Abstract

Stein has characterized one of the central problems in accounting for our knowledge in physics as that of getting the laboratory, or observatory, inside the theory — that is, of understanding how the mathematical structures of fundamental physical theories have empirical content. He has argued that physicists respond to this problem by giving schematic representations of observers and experiments. In addition, Stein emphasizes the importance of regarding knowledge as an enterprise, with current theories providing guidance for future inquiry. I will explore some ramifications of this way of thinking about the structure of scientific theories for contemporary cosmology. One goal of observational cosmology is to measure the six basic parameters appearing in the standard model of cosmology. These parameters are well-defined if the universe is suitably approximated at some scale by a perturbed FLRW model. The enormous extrapolations involved in the standard model are often justified by the consistent determination of these parameters via a variety of methods. Here I will consider two recent debates regarding this approach to cosmology, inspired by Stein's work. The first debate regards the impact of different ways of characterizing the propagation of light through a cosmological spacetime on the determination of cosmological parameters (such as H_0). The second regards how the highly symmetric FLRW models relate to describing the real universe, at small scales where it is very lumpy.

Some Reflections on the Structure of Cosmological Knowledge

Cosmologists represent the spacetime geometry of the universe at the largest scales, and its evolution, with simple expanding universe models. Given their essential role in our current understanding of the universe, these models should qualify as part of cosmological knowledge, if anything does. In what sense do they provide knowledge of spacetime geometry and dynamics at the largest scales? More precisely, following Stein (1994) in distinguishing three different aspects of our knowledge, in what sense do the expanding universe models, also known as the Friedmann-Lemaître-Robertson-Walker (FLRW) models, constitute a set of systematic claims about a domain of phenomena, amenable to justification, and able to guide further inquiry?

But perhaps I must first address a prior question: is such knowledge even possible? Kant, and other philosophers, have attempted to establish limitations on what we can know about the universe as a whole. Torretti (2000)'s lucid discussion of the discovery of expanding universe models, for example, ends on a sharply critical note, comparing contemporary cosmology unfavorably to other areas of physics. Torretti's main reason for assigning a low "final grade," and a common starting point for philosophical critiques, is that cosmology has an unusual subject matter: the unique universe.[1] Torretti holds that the methodology of physics, appropriate for developing mathematical models of subsystems of the universe, does not extend to theories of the universe-as-a-whole. In my view, Torretti et al. overstate the significance of this challenge. Cosmologists have had considerable success treating the universe at large scales as, in Bondi's apt formulation, "the largest workshop in which we may assemble equipment, the elements of which are entirely composed of terrestrially verified laws of physics" (Bondi, 1960, p. 4). This leads to models of sub-systems of the universe, just as in other areas of physics, albeit of an unusual kind: the expanding universe models describe, not "totality," but the large-scale geometrical features of spacetime and their dynamics. This is not to deny that cosmology differs from other areas of physics. With regard to some questions, such as those regarding the "initial state" or the "origin of the universe," the small sample size does pose challenges. Yet such differences do not support a general argument that has any bearing on the status of the expanding universe models. Cosmology bolsters Stein (1958)'s remark: "philosophical refutations of the possibility of science accounting for some domain of phenomena have notoriously been the preludes of scientific conquest" (p. 13).

The main challenge to using expanding universe models for further conquest resembles that facing other dynamical theories in physics. The FLRW models describe an extremely simple, symmetric spacetime geometry, a universe filled uniformly with matter and devoid of landmarks of any sort. Yet astronomers observe complicated structures, over a wide range of scales, with large variations in density. How can we discern a simple global geometry in the rich complexity of real phenomena, and what justifies favoring FLRW models over alternatives? Furthermore, in what sense can the assumption that the FLRW models apply guide further inquiry effectively?

