizon delimits the spacetime region from which signals emitted at some time t_e traveling at or below the speed of light could reach a given point. The existence of particle horizons in the FLRW models indicates that distant regions are not in causal contact.[59] There are observed points on the CBR separated by a distance greater than the particle horizon at that time (see Fig. 2). The Standard Model assumes that these regions have the same properties — e.g., the same temperature to within 1 part in 10^5 — even though they were not in causal contact. In slightly different terms, if one expects no correlations between the causally disjoint regions it is mysterious how the observable universe could be so well approximated by an FLRW model.

The flatness problem arises because the energy density at early times has to be *very* close to the value of the critical density $\Omega = 1$.[60] An FLRW model close to the "flat" $k = 0$ model, with nearly critical density, at some specified early time is driven rapidly away from critical density under FLRW dynamics; the flat model is an unstable fixed point under dynamical evolution.[61] This aspect of the dynamics makes it extremely puzzling to find that the universe is still close to the critical density — this requires an extremely finely-tuned choice of the energy density at the Planck time $\Omega(t_p)$, namely $|\Omega(t_p) - 1| \leq 10^{-59}$.

The horizon and flatness problems both reflect properties of the FLRW models. There are other similar "fine-tuning" problems related to other aspects of the Standard Model. The account of structure formation requires a set of "seed" perturbations that have two troubling features: first, the perturbations have to be coherent on super-horizon length scales, and, second, the amplitude of the perturbations was much smaller than one would expect for natural possibilities such as thermal

[59]Following Rindler (1956), a horizon is the surface in a time slice t_0 separating particles moving along geodesics that could have been observed from a worldline γ by t_0 from those which could not. The distance to this surface, for signals emitted at a time t_e, is given by:

$$d = R(t_0) \int_{t_e}^{t_0} \frac{dt}{R(t)} \qquad (3)$$

Different "horizons" correspond to different choices of limits of integration, with the "particle horizon" defined as the limit $t_e \to 0$. The integral converges for $R(t) \propto t^n$ with $n < 1$, which holds for matter or radiation-dominated expansion, leading to a finite horizon distance. See Ellis and Rothman (1993) for a clear introduction to horizons.

[60]$\Omega =: \frac{\rho}{\rho_c}$, where the critical density is the value of ρ for the flat FLRW model, $\rho_c = \frac{3}{8\pi}\left(H^2 - \frac{\Lambda}{3}\right)$.

[61]It follows from the FLRW dynamics that $\frac{|\Omega - 1|}{\Omega} \propto R^{3\gamma-2}(t)$. $\gamma > 2/3$ if the strong energy condition holds, and in that case an initial value of Ω not equal to 1 is driven rapidly away from 1.

fluctuations.[62] There are other puzzling features not related to the seed perturbations. It is not clear, for example, why the baryon-to-photon ratio, relevant to nucleosynthesis calculations, has the particular value it does. (This list could be extended.) The general complaint is that the Standard Model requires a variety of seemingly implausible assumptions regarding the initial state. Why did the universe start off with such a glorious pre-established harmony between causally disjoint regions? How was the initial energy density so delicately chosen that we are still close to the flat model? (And so on.) Although these features are all *possible* according to the Standard Model, the fact that they obtain seems, intuitively, to be incredibly improbable. The Standard Model treats these posits as brute facts not subject to further explanation.

By contrast, Guth proposed to supplement the Standard Model by modifying the very early expansion history of the universe, drawing on ideas in particle physics. Guth proposed that the universe underwent a transient phase of Λ-dominated, exponential expansion at roughly 10^{-35}s. Introducing this inflationary stage eases the conflict between a "natural" or "generic" initial state and the observed universe, in the following sense. Imagine choosing a cosmological model at random from among the space of solutions of EFE. Even without a good understanding of this space of solutions or how one's choice is to be "actualized," it seems clear that one of the maximally symmetric FLRW models must be an incredibly "improbable" choice.[63] New dynamics in the form of inflation makes it possible for "generic" pre-inflationary initial conditions to evolve into the uniform, flat state required by the Standard Model.[64] According to the Standard Model alone, what we observe is incredibly improbable; according to the Standard Model *plus inflation*, what we observe is to be expected.

This is an example of a general strategy, which I will call the "dynamical approach": given a theory that apparently requires special initial conditions, augment the theory with new dynamics such that the dependence on special initial conditions is reduced. McMullin (1993) describes a preference for this approach as accepting an "indifference principle," which states that a theory that is indifferent to the initial state, i.e. robust under changes of it, is preferable to one that requires special initial conditions. Theorists who accept the indifference principle can identify fruitful problems by considering the contrast between a "natural" initial state and the observed universe, and then seek new dynamics to reconcile the two.

This line of reasoning is frequently endorsed as a motivation for inflation in the huge literature

[62] One can evolve observed fluctuations backwards to determine the amplitude of the fluctuation spectrum at a given "initial" time t_i. For t_i on the order of the Planck time, for example, Blau and Guth (1987) calculate that the fluctuations obtained by evolving backwards from the time of recombination implies a density contrast of $\approx 10^{-49}$ at t_i, nine orders of magnitude *smaller* than thermal fluctuations. The comparison depends on the choice of the time t_i: if this is treated as a free variable, then there will be some time at which the fluctuations are comparable to thermal fluctuations.

[63] For any reasonable choice of measure over the space of solutions, these models are presumably a measure-zero subset.

[64] Inflation solves the horizon problem because the horizon distance increases exponentially during inflation; for a sufficiently long period of inflation, all the points . The inflationary phase also reverses the dynamical feature of the FLRW models responsible for the flatness problem. Because $\gamma = 0$ (in the equation in f.n. 61) for most models of inflation, inflationary expansion drives Ω towards 1, enlarging the range of choices $\Omega(t_p)$ compatible with observations.

on the topic following Guth's paper. However, a number of skeptics have challenged the dynamical approach as a general methodology and as a motivation for accepting inflation.[65] One line of criticism concerns whether inflation achieves the stated aim of eliminating the need for special initial conditions, as opposed to merely shifting it to a different aspect of the physics. In effect inflation exchanges the degrees of freedom associated with the spacetime geometry of the initial state for the properties of a field (or fields) driving an inflationary stage. This exchange has obvious advantages if physics can place tighter constraints on the relevant fields than on the initial state of the universe. What is gained, however, if the field (or fields) responsible for inflation have to be in a special state to trigger inflationary expansion, or have to be finely-tuned properties to be compatible with observations?