My title alludes to (Stein, 1994), where (among other things) Stein analyzes similar questions from an earlier era, namely how Newton's *Principia* gives us a "theory of a mathematical struc-

[1] Torretti states a further, related criticism: cosmologists must employ quantum theory and particle physics (in describing the early universe, for example), despite their "logical incompatibility" with general relativity. He acknowledges that we often combine incompatible theories in practice, taking into account domains of applicability and margins of error. For reasons that are obscure to me, he comments that: "I doubt, however, this method [of dealing with hybrid theories] can lead to interesting results about the totality of things" (Torretti, 2000, p. 182).

ture discernible in the world of phenomena, of observations, of experience" (p. 393). Based on this and other exemplary historical cases, Stein develops a distinctive account of the structure of scientific theories, and the status of our knowledge of physics (see also Stein, 1989, 1992). There are two aspects of Stein's position I take to be particularly valuable. The first regards the relationship between the mathematical formulation of a theory and its observational or experimental content. Stein proposes that connecting theory and observation requires representing an observer schematically within a given theory. The aim is not to "deduce" an observation, as Carnap would have it, but to give an account of how an observer's experience relates to the fundamental quantities and dynamics introduced by a theory — with the observer, in effect, modeled as a "measurement device" interacting with other systems.[2] Determining a theory's content is treated as a problem of applied mathematics and modeling rather than logic. Second, theories provide guidance for developing more detailed descriptions of systems falling within their domains, in light of ongoing experiments and observations, and they should be evaluated based on whether they can be consistently and fruitfully applied as inquiry proceeds (Stein, 1989, 2014). These two aspects of Stein's position are essential to prospective evaluation of contemporary science, in addition to guiding retrospective reconstructions such as his seminal treatments of Newton.

Cosmologists sometimes treat the striking ability of the standard model to accomodate a wide variety of data, with only six parameters, as the strongest rationale for the model. The paper begins in §1 with a brief review of the ΛCDM cosmological model. The model's consistency with observations is impressive, but success comes at the cost of introducing dark matter and dark energy. I discuss three alternative justifications for employing FLRW models in §2, and argue that the most persuasive case for taking the FLRW models to accurately represent spacetime geometry combines CMBR observations with the Copernican principle. Despite this empirical argument that the FLRW models apply to the early universe, the physical status of these models is still ambiguous (§3). How do late universe measurements of cosmological quantities relate to the parameters appearing in the ΛCDM model? §4 considers one line of work, regarding the optical properties of "lumpy" models, that ends up supporting the use of FLRW models as reliable approximations. §5 considers how the "backreaction" of inhomogeneities effects the inferences cosmologists have drawn based on observations, such as that of accelerated expansion. These issues have been discussed extensively by cosmologists, but it is hard to find a place for such debates in many philosophical accounts of the structure and content of theories. I aim to show below how Stein's approach to understanding the empirical content of theories leads to a much more compelling account of how scientists discern the FLRW models in experience and use them to guide the ongoing construction of cosmology's standard model.

1 The Search for Six Numbers

Sandage characterized his observational program as the search for just two numbers (Sandage, 1961, 1970): the Hubble constant H_0 and the deceleration parameter q_0, determining the "best-fit" FLRW model (assuming that $\Lambda = 0$). Although the complexity and diversity of observations taken into account has increased dramatically, observational cosmologists still seek optimal parameter values, now for the ΛCDM model. (The name refers to two contributions to mass-energy posited by the model: non-zero cosmological constant Λ, and *c*old *d*ark *m*atter.) Despite

[2]Despite sympathy with many of Carnap's views, Stein rejects his proposal that theoretical and experimental physics can be treated within a single formal language. If such an overarching language existed, the content of a theory could be characterized by its deductive consequences within the "observational part." Yet, as Stein notes, nothing like this appears in journal articles or physics textbooks – "I cannot think of any case in which one can honestly *deduce* what might honestly be called an observation" (Stein, 1992, p. 290, original emphasis). Stein qualifies this criticism as a *de facto* claim about the character of physics, rather than one that follows from some basic principle (see also Stein, 1994). An "honest deduction" would be enormously complex, requiring a synthesis of several different subfields; even if that possibility might some day be realized, it would be remote from scientific practice.

some differences in detail, observational programs typically fix optimal values for 6-9 parameters (see, for example, Aghanim et al., 2018; Tanabashi et al., 2018). We cannot evaluate the results of this search for six(-ish) numbers without first analyzing the framework used to constrain parameters based on observations.

This framework makes two main assumptions. First, gravity, as described by general relativity, governs the evolution of the universe at the largest scales. This is an enormous extrapolation of general relativity, extending beyond the solar-system scales where it is subject to high-precision tests by roughly 14 orders of magnitude. Second, and more significantly, the framework relies on a specific class of solutions to Einstein's field equations (EFE): it takes perturbed FLRW models to describe the universe at large length scales.[3] Observational campaigns target parameter values defined as features of perturbed-FLRW models. The physical significance of these parameter values rests on the assumption that the spacetime geometry of these models accurately represents the universe.

Friedmann and Lemaître found these exact solutions to EFE by imposing symmetries to simplify the mathematics: specifically, *isotropy* and *homogeneity*.[4] Roughly put, isotropy holds if there are no geometrically preferred spatial directions at any given point, whereas homogeneity requires that points on spatial hypersurfaces representing events at the same cosmic time "look the same" geometrically. The FLRW models have a particularly simple structure: spacetime consists of a collection of three-dimensional hypersurfaces of constant cosmic time $\Sigma(t)$ (topologically, $\Sigma \times \mathbb{R}$). Einstein's equations usually lead to a set of coupled partial differential equations, but for the FLRW models the dynamics is fully captured by two ordinary differential equations for the scale factor $R(t)$, the Friedmann equation and the isotropic form of the Raychaudhuri equation. The scale factor represents the spatial distance in Σ between nearby observers moving along timelike geodesics. In an expanding universe, for example, $R(t)$ increases as freely falling bodies move apart. The curvature of the FLRW model and the equation of state of its material constituents, including a cosmological constant, Λ, determine the evolution of $R(t)$.

The FLRW models alone describe a universe without any structure, such as galaxies or clusters of galaxies. The current ΛCDM model accounts for the formation of such large-scale structures by introducing small perturbations away from strict uniformity in the early universe. Gravity enhances density contrasts, so small initial perturbations away from uniformity grow with time. For sufficiently small fluctuations, this evolution can be described in terms of linear perturbations to a background cosmological model. The dynamics of the evolution of these small fluctuations follows from EFE. Eventually, of course, the density contrast grows large enough that linearized perturbation theory fails to apply. A full account of structure formation for later epochs, down to the length scales of galaxies, requires going beyond linear perturbation theory, as well as incorporating a variety of physical effects in addition to gravity.

The assumption that a perturbed FLRW model approximates the real universe makes it possible to bring several distinct lines of evidence to bear on the model's free parameters. These parameters fall into three broad groups. The Friedmann equation, governing the evolution of $R(t)$, includes contributions from different types of matter (radiation, baryonic matter, non-baryonic matter, neutrinos, ...), spacetime curvature, and the cosmological constant. One set of parameters characterizes each of these contributions, in terms of density parameters Ω_i defined as ratios with respect to the "critical density."[5] In principle this leads to several independent

[3]This second assumption is more significant for two reasons. First, many of the observational relationships follow from the spacetime geometry of the FLRW models, and are in that sense independent of the dynamics for the gravitational field specified by general relativity. Second, many alternative theories of gravity also admit the FLRW models as solutions.

[4]Lemaître assumed a "perfect fluid" as the source – idealizing away the non-uniformities of a normal fluid that would be incompatible with these two symmetries. Robertson and Walker, in later work, clarified the geometrical properties of these models.

[5]The flat model has precisely the energy density needed to counteract initial expansion, so that $\dot{R}(t) \rightarrow 0$ as $t \rightarrow \infty$. From the Friedmann equation, this holds at the critical density: $\rho_c = \frac{3}{8\pi}\left(H^2 - \frac{\Lambda}{3}\right)$, where $H = \frac{\dot{R}}{R}$ is the Hubble constant. The density parameters for matter species are then conventionally defined with respect to the critical density, $\Omega_i = \frac{\rho_i}{\rho_c}$, with Λ set to zero; $\Omega_\Lambda = \frac{\Lambda}{3H^2}$; and $\Omega_k = -\frac{k}{R^2 H^2}$ (where $k = \{-1, 0, +1\}$ for negative,

	Planck TT+lowP+lensing	*Planck* TT+lowP+lensing+ext
$\Omega_{\rm b}h^2$	0.02226 ± 0.00023	0.02227 ± 0.00020
$\Omega_{\rm c}h^2$	0.1186 ± 0.0020	0.1184 ± 0.0012
$100\,\theta_{\rm MC}$	1.0410 ± 0.0005	1.0411 ± 0.0004
$n_{\rm s}$	0.968 ± 0.006	0.968 ± 0.004
τ	0.066 ± 0.016	0.067 ± 0.013
$\ln(10^{10}\Delta^2_{\mathcal{R}})$	3.062 ± 0.029	3.064 ± 0.024
h	0.678 ± 0.009	0.679 ± 0.006
σ_8	0.815 ± 0.009	0.815 ± 0.009
$\Omega_{\rm m}$	0.308 ± 0.012	0.306 ± 0.007
Ω_Λ	0.692 ± 0.012	0.694 ± 0.007