There are also direct challenges to the dynamical approach itself, sometimes presented in concert with advocacy of an alternative "theory of initial conditions" approach. First, why should we assume that the initial state of the universe is "generic"? Penrose, in particular, has argued that this proposal is not compatible with a neo-Boltzmannian account of the second law of thermodynamics (cf. Albrecht 2004). Penrose (1979) treats the second law as arising from a law-like constraint on the initial state of the universe, requiring that it has low entropy. Rather than introducing a subsequent stage of dynamical evolution that erases the imprint of the initial state, we should aim to formulate a "theory of initial conditions" that accounts for its special features. Second, how should we make sense of the implicit probability judgments employed in these arguments? The assessment of an initial state as "generic," or, on the other hand, as "special," is based on a choice of measure over the allowed initial states of the system. But on what grounds is one measure to be chosen over another? Furthermore, how does a chosen measure relate to the probability assigned to the actualization of the initial state? It is clear that the usual way of rationalizing measures in statistical mechanics, such as appeals to ergodicity, do not apply in this case because the state of the universe does not "sample" the allowed phase space.[66]

Assessing the dynamical approach depends on a number of central issues in philosophy of science. Philosophers steeped in debates regarding scientific explanation may find it exciting to discover a major scientific research program motivated by explanatory intuitions. Proponents of inflation often sound as though their main concern is to make the early universe safe for Reichenbach's principle of the common cause. Or, they emphasize the unification between particle physics and cosmology achieved in their models. While these connections are intriguing, they both must be treated with a grain of salt.[67] A more general question is whether the explanatory intuitions betray an overly strong rationalistic tendency to demand explanations of *everything*. Callender (2004) argues in favor of accepting a posited initial state as a brute fact, in part by showing that purported "explanations" of it are mostly vacuous.[68]

[65] One of the main lines of criticism of inflation is due to Roger Penrose; see Penrose (2004), Chapter 28 for a recent exposition. See Earman and Mosterin (1999) for a philosopher's take on inflation, Linde (2007), for example, for a recent review and Turok (2002) for a critical assessment.

[66] For further discussion, see, e.g., Callender (2004a); Earman (2006); Wald (2006); Wallace (2011).

[67] For further discussion of causality in relation to the horizon problem, see Earman (1995), and for a critical assessment of unification claims see Zinkernagel (2000).

[68] See Price (2004) for a defense of the opposing point of view, in an exchange with Callender (2004).

A quite different approach purports to explain various features of the universe as necessary conditions for our presence as observers, to which we now turn.

7 Anthropic Reasoning

There has been a great deal of controversy regarding anthropic reasoning in cosmology in the last few decades.[69] Weinberg (2007) describes the acceptance of anthropic reasoning as a radical change for the better in how theories should be assessed, comparable to the introduction of symmetry principles. In assessing cosmological theories we need, on this view, to account for selection effects due to our presence as observers and to consider factors such as the number of observers predicted to exist by competing theories. How exactly this is to be done remains a matter of dispute. There is no widely accepted standard account of anthropic reasoning. Critics of this line of thought argue that insofar as anthropic reasoning introduces new aspects of theory assessment, as opposed to merely putting an anthropic gloss on some accepted inductive methodology, it is ill-motivated or even incoherent. A methodology that is itself controversial is not particularly useful in forging consensus, so the articulation and assessment of anthropic reasoning is clearly an essential task. Philosophers have already contributed to this effort and should continue to do so. My aim here is to provide a brief overview of the debate, with an emphasis on connections with the philosophical literature.

Two exemplary cases should suffice to introduce anthropic reasoning. Dirac (1937) noted that various "large numbers" defined in terms of the fundamental constants have the same order of magnitude. This coincidence (and others) inspired his "Large Number Hypothesis": dimensionless numbers constructed from the fundamental constants "are connected by a simple mathematical relation, in which the coefficients are of the order of magnitude unity" (Dirac 1937, p. 323). Since one of these numbers includes the age of the universe t_0, so must they all. This implies time variation of the gravitational "constant" G. Dicke (1961) argued that attention to selection effects undermined the evidential value of this surprising coincidence. Surprise at the coincidence might be warranted if t_0 could be treated as "a random choice from a wide range of possible values" (Dicke 1961, p. 440), but there can only be observers to wonder at the coincidence for some small range of t. Dicke (1961) argued that the value of t must fall within an interval such that Dirac's coincidence automatically holds given two necessary conditions for the existence of observers like us.[70] The evidence allegedly provided by the large number coincidence bears no relation to the truth or falsity of Dirac's hypothesis or the Standard Model.[71] Taking the coincidence as evidence for the large number hypothesis would be as misguided as concluding (recycling Eddington's example) that there

[69] Barrow and Tipler (1986) is an influential early survey of the field; see Carr (2007) for a recent collection of essays.

[70] These necessary conditions are: (1) that main sequence stars are still burning, and (2) that an earlier generation of red giants had time to produce carbon in supernovae.

[71] Bayesians can account for this by explicitly conditionalizing on some claim characterizing the selection effect A: $P_s(\cdot) = P(\cdot|A)$. The selection effect may render an originally "informative" piece of evidence E useless, in that $P_s(E|H) \approx P_s(E|\neg H)$. In these terms, Dicke's argument shows that $P_s(LN|H_D) \approx P_s(LN|H_{SM}) \approx 1$, where LN is the large number coincidence, H_D is Dirac's cosmological theory, and H_{SM} is the Standard Model.

are no fish smaller than 6 inches in a pond based on the absence of such small fish in a fisherman's basket, even though the fisherman's net has gaps too large to hold these fish.

Attention has recently focused on a different kind of anthropic reasoning exemplified by Weinberg (1987)'s prediction for Λ.[72] Just as in Dicke's arguments regarding t, within the Standard Model the value of Λ cannot be freely chosen. Because a Λ term does not dilute with expansion, a cosmological model with $\Lambda > 0$ will transition from matter-dominated to vacuum-dominated expansion. Weinberg showed that structure formation via gravitational enhancement stops in the vacuum-dominated stage. The existence of large gravitationally bound systems (large enough to lead to the formation of stars) then imposes an upper bound on possible values of Λ, keeping other aspects of the Standard Model fixed.[73] It is plausible to take the existence of gravitationally bound systems as a necessary precondition for the existence of observers. There is also a lower bound: a negative Λ term contributes to EFE like normal matter and energy, and adding a large negative Λ term leads to a model that recollapses before there is time for observers to arise.