Table 1: Recommended cosmological parameter values from Tanabashi et al. (2018). The first column indicates values determined from low-multipole Planck measurements plus lensing. The second column adds several constraints from non-CMB observations. Uncertainties are at the 68% confidence level. The six parameters at the top are fit to the data; those at the bottom are derived. The analysis assumes spatial flatness.

parameters, each to be fixed by observations. In practice, cosmologists typically assume a flat FLRW model (setting $\Omega_k = 0$), use observations to find the optimal values for a subset (typically for baryonic Ω_b and cold dark matter Ω_c), and treat the rest as derived (or fixed by other constraints). The second group of parameters characterize the small perturbations away from an FLRW model. These perturbations can be fully characterized in terms of a dimensionless power spectrum, provided that they obey Gaussian statistics (as is usually assumed). Two free parameters specify the power spectrum: the spectral index (n_s), which is equal to unity if there is no preferred scale, and a parameter fixing the amplitude of perturbations (such as Δ_R). Finally, a third more heterogenous group of parameters includes properties that are essential to interpreting the observations but cannot be constrained from theory. This includes, for example, a parameter (τ) specifying the optical depth to the surface of last scattering (a measure of the ionization state of the universe), and the size of the acoustic horizon at the surface of last scattering (θ_{mc}).

Cosmologists often celebrate the advent of "precision cosmology," due to the dramatic increase in the amount, and diversity, of data that can be used to constrain these parameters. Ongoing observations of the cosmic microwave background radiation (CMBR) have been particularly important in giving precise determinations of the cosmological parameters, since it can be treated with high confidence using linearized perturbation theory and well-understood physics. Longair (2019) provides a historical overview of the various new techniques and observational campaigns devoted to precision measurements of the cosmological parameters pursued in the last three decades. As he describes, various lines of work have led to an impressive network of tests and consistency checks, leading to the optimal parameter values summarized in Table 1. The baryonic mass density Ω_b can be tightly constrained, for example, by the light element abundances based on big bang nucleosynthesis, as well as by the relative magnitude of the first and second acoustic peaks in the CMBR. The physics involved in spelling out the role of Ω_b in

flat, and positive curvature). Optimal values are often found for "physical density parameters" $\Omega_i h^2$, where h is the reduced Hubble constant ($= H_0/(100 km\ s^{-1}\ Mpc^{-1})$).

these two different phases of the universe's history are quite different, yet the two determinations are in very close agreement (with $\Omega_b \approx .045$). Similar arguments can be made for the other parameters. Peebles (2005), for example, reviews 13 different ways to measure the mean total mass density (including cold dark matter) on a wide variety of scales. The persistent contrast between these two estimates, with an overall mass density at $\Omega_m \approx .30$, was one of the factors leading to the acceptance of non-baryonic dark matter. (The energy density of cold dark matter is usually now treated as an independent parameter, Ω_c.)[6]

This last point underscores the flexibility highlighted in the very name of the ΛCDM model: allowing Ω_m to deviate substantially from Ω_b requires the introduction of non-baryonic dark matter, with an overall energy density roughly six times greater than that of baryonic matter. The cosmological constant, or "dark energy," contributes an even more remarkable amount of the total energy density: $\Omega_\Lambda \approx .70$. This should perhaps be acknowledged as an important third assumption of this framework: cosmologists do not limit their models to known types of matter and energy. Treating $\{\Omega_c, \Omega_\Lambda\}$ as free parameters amounts to taking out a mortgage against future physics. If this mortgage had been turned down, the same observations taken to vindicate the use of perturbed FLRW models would instead show that they cannot consistently incorporate all of the available observational data. The consistent, independent measurements of properties of the "dark sector" based on its cosmological impact suffices for a down payment, especially since currently understood physics does not demand setting either of these parameters to zero. But the bankers waiting for repayment, in the form of a compelling physical account of the dark sector, susceptible to non-cosmological tests, must be getting anxious.

Current parameter estimates rely almost entirely on CMBR observations. In Table 1, for example, the authors include the second column to show that adding other types of observations has minimal impact.[7] They take the parameter values determined from Planck as the "preferred" values, reflecting confidence in the physical understanding of the CMBR and observations of it. Yet if we take the ΛCDM to approximate the entire observable universe, rather than merely its state at the time the CMBR was emitted, various independent methods of measuring parameter values must coincide.