So far the argument is similar to Dicke's elucidation of anthropic bounds on t. But Weinberg next predicted that Λ's observed value should be close to the mean of the values suitable for life. If we inhabit a "multiverse" in which the value of Λ varies in different regions,[74] the prediction is obtained by using the presence of observers as a selection effect. Weinberg assumed that the probability distribution for values of Λ in the multiverse is uniform within the anthropic bounds, and that we are typical members of the reference class of observers in the universe Vilenkin (1995) calls this the "principle of mediocrity" (PM). In Bayesian terms, an initially flat probability distribution for the value of Λ is turned into a prediction – a sharply peaked distribution around a preferred value – by conditionalizing on the existence of large gravitationally bound systems, serving as a proxy for observers. Each of these assumptions is controversial. I will postpone more detailed discussion of the multiverse until the next section, and take up the PM shortly. The first assumption is often justified by appeals to simplicity or naturalness, but it is on unsure footing without further specification of how the multiverse is generated.[75] Nonetheless, Weinberg's prediction of a positive value of Λ within two orders of magnitude of currently accepted values has been widely cited as a striking success of anthropic reasoning.[76]

[72]This is not to say that Weinberg's paper is the first appearance of this kind of anthropic reasoning in contemporary cosmology; Collins and Hawking (1973) is an earlier influential example, in which similar reasoning is used to account for the isotropy of the universe.

[73]More precisely, the upper bound relates the Λ term to the total energy density in matter at the time when most galaxies formed; the upper bound on Λ is ≈ 200 times the present matter density. Considering variation of multiple parameters may undermine this bound; larger values of Λ can be tolerated if one increases the amplitude of the initial spectrum of density perturbations, for example. See Aguirre (2007) for a discussion of the problems associated with considering a single parameter.

[74]Weinberg (1987) did not base his suggestion on a particular multiverse proposal, instead listing four proposals that would provide a suitable setting for his argument.

[75]There have been calculations for the prior probability distribution over Λ in different proposed multiverses; the assumption holds in some but not all of them (see, e.g., Garriga and Vilenkin 2000).

[76]This is a vast improvement on the estimates produced by particle physics, which are off by up to 120 orders of magnitude. In a later treatment, Weinberg argues for a lower anthropic bound, such that the probability assigned to current observations is either 5 or 12 % (depending on other assumptions); see Weinberg (2007) for an overview and references.

Different views regarding anthropic reasoning can be characterized in part by whether they take Weinberg's argument as a valid extension of Dicke's. Many anthropic skeptics accept Dicke's reasoning but see it as an illustration of how to take selection effects into account, without any truly anthropic elements (e.g., Earman 1987; Smolin 2007). Dicke simply follows through the consequences of the existence of main sequence stars and heavy elements. The nature of "observers" and whether they are typical members of a given reference class play no role. Furthermore, as Roush (2003) emphasizes, Dicke's argument *devalues* a particular body of evidence. The apparent coincidences that troubled Dirac reflect deep biases in the evidence available to us, and as a result have no value in assessing his hypothesis. Weinberg's argument, by contrast, takes the successful "prediction" of a surprising value for a particular parameter as evidence in favor of a multiverse. Thus it is more in line with Dirac's idea that such coincidences can be revealing rather than with Dicke's response. It also depends on assumptions regarding our "typicality" among members of a reference class, raising a number of issues that Dicke's argument avoids. Proponents of anthropic reasoning argue that these issues have to be dealt with in order to assess cosmological theories.

Some have argued that the PM must be assumed in order to extract any predictions at all from cosmological theories that describe an infinite universe.[77] Consider an observation O, for example that the CBR has an average temperature within the observer's Hubble volume of $\mathsf{T} = 3.14159...\mathsf{K}$, in agreement with the decimal expansion of π to some specified number of digits. Suppose we have a cosmological theory T that predicts the existence of an open FLRW model with infinite spatial slices Σ and also assigns a non-zero probability to O. Then there is an observer for whom O is true *somewhere* in the vast reaches of the infinite universe. The point generalizes to other observations and threatens to undermine the use of *any* observations to assess cosmological theories.[78] (This challenge arises even in the Standard Model, provided that the universe is not closed, and does not depend on more speculative multiverse proposals.) This skeptical conclusion can only be evaded by accepting the principle of mediocrity, according to this line of thought: we are interested not in the reports of such improbable "freak observers," but rather in *our* observations — where we regard ourselves as randomly selected from an appropriate reference class. Even "infinite universe" theories can make predictions by employing the PM, once the appropriate reference class has been specified.

The PM leads, unfortunately, to absurd results in other cases. These problems are arguably due to the explicit reliance on the choice of a reference class. This choice does not reflect a factual claim about the world, yet it can lead directly to striking empirical results, as illustrated in the Doomsday Argument (e.g., Leslie 1992, Gott 1993, Bostrom 2002). The argument follows from applying the PM to one's place in human history, in particular by asserting that one should occupy a "typical" birth rank among the reference class consisting of all humans who have ever lived. This implies that there are roughly as many humans born before and after one's own birth. For this to be true, given the current rate of population growth, "doomsday" — a rapid drop in the growth rate of the

[77] See, e.g., Vilenkin (1995), Bostrom (2002).

[78] Obviously this argument requires some assumptions regarding methodology; it is typically formulated within a Bayesian approach, and the conclusion need not follow on other accounts of inductive method. Shortly I will return to the question of whether this is a good argument even on a Bayesian approach.

human population — must be just around the corner.[79] The conclusion of the argument depends critically on the reference class. Starkman and Trotta (2006) argue that Weinberg's prediction of Λ is similarly sensitive to the reference class used in applying the PM.

Philosophers have discussed a number of other cases, from Sleeping Beauties to Presumptuous Philosophers, meant to test principles proposed for anthropic reasoning.[80] Stated more generally, these proposals regard how to incorporate indexical information (about, e.g., one's location in the history of mankind) in evidential reasoning. Straightforward modifications of the PM to avoid the Doomsday argument lead to counter-intuitive results in these other cases. Bostrom (2002) advocates responding to the Doomsday argument by considering a different reference class when applying the PM, but his arguments that there is a unique reference class that resolves the problems are unconvincing. An alternative response is to take the number of observers in the reference class into account, by weighting the prior probability by this number.[81] For example, if a theory predicts that there will be 10^6 more observers (in the appropriate reference class) than a competing theory, then the prior probabilities should have this same ratio. This effectively blocks the Doomsday argument. It has unpalatable consequences of its own, however, if it is taken as a general methodological principle: it implies nearly unshakeable confidence in theories that predict large numbers of observers.[82]

The combined effect of accepting PM and adjusting the priors to take account of the number of observers is to eliminate the dependence on a choice of a particular reference class, as Neal (2006) shows. Rather than introduce the reference class only to eliminate its impact, why not simply apply Bayesian conditionalization? Neal (2006) argues that standard Bayesian conditionalization on *all* non-indexical evidence available resolves the various puzzles associated with anthropic reasoning, with one caveat. On this approach anthropic reasoning is just a species of Bayesian conditionalization, and there is no need to introduce further methodological principles.[83] (It is crucial to conditionalize on *everything* because, as analyses of selection effects like Dicke's show, it is not always transparent which aspects of our evidence are relevant.)

This approach leads to the following assessment of anthropic predictions, such as Weinberg's prediction of Λ. Consider a multiverse theory T_M in which the value of Λ (and perhaps other

---

[79] There are various different formulations of the argument (see Bostrom 2002 for an entry point into this literature). One formulation starts with the assumption that the probability of one's own birth rank being r is given by $\Pr(r|N) = 1/N$, where N is the total number of humans ever born (assuming that $N \geq r$). If one further assigns a prior probability $\Pr(N) = k/N$ (with a constant k), then the posterior probability obtained using Bayes's theorem is $\Pr(N|r) = k/N^2$. It follows that there is a less than 5% probability that the total number of humans ever born will exceed $20r$. The argument is entirely general, and results from invoking the PM in choosing a time within a process that extends over some finite duration.

[80] See Bostrom (2002) and Neal (2006) for discussions of the different versions of "anthropic reasoning" and the various puzzles they are meant to address.

[81] This was proposed by Dieks (1992) in response to the Doomsday argument; see Bostrom (2002) and Dieks (2007) for further discussion. The idea has also been discussed in light of Elga (2000)'s Sleeping Beauty problem.

[82] Hence the Presumptuous Philosopher (see Bostrom 2002), whose posterior probability in the theory with more observers remains high despite receiving disconfirming evidence.

[83] This is not to say that various considerations emphasized in the anthropic literature, such as the number of observers predicted to exist in a particular situation, are irrelevant. Rather, such factors can be accounted for in a Bayesian approach by paying careful attention to the details without adding further general principles.

parameters) takes on different values in different regions, contrasted with a theory T_1 in which the value of Λ is not fixed by theoretical principles, but does not vary in different regions. Suppose that Δ is the range of values of Λ compatible with all available evidence (including, for example, the existence of galaxies at high redshifts), and that according to T_M the fraction of regions with a value of Λ within Δ is given by f,[84] whereas T_1 assigns a probability of g to Δ. If one assigns equal priors to the two theories, the odds ratio for T_M to T_1 upon conditionalization will be given by f/g. The evaluation of the two theories depends on the probability they assign to a value of Λ within Δ. Whether the theory involves a "multiverse" with Λ varying in different regions is irrelevant to the comparison. The assessment also does not depend on considering how Δ compares to Δ', the range of parameter values of Λ compatible with "intelligent life" (or "advanced civilizations," etc.).

The caveat is that this analysis applies to universes in which the evidence is sufficiently rich to single out a unique observer. Neal acknowledges that in an infinite universe the argument above regarding "freak observers" poses a threat, given that there will be multiple observers with the same total body of evidence. He goes on to argue, however, that it is implausible that our evidential reasoning should depend on whether the universe is large enough to contain observers with exactly the same evidence. (This is, of course, exactly the context in which cosmologists feel the need to invoke the PM — see, e.g., Garriga and Vilenkin 2007.)

Philosophers have rejected the use of PM on other grounds. Norton (2010) has challenged the employment of probability distributions as a way of representing neutrality of evidential support, as part of a more general criticism of Bayesianism. He argues that the ability to get something from nothing — a striking empirical result from innocuous assumptions, as in the Doomsday argument — reflects the extra representational baggage associated with describing ignorance using a probability measure. Probability measures are assumed to be countably additive, and this prevents them from expressing complete evidential neutrality. Assigning a uniform prior probability over the values of some parameter such as Λ implies that a value in a finite interval is disfavored by the evidence, rather than treating all of these values neutrally. One might hope that invoking a "random" choice among members of a reference class can underwrite ascriptions of probability. Norton counters that invocations of indifference principles such as PM actually support the ascription of neutral evidential warrant rather than uniform probability.

This brief survey has sketched three different lines of thought regarding anthropic reasoning. The most conservative option is to apply standard Bayesian methodology to cases where anthropic issues arise. The hope is that these cases can be treated by carefully attending to details without introducing new principles of general scope, and without invoking reference classes. One advantage of the conservative position is the availability of arguments in favor of the basic tenets of Bayesianism. It would be surprising if the validity of these methodological principles were in fact sensitive to whether we live in a vast, finite universe or a truly infinite universe. Against the conservatives, Norton directly attacks the use of probability to represent degrees of belief in cases of neutral support, such as undetermined parameters. This general criticism of Bayesianism has implications much broader than anthropic reasoning, but the conclusions it leads to in this case

[84] How to calculate this fraction depends upon the measure assigned over the multiverse, so that one can count regions. Here for the sake of illustration I will simply assume that such a fraction is well defined, and that it yields a finite result.

are similar to those of the conservative Bayesian: a rejection of the need to provide anthropic explanations of particular parameter values. Finally, a third position is that there are important and new methodological principles required to handle indexical information and selection effects. One goal of such an account would be to clarify this style of reasoning, which is widely employed within contemporary cosmology. What is lacking so far, in my view, is a compelling account of what these principles are and a motivation for accepting them.

8 Multiverse

Anthropic reasoning is often discussed in tandem with the multiverse. Weinberg's anthropic prediction for Λ is based on applying a selection effect to a multiverse in which the value of Λ varies in different regions. The multiverse idea has gained traction in part because Weinberg's approach is widely regarded as the only viable solution to the cosmological constant problem, and other similar problems may also admit only anthropic solutions.

Two different lines of thought in physics also support the introduction of the multiverse. First, within inflationary cosmology the same mechanism that produces a uniform, homogeneous universe on scales on the order of the Hubble radius leads to a dramatically different global structure of the universe. Inflation is said to be "generically eternal" in the sense that inflationary expansion continues in different regions of the universe, constantly creating bubbles such as our own universe, in which inflation is followed by reheating and a much slower expansion.[85] The individual bubbles are effectively causally isolated from other bubbles, and are often called "pocket universes." The second line of thought relates to the proliferation of vacua in string theory. Many string theorists now expect that there will be a vast landscape of allowed vacua, with no way to fulfill the original hope of finding a unique compactification of extra dimensions to yield low-energy physics.

Both of these developments suggest treating the low-energy physics of the observed universe as partially fixed by parochial contingencies related to the history of a particular pocket universe. Other regions of the multiverse may have drastically different low-energy physics because, for example, the inflaton field tunneled into a local minima with different properties.[86] Here my main focus will be on a philosophical issue that is relatively independent of the details of implementation: in what sense does the multiverse offer satisfying explanations?