Accepting the ΛCDM model brings with it a commitment to account for how the parameters determined via "global" measurements of the CMBR relate to the "local" measurements of these same parameters.[8] As systematic uncertainties have narrowed for each of these different measurement techniques, this commitment is now being put to the test and will come under increasing pressure with the next generation of observations. As we will see below, "schematizing the observer" in Stein's sense plays a crucial role in understanding how these different measurements relate to one another, and in determining whether the simple six-parameter ΛCDM model suffices for describing the observed universe.

2 Status of FLRW Models

The advantages of adopting the FLRW models as a framework for interpreting cosmological observations should be clear, even from the very brief overview above. Yet it is also clear that the observed universe departs from the simple FLRW models, below extremely large length scales and at later stages of evolution (at the very least). How can we discern the FLRW models in the observed universe?

[6]The quoted value for Ω_b is based on a reduced hubble parameter $h = .70$, and the various determinations of the mass density also depend on the value of h; see Longair (2019) for further discussion and references.

[7]These include determinations of the Hubble parameter from the HST key project, constraints from supernovae measurements (JLA analysis), and observations of baryon acoustic oscillations (BAOs, from the SDSS, BOSS, and 6dF surveys).

[8]"Global" measurements include large-scale structure surveys and observations of the CMBR, and regard the universe at large length scales and/or early times (high redshifts). By contrast, "local" measurements, such as measurements of the Hubble parameter and the expansion of the universe based on Cepheids or supernovae, focus on smaller length scales and lower redshifts.

The interpretation of cosmological observations can hardly proceed without a variety of assumptions. What is the character of these assumptions, and how much do we need to assume *a priori* in order to interpret observations? This raises a concern of circularity. Do we need to in fact assume that the FLRW models hold, as in the previous section, or can we proceed with weaker assumptions? Furthermore, accepting the FLRW geometry forces the introduction of dark energy and dark matter, which together contribute the vast majority of the universe's energy density. Critics take this flexibility to indicate a flaw of the standard model (e.g., McGaugh, 2014; Merritt, 2017). In this section, I will consider three main lines of justification for employing the FLRW models distinct from the successful model building described above. According to the first, we should expect the universe to be almost FLRW as a result of early universe physics. I will argue that a second approach is more promising: if we take the Copernican principle (defined below) as constitutive of how to pursue physical cosmology, there is a more direct empirical case in favor of the FLRW models based on extrapolating from the isotropy of the CMBR. Finally, a third approach aims to test spatial homogeneity observationally with remarkably weak assumptions.

Before turning to these justifications, I should mention a minimalist approach that aims to use observations to determine spacetime geometry without restricting attention to a specific class of models. The "ideal observational cosmology" program pursued by George Ellis and various collaborators established that, in principle, the spacetime metric throughout the past lightcone of an observer can be determined based on ideal observations. These ideal observations include, for example, area distances, lensing distortion, and number counts for discrete sources. Yet this program has limited scope.[9] It proceeds as if there is a sharp separation between the characterization of the systems appearing in the ideal data set (such as stars that can be used as "standard candles") and cosmological assumptions. The direct observational approach cannot succeed unless the physical properties of the systems used to trace spacetime geometry are independently given. This is not the case for the "dark sector." Without substantive modeling assumptions, the distribution of dark matter, for example, cannot be determined from observations of luminous matter. Despite the value of this line of work, allowing for the possibility of dark matter and dark energy undercuts the possibility of directly justifying use of the FLRW models in this fashion.

The status of the FLRW models began to change in 1965, with the first observations of the CMBR. Subsequent observations placed increasingly tight bounds on the variations in temperature of the CMBR: the first evidence of anistropy was not detected until the COBE satellite (1992) measured temperature variations of 1 part in 10^5, once the dipole moment associated with the Earth's proper motion is subtracted.[10] From the 1930s onward, the case for the FLRW models relied in part on their ability to give a qualitative explanation of the Lemaître-Hubble recession of galaxies, and in part on the more controversial claim that the observed universe appears uniform above some large length scale. The statistical properties of large-scale distributions of galaxies and quasars provided the best case in favor of taking the FLRW models seriously, but they impose significantly weaker bounds on anisotropy than the CMBR. Before 1965, the FLRW models were typically regarded as, at best, a rough approximation to the large-scale properties of the universe above some length scale, ℓ — with ongoing debates about its value, currently estimated to be $\ell \approx 100\,Mpc$. The CMBR observations led cosmologists to conclude, by contrast, that the FLRW models provide an astonishingly accurate description of the entire observable universe when the CMB photons decoupled from matter. Although this

[9]See Ellis et al. (1985) for a full characterization of the ideal data set and a proof of the main theorem, and Ellis et al. (2012, Chapter 8) for a more recent discussion with further references, which emphasizes the limitations of the program if dark matter and dark energy exist. The other main challenge faced by the program is finding astrophysical systems that have appropriate features to be used in the ideal data set. Finally, the results constrain spacetime geometry on the past light cone of an observer, so making the case in favor of the FLRW geometry for other regions of spacetime would require some version of the Copernican principle (discussed below).

[10]Due to the Sachs-Wolfe effect, these temperature variations reflect variations in the gravitational potential on the surface of last scattering; subsequent observations of the polarization of CMB photons also constrain the peculiar velocity of the plasma.

discovery made it possible to apply the simple FLRW models with much greater confidence, cosmologists responded with perplexity: there was no obvious physical reason to expect that the early universe would be so highly symmetric.[11]

Many cosmologists now regard inflationary cosmology as providing a physical rationale for the FLRW models: we should expect an almost-FLRW bubble, of a scale much greater than the observed universe, as the "generic" output of a transient phase of exponential expansion in the very early universe.[12] The homogeneity and isotropy of spacetime within the bubble, along with the pertubations, all result from the dynamical evolution of a scalar field (or fields) driving this rapid expansion. There are several open foundational questions regarding inflationary cosmology (see, e.g., Brandenberger, 2014), but even setting these to one side, appealing to inflation (as it is currently understood) as a justification for using perturbed FLRW models begs the question. Suppose, to illustrate the point, that we had arrived at the current state of development of high energy particle physics prior to cosmology's golden age, and remained ignorant of the properties of the early universe. Would theorists be in a position to predict what we observe in the CMBR? Inflation arguably doesn't get us to a clear yes.[13] It leads to a reformulation of the question: rather than asking why the FLRW models apply, we ask why does an appropriate inflaton field (or fields) exist? This reformulation is fruitful. But currently candidates for the inflaton field are chosen for their compatibility with cosmological observations, rather than following from other considerations in high energy theory or phenomenology. As long as this is the case, even if inflation provides a plausible account of how to generate almost-FLRW bubbles, justifying the FLRW models based on inflation would be circular.

The question of whether inflation explains why the FLRW models apply brings to mind Stein (1989)'s trenchant remarks about Huygens's vortex theory. Huygens defended a theory of terrestrial weight, intended to account for gravity with only action-by-contact interactions. He proposed that a whirling vortex of aether would press on bodies, leading them to move just as Galileo had described. Ideology regarding what constitutes a good explanation drives theory development in both cases. In the case of inflation, cosmologists prefer "dynamical" explanations that do not rely on "finely-tuned" initial conditions. But, just as in politics, ideology often leads one to turn a blind eye to legitimate problems. The demand for an action-by-contact account of gravity requires incredible assumptions regarding the aether making up the vortex. Stein remarks that "a judgment not seduced by the great desire to *feel* that it understands ... will regard the explanation as quite arbitrary; as, so to speak, a quasi-pseudo-answer" (Stein, 1989, p. 55). Inflation, by contrast, does provide a promising way to understand observed features of the early universe, leading to many further open questions about the field(s) driving inflation. Yet some responses to these further questions, such as "eternal inflation," are quasi-pseudo-answers in just this sense.

In any case, cosmologists need not appeal to early universe physics. A more direct justification for using the FLRW models follows from the isotropy of the CMBR in conjunction with the so-called "Copernican principle." Isotropy observed from a single point, or along a single world-line, does not imply homogeneity.[14] Yet we are confined to observations from here and now, in cosmological terms. The Copernican principle, as it is usually formulated, asserts that our position is "typical" rather than "special." It is intended to justify the claim that other observers, in galaxies far, far away, see the universe more or less as we do. We can then appeal

[11]This response is particularly clear in Misner (1968).

[12]See, e.g., Martin (2018) for a concise recent overview of inflation, and Liddle and Lyth (2000) for a textbook account of inflation's implications with respect to large-scale structure.