But, first, what do we mean by a "multiverse" in this setting?[87] These lines of thought lead to a multiverse with two important features. First, it consists of causally isolated pocket universes, and second, there is significant variation from one pocket universe to another. There are other ideas of a multiverse, such as an ensemble of distinct possible worlds, each in its own right a topologically connected, maximal spacetime, completely isolated from other elements of the ensemble. But in

[85]Note that arguments to this effect usually involve a lot of hand-waving.

[86]The Everett interpretation of quantum mechanics attributes a branching structure to the universal wave function of the universe, and the individual branches can be regarded as something akin to pocket universes (see Wallace, this volume, for a discussion of the Everett interpretation). However, unlike the other accounts the laws of physics do not vary in the different branches. There is a clear distinction between the two cases, although recently there has been interest in exploring connections between these two lines of thought.

[87]See Tegmark (2009) for an influential classification of four different types or levels of the multiverse.

contemporary cosmology, the pocket universes are all taken to be effectively causally isolated parts of a single, topologically connected spacetime — the multiverse. Such regions also occur in some cosmological spacetimes in classical GR. In De Sitter spacetime, for example, there are inextendible timelike geodesics γ_1, γ_2 such that $J^-(\gamma_1)$ does not intersect $J^-(\gamma_2)$. In cases like this the definition of "effectively causally isolated" can be cashed out in terms of relativistic causal structure, but for a quantum multiverse the definition needs to be amended.

The example of pocket universes within De Sitter spacetime lacks the second feature, variation from one pocket universe to another. This can take several forms, from variation in the constants appearing in the Standard Models of cosmology and particle physics to variation of the laws themselves. Within the context of eternal inflation or the string theory landscape, what were previously regarded as "constants" may instead be fixed by the dynamics. For example, Λ is often treated as the consequence of the vacuum energy of a scalar field displaced from the minimum of its effective potential. The variation of Λ throughout the multiverse may then result from the scalar field settling into different minima. Greater diversity is suggested by the string theory landscape, according to which the details of how extra dimensions are compactified and stabilized are reflected in different low-energy physics.

In the multiverse some laws will be demoted from universal to parochial regularities. But presumably there are still universal laws that govern the mechanism that generates pocket universes. This mechanism for generating a multiverse with varying features may be a direct consequence of an aspect of a theory that is independently well-tested. Rather than treating the nature of the ensemble as speculative or conjectural, one might then have a sufficiently clear view of the multiverse to calculate probability distributions of different observables, for example. In this case, there is a direct reply to multiverse critics who object that the idea is "unscientific" because it is "untestable": other regions of the multiverse would then have much the same status as other unobservable entities proposed by empirically successful theories.[88] Unfortunately for fans of the multiverse, the current state of affairs does not seem so straightforward. Although multiverse proposals are motivated by trends in fundamental physics, the detailed accounts of how the multiverse arises are typically beyond theoretical control. As long as this is the case, there is a risk that the claimed multiverse explanations are just-so stories where the mechanism of generating the multiverse is contrived to do the job. This strikes me as a legitimate worry regarding current multiverse proposals, but I will set this aside for the sake of discussion.

Suppose, then, that we are given a multiverse theory with an independently motivated dynamical account of the mechanism churning out pocket universes. What explanatory questions might this theory answer, and what is the relevance of the existence of the multiverse itself to its answers?[89] Here we can distinguish between two different kinds of questions. First, should *we* be surprised to measure a value of a particular parameter X (such as Λ) to fall within a particular range? Our surprise ought to be mitigated by a discussion of anthropic bounds on X, revealing various unsuspected connections between our presence and the range of allowed values for the parameter in question. But, as with Dicke's approach discussed above, this explanation can be taken

[88]This line of argument has appeared numerous times in the literature; see, e.g., Livio and Rees (2005) for a clear formulation.

[89]Here I am indebted to discussions with John Earman.

to demystify the value of X without also providing evidence for a multiverse. The value of this discussion lies in tracing the connections between, e.g., the time-scale needed to produce carbon in the universe or the constraints on expansion rate imposed by the need to form galaxies. The existence of a multiverse is irrelevant to this line of reasoning.

A second question pertains to X, without reference to our observation of it: why does the value of X fall within some range in a particular pocket universe? The answer to this question offered by a multiverse theory will apparently depend on contingent details regarding the mechanism which produced the pocket universe. This explanation will be *historical* in the sense that the observed values of the parameter will ultimately be traced back to the mechanism that produced the pocket universe.[90] It may be surprising that various features of the universe are given this type of explanation rather than following as necessary consequences of fundamental laws. However, the success of historical explanations does not support the claim that other pocket universes must exist. Analogously, the success of historical explanations in evolutionary biology does not imply the existence of other worlds where pandas have more elegant thumbs.

To put the point in a slightly different form, the value of converting questions about modalities in cosmology into questions about location within a vastly enlarged ontology is not clear. Both types of questions can apparently be answered adequately without making the further ontological commitment to the actual existence of other pocket universes.

9 Conclusion

One theme running through the discussion above is the attempt to identify distinctive evidential challenges faced in cosmology. There is an echo of skepticism regarding the possibility of knowledge of the universe-as-a-whole in the discussion of global properties of the universe (§5). Local observations are not sufficient to warrant conclusions regarding global properties without help from general principles like the cosmological principle, which is itself on unsure footing. This does not, however, support a general skepticism about cosmology. Most contemporary research in cosmology is compatible with agnosticism regarding the global properties of the universe. The challenges arise, not from the limits imposed by the causal structure of GR, but from the difficulty in gaining access to the relevant phenomena via independent routes. As the discussion in (§3) illustrates, assuming that the Standard Model is basically correct makes it possible to infer the presence of dark matter and dark energy. It is difficult to rule out the possibility that the same observations used as the basis for this inference instead reveal flaws in the Standard Model. Yet this does not mean that all the responses to the observations should be given equal credence. Philosophers of science ought to offer an account of empirical support that clarifies the assessment of different responses. Regarding early universe cosmology (§6), the theory being used to describe the underlying physics is tested through its role in providing a reconstruction of the universe's history. The field has been partially driven by strong explanatory intuitions favoring a theory that renders the observed history probable or expected, although it is unclear how to move beyond intuitive discussions of probability.

[90] The explanation may also be path-dependent in the sense of depending not just on an initial state, but on various stochastic processes leading to the formation of the pocket universe.

Cosmologists have to face the possibility that the data they use to assess theories is subject to unexpected anthropic selection effects (§7). Whether these selection effects can be treated within standard approaches to confirmation theory or require new principles of anthropic reasoning is currently being debated. Finally, cosmologists may also see their explanatory aims change, with various features of the universe traced to environmental features of our pocket universe rather than being derived from dynamical laws (§8).

References

Aguirre, A. (2007). "Making predictions in a multiverse: conundrums, dangers, coincidences." *Universe or Multiverse?* 1, 367.

Albrecht, A. (2004). "Cosmic inflation and the arrow of time." *Science and Ultimate Reality* , 363–401.

Alpher, Ralph A., Bethe, Hans and Gamow, George (1948). "The origin of chemical elements." *Physical Review* 73, 803–804.

Balashov, Y. (2002). "Laws of physics and the universe." *Einstein Studies in Russia. Boston, MA/Basel/Berlin: Birkhäuser* , 107–148.

Barrow, John D. and Tipler, Frank J. (1986). *The anthropic cosmological principle*. Oxford: Oxford University Press.

Beisbart, C. (2009). "Can we justifiably assume the cosmological principle in order to break model underdetermination in cosmology?" *Journal for general philosophy of science* 40, 175–205.

Bekenstein, J. D. (2010). "Alternatives to dark matter: Modified gravity as an alternative to dark matter." *ArXiv e-prints* `1001.3876`.

Bertone, Gianfranco, Hooper, Dan and Silk, Joseph (2005). "Particle dark matter: Evidence, candidates and constraints." *Phys. Rept.* 405, 279–390. `hep-ph/0404175`.

Blau, Steven K. and Guth, Alan (1987). "Inflationary cosmology." In *300 years of gravitation*, S. W. Hawking and Werner Israel, eds. Cambridge: Cambridge University Press, 524–603.

Bostrom, Nick (2002). *Anthropic Bias: Observation Selection Effects in Science and Philosophy*. New York: Routledge.

Caldwell, RR and Stebbins, A. (2008). "A test of the copernican principle." *Physical review letters* 100, 191302.

Callender, C. (2004a). "Measures, explanations and the past: Should specialinitial conditions be explained?" *The British journal for the philosophy of science* 55, 195–217.

— (2004b). "There is no puzzle about the low-entropy past." *In: Contemporary Debates in Philosophy of Science, Oxford* , 240–256.

Callender, Craig and Huggett, Nick, eds. (2001). *Philosophy Meets Physics at the Planck Scale*. Cambridge University Press.

Carr, B. (2007). *Universe or multiverse?*. Cambridge Univ Pr.

Carrera, M. and Giulini, D. (2010). "Influence of global cosmological expansion on local dynamics and kinematics." *Reviews of Modern Physics* 82, 169.

Clarkson, C. and Maartens, R. (2010). "Inhomogeneity and the foundations of concordance cosmology." *Classical and Quantum Gravity* 27, 124008.

Clifton, T., Clarkson, C. and Bull, P. (2011). "The isotropic blackbody cmb as evidence for a homogeneous universe." *Arxiv preprint arXiv:1111.3794* .

Coles, Peter and Ellis, G. F. R. (1997). *Is the universe open or closed?*. Cambridge: Cambridge University Press.

Collins, C. B. and Hawking, S. W. (1973). "Why is the universe isotropic?" *Astrophysical Journal* 180, 317–334.

Cooperstock, FI and Tieu, S. (2007). "Galactic dynamics via general relativity: a compilation and new developments." *International Journal of Modern Physics A* 22, 2293–2326.

Curiel, Erik N. (1999). "The analysis of singular spacetimes." *Philosophy of Science* 66, S119–S145.

Dicke, Robert (1961). "Dirac's cosmology and Mach's principle." *Nature* 192, 440–441.

Dieks, Dennis (1992). "Doomsday, or: the dangers of statistics." *Philosophical Quarterly* 42, 78–84.

— (2007). "Reasoning about the future: Doom and beauty." *Synthese* 156, 427–439. ISSN 00397857. URL `http://www.jstor.org/stable/27653528`.

Dirac, Paul A. M. (1937). "The cosmological constants." *Nature* 139, 323.

Dodelson, S. (2003). *Modern cosmology*. Academic Pr.

Earman, J. (2006). "The past hypothesis: Not even false." *Studies in History and Philosophy of Modern Physics* 37, 399–430.

Earman, John (1987). "The SAP also rises: A critical examination of the anthropic principle." *American Philosophical Quarterly* 24, 307–317.

— (1995). *Bangs, Crunches, Whimpers, and Shrieks*. Oxford: Oxford University Press.

— (2001). "Lambda: The constant that refuses to die." *Archive for History of Exact Sciences* 55, 189–220.

Earman, John and Mosterin, Jesus (1999). "A critical analysis of inflationary cosmology." *Philosophy of Science* 66, 1–49.

Earman, John and Roberts, John (1999). "*Ceteris Paribus*, there is no problem of provisos." *Synthese* 118, 439–478.

Ehlers, J., Geren, P. and Sachs, R. K. (1968). "Isotropic solutions of the Einstein–Liouville equations." *Journal of Mathematical Physics* 9, 1344–1349.

Einstein, Albert (1917). "Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie." *Preussische Akademie der Wissenschaften (Berlin). Sitzungsberichte* , 142–152.

Elga, Adam (2000). "Self-locating belief and the sleeping beauty problem." *Analysis* 60, 143–147.

Ellis, G. F. R. (1971). "Topology and cosmology." *General Relativity and Gravitation* 2, 7–21.

— (1980). "Limits to verification in cosmology." In *9th Texas Symposium on Relativistic Astrophysics*, vol. 336 of *Annals of the New York Academy of Sciences*. 130–160.

— (2007). "Issues in the philosophy of cosmology." In *Handbook for the Philosophy of Physics*, J. Earman and J. Butterfield, eds., vol. Part B. Elsevier, 1183–1286.

Ellis, G. F. R., Maartens, R. and Nel, S. D. (1978). "The expansion of the universe." *Monthly Notices of the Royal Astronomical Society* 184, 439–465.

Ellis, G. F. R., Nel, S. D., Maartens, R., Stoeger, W. R. and Whitman, A. P. (1985). "Ideal observational cosmology." *Physics Reports* 124, 315–417.

Ellis, G. F. R. and Rothman (1993). "Lost horizons." *American Journal of Physics* 61, 883–893.

Ellis, G. F. R. and Sciama, D. W. (1972). "Global and non-global problems in cosmology." In *General Relativity: Papers in Honour of J. L. Synge*, L. O'Raifeartaigh, ed. Oxford: Clarendon Press, 35–59.

Ellis, G.F.R. (1999). "Before the beginning: Emerging questions and uncertainties." *Astrophysics and Space Science* 269, 691.

— (2011). "Inhomogeneity effects in cosmology." *Classical and Quantum Gravity* 28, 164001.

Frieman, J., Turner, M. and Huterer, D. (2008). "Dark energy and the accelerating universe." *Arxiv preprint arXiv:0803.0982* .

Garriga, Jaume and Vilenkin, Alexander (2008). "Prediction and explanation in the multiverse." *Phys. Rev.* D77, 043526. `0711.2559`.

Geroch, Robert, Can-bin, L. and Wald, Robert (1982). "Singular boundaries of space-times." *Journal of Mathematical Physics* 23, 432–35.

Geroch, Robert and Horowitz, Gary (1979). "Global structure of spacetimes." In Hawking and Israel (1979), 212–293.

Glymour, Clark (1977). "Indistinguishable space-times and the fundamental group." In *Foundations of Space-Time Theories*, John Earman, Clark Glymour and John Statchel, eds., vol. VIII of *Minnesota Studies in the Philosophy of Science*. University of Minnesota Press, 50–60.

Gott, J.R. (1993). "Implications of the copernican principle for our future prospects." *Nature* 363, 315–319.

Guth, Alan (1981). "Inflationary universe: A possible solution for the horizon and flatness problems." *Physical Review D* 23, 347–56.

Hawking, Stephen and Israel, Werner, eds. (1979). *General Relativity: An Einstein Centenary Survey*. Cambridge: Cambridge University Press.

Hubble, E. (1929). "A relation between distance and radial velocity among extra-galactic nebulae." *Proceedings of the National Academy of Sciences* 15, 168–173.

Huterer, Dragan (2010). "The Accelerating Universe." `1010.1162`.

Jeans, J.H. (1902). "The stability of a spherical nebula." *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 199, 1–53.

Kirshner, R.P. (2009). "Foundations of supernova cosmology." *Arxiv preprint arXiv:0910.0257* .

Kragh, Helge (1996). *Cosmology and Controversy*. Princeton: Princeton University Press.

Krauss, Lawrence M. and Turner, Michael S. (1999). "Geometry and destiny." *General Relativity and Gravitation* 31, 1453–1459. `astro-ph/9904020`.

Lachieze-Rey, M. and Luminet, J.P. (1995). "Cosmic topology." *Physics Reports* 254, 135–214.

Leslie, John (1992). "Doomsday revisited." *The Philosophical Quarterly* 42, 85–89. URL `http://www.jstor.org/stable/2220451`.

Li, A. (2003). "Cosmic needles versus cosmic microwave background radiation." *The Astrophysical Journal* 584, 593.

Lifshitz, Yevgeniy M. (1946). "On the gravitational stability of the expanding universe." *Journal of Physics USSR* 10, 116–129.

Livio, M. and Rees, M. (2005). "Anthropic reasoning." *Science* 309, 1022–1023.

Longair, Malcolm (2006). *The Cosmic Century*. Cambridge: Cambridge University Press.

Malament, David (1977). "Obervationally indistinguishable space-times." In *Foundations of Space-Time Theories*, John Earman, Clark Glymour and John Statchel, eds., vol. VIII of *Minnesota Studies in the Philosophy of Science*. University of Minnesota Press, 61–80.

Manchak, J.B. (2009). "Can we know the global structure of spacetime?" *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics* 40, 53–56.

McMullin, Ernan (1993). "Indifference principle and anthropic principle in cosmology." *Studies in the History and Philosophy of Science* 24, 359–389.

Milgrom, M. (1983). "A modification of the Newtonian dynamics: implications for galaxy systems." *Astrophys. J.* 270, 384–389.

Munitz, Milton K. (1962). "The logic of cosmology." *British Journal for the Philosophy of Science* 13, 34–50.

North, John D. (1965). *The Measure of the Universe*. Oxford: Oxford University Press.

Norton, J.D. (2010). "Cosmic confusions: Not supporting versus supporting not." *Philosophy of Science* 77, 501–523.

— (2011). "Observationally indistinguishable spacetimes: A challenge for any inductivist." *Philosophy of Science Matters: The Philosophy of Peter Achinstein* , 164.

Olive, K.A., Steigman, G. and Walker, T.P. (2000). "Primordial nucleosynthesis: Theory and observations." *Physics Reports* 333, 389–407.

Ostriker, J. P. and Steinhardt, P. J. (1995). "The observational case for a low-density Universe with a non-zero cosmological constant." *Nature* 377, 600–602.

Ostriker, J.P. and Peebles, P.J.E. (1973). "A numerical study of the stability of flattened galaxies: or, can cold galaxies survive?" *The Astrophysical Journal* 186, 467–480.

Pauri, M. (1991). "The universe as a scientific object." In *Philosophy and the Origin and Evolution of the Universe*, E. Agazzi and A. Cordero, eds. Dordrecht: Kluwer Academic Publishers, 291–339.

Peebles, Phillip James Edward (1993). *Principles of Physical Cosmology*. Princeton: Princeton University Press.

Penrose, R. (2004). *The road to reality*. London: Jonathan Cape.

Penrose, Roger (1979). "Singularities and time-asymmetry." In Hawking and Israel (1979), 581–638.

Polchinski, J. (2006). "The cosmological constant and the string landscape." *Arxiv preprint hep-th/0603249* .

Price, H. (2004). "On the origins of the arrow of time: Why there is still a puzzle about the low entropy past." *Contemporary Debates in the Philosophy of Science. Blackwell* .

Rindler, Wolfgang (1956). "Visual horizons in world models." *Monthly Notices of the Royal Astronomical Society* 116, 662–677.

Roush, S. (2003). "Copernicus, Kant, and the anthropic cosmological principles." *Studies In History and Philosophy of Modern Physics* 34, 5–35.

Rugh, S.E. and Zinkernagel, H. (2002). "The quantum vacuum and the cosmological constant problem." *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics* 33, 663–705.

— (2009). "On the physical basis of cosmic time." *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics* 40, 1–19.

Sanders, R. H. and McGaugh, S. S. (2002). "Modified Newtonian Dynamics as an Alternative to Dark Matter." *Anual Reviews of Astronomy and Astrophysics* 40, 263–317. `arXiv:astro-ph/0204521`.

Saunders, S. (2002). "Is the Zero-Point Energy Real?" In *Ontological Aspects of Quantum Field Theory*, Meinard Kuhlmann, Holger Lyre and Andrew Wayne, eds. Singapore: World Scientific, 313–343.

Scheibe, E. (1991). "General laws of nature and the uniqueness of the universe." In *Philosophy and the Origin and Evolution of the Universe*, E. Agazzi and A. Cordero, eds. Dordrecht: Kluwer Academic Publishers, 341–360.

Smeenk, Chris (2012). "Einstein's role in the creation of relativistic cosmology." In *Cambridge Companion to Einstein*, Michel Janssen and Christoph Lehner, eds. Cambridge: Cambridge University Press.

Smith, Sheldon (2002). "Violated laws, *ceteris paribus* clauses, and capacities." *Synthese* 130, 235–264.

Smolin, L. (2007). "Scientific alternatives to the anthropic principle." *Universe or multiverse* , 323–366.

Smolin, Lee (2000). "The present moment in quantum cosmology: Challenges to the arguments for the elimination of time." In *Time and the Instant*, R. Durie, ed. Manchester: Clinamen Press, 112–143.

Sotiriou, T.P. and Faraoni, V. (2010). "f (r) theories of gravity." *Reviews of Modern Physics* 82, 451.

Starkman, Glenn D. and Trotta, Roberto (2006). "Why anthropic reasoning cannot predict Lambda." *Phys. Rev. Lett.* 97, 201301. `astro-ph/0607227`.

Stoeger, W. R., Maartens, R. and Ellis, G. F. R. (1995). "Proving almost-homogeneity of the universe: an almost Ehlers-Geren-Sachs theorem." *Astrophysical Journal* 443, 1–5.

Tegmark, M. (2009). "The multiverse hierarchy." *Arxiv preprint arXiv:0905.1283* .

Torretti, Roberto (2000). "Spacetime models of the world." *Studies in the History and Philosophy of Modern Physics* 31, 171–186.

Trimble, V. (1987). "Existence and nature of dark matter in the universe." *Annual Review of Astronomy and Astrophysics* 25, 425–472.

Turok, N. (2002). "A critical review of inflation." *Classical and Quantum Gravity* 19, 3449.

Uzan, J. P. (2010). "Dark energy, gravitation, and the Copernican principle." *Dark Energy: Observational and Theoretical Approaches*, ed. by P. Ruiz-Lapuente. Cambridge: Cambridge University Press, 3–47.

Uzan, J.P., Clarkson, C. and Ellis, G.F.R. (2008). "Time drift of cosmological redshifts as a test of the copernican principle." *Physical review letters* 100, 191303.

Vanderburgh, W.L. (2003). "The dark matter double bind: Astrophysical aspects of the evidential warrant for general relativity*." *Philosophy of Science* 70, 812–832.

— (2005). "The methodological value of coincidences: Further remarks on dark matter and the astrophysical warrant for general relativity." *Philosophy of Science* 72, 1324–1335.

Vilenkin, A. (1995). "Predictions from quantum cosmology." *Physical Review Letters* 74, 846–849.

Wald, R.M. (2006). "The arrow of time and the initial conditions of the universe." *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics* 37, 394–398.

Wald, Robert (1984). *General Relativity*. Chicago: University of Chicago Press.

Wallace, D. (2011). "The logic of the past hypothesis." .

Weinberg, Steven (1972). *Gravitation and Cosmology*. New York: John Wiley & Sons.

— (1987). "Anthropic Bound on the Cosmological Constant." *Physical Review Letters* 59, 2607.

— (1989). "The cosmological constant problem." *Reviews of Modern Physics* 61, 1–23.

— (2007). "Living in the multiverse." *Universe or multiverse* , 29–42.

— (2008). *Cosmology*. Oxford University Press, USA.

Wolf, J.A. (2011). *Spaces of constant curvature*. 6th ed. Providence, RI: American Mathematical Society.

Zinkernagel, Henrik (2002). "Cosmology, particles, and the unity of science." *Studies in the History and Philosophy of Modern Physics* 33, 493–516.

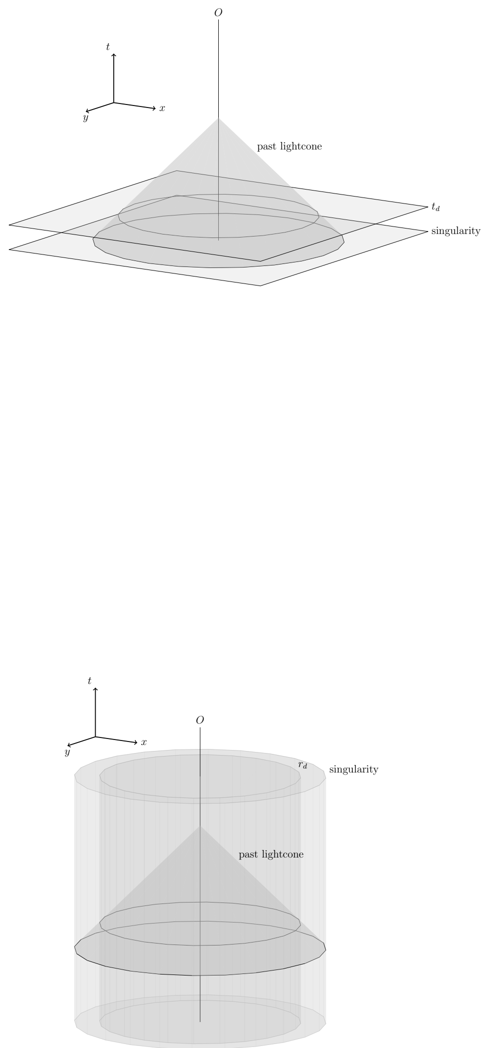

Figure 1: This figure contrasts the standard big bang model (a) and Ellis et al. (1978)'s model (b); in the latter, a cylindrical timelike singularity surrounds an observer O located near the axis of symmetry, and the constant time surface t_{D} from which the CBR is emitted in the standard model is replaced with a surface r_{D} at fixed distance from O.

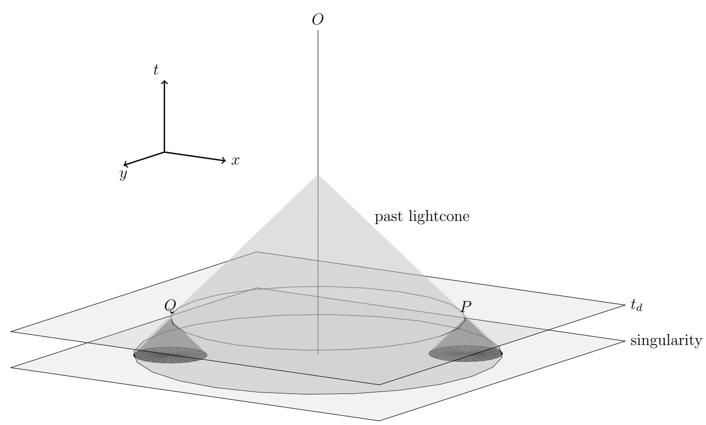

Figure 2: This figure illustrates the horizon problem. Lightcones are at 45° but distances are distorted, much like a Mercator projection. Two points P, Q on the surface of last scattering $\mathtt{t_d}$, both falling within our past light cone, do not have overlapping light cones.