[13]A prediction in this sense would not depend on observed features of the early universe. Currently, detailed knowledge of what has been established observationally regarding the early universe enters in at the design phase for inflationary models.

[14]For example, Ellis et al. (1978) construct a cylindrical spacetime which replaces variation in time in the FLRW models with spatial variation away from the axis of symmetry. In this model, the surface of last scattering is a timelike surface at a given radial distance, rather than a spacelike surface of constant time. The universe would appear to be isotropic to observers sufficiently close to the axis of symmetry in this model, but the model is obviously not spatially homogeneous.

to theorems clarifying how the observations of a group of "equivalent" observers spread through a region can constrain its spacetime geometry.[15]

These results are often called EGS-style theorems after the seminal result (Ehlers et al., 1968), which shows that if a collection of observers, moving along a congruence of timelike curves, all see an exactly isotropic radiation field, the spacetime geometry in the region filled by the congruence is conformal to a stationary solution. A similar result establishes that an isotropic radiation field (such as the CMBR), as seen by observers moving along timelike geodesics, implies FLRW geometry (Theorem 11.1 in Ellis et al., 2012). Ideally this line of thought would lead to a theorem that is robust — in the sense that, for example, "almost" isotropic radiation plus the Copernican principle implies that the universe is "almost" FLRW. Several further theorems have been established, but they still all require substantive physical assumptions that cannot be justified directly on observational grounds.[16] In addition, these theorems apply to "fundamental observers" – co-moving with the cosmic expansion along timelike geodesics. Comparing our observations to those of such idealized observers requires a model-dependent correction, subtracting our proper motion with respect to the CMBR and the associated dipole moment. Although there is still further work to be done, this is a promising approach to establishing almost FLRW geometry based on what a class of "equivalent" observers see.[17]

The status of the Copernican principle is tied up with fundamental questions about the aims of cosmology. Cosmology seeks to give an account of the world as a whole, at times long before we appeared on the scene, and at length scales enormously larger than those most directly relevant to experience. Treating cosmology as part of physics requires that we seek suitably objective explanations. The name of the principle invokes Copernicus's achievement in removing us from a priviliged position in the cosmos. Modern formulations of the Copernican principle typically rule out explanations that invoke facts about our location, on the grounds that they are not sufficiently objective. Despite the appeal of this line of thought, formulating the Copernican principle is surprisingly subtle. Facts about us *do* play a role in explaining some cosmological observations. Observers like us can only be present in extraordinarily rare physical environments. Selection effects related to the physical conditions necessary for our existence have to be taken into account in evaluating evidence.

I follow Roush (2003) in taking the appropriate formulation of the Copernican principle to involve, rather than a loss of privilege, a claim regarding the nature of our evidence: *we can plausibly take some types of observations as a representative sample from an ensemble of possible observations*. We can assess, case by case, proposals to take particular types of observations as unbiased. In some cases this is challenging due to "cosmic variance." Consider a property of the universe that we can only measure a limited number of times within the observable universe, such as low multipoles in the CMBR anisotropy spectrum. These observations are "cosmic variance limited" because we cannot use repeated observations to yield better statistics. The implications of an anomalous value compared to a theoretical ensemble are also unclear. We may have confidence in the physical description of stochastic processes generating an ensemble of possible realizations, for example of the CMB sky. But have we neglected physical effects whose only imprint is on extremely large-scale properties? Applying the Copernican principle in such cases requires facing up to difficult questions regarding how to distinguish laws and initial conditions in cosmology.

Several other cases are, however, more tractable. Obviously we are not an "unbiased sample" from an ensemble of observers sprinkled evenly throughout the galaxy with respect to spacetime volume. We must be located on a habitable planet, and this selects a specific "local" environment. By contrast, in the arguments in favor of the FLRW models the relevant features are

[15]The theorems make minimal initial assumptions about the spacetime, constrasting starkly with the parameter fitting approach described in the previous section.

[16]See Clarkson (2012) and Chapter 11 of Ellis et al. (2012) for reviews and references. One of the most general results, Theorem 11.3 in Ellis et al. (2012), requires, for example, that derivatives of the multipoles for the photon distribution are small, which does not follow from the multipoles themselves being small.

[17]Similar theorems make use of other kinds of observations in place of the CMBR, such as istropic expansion – see Clarkson (2012).

large-scale structure and the CMBR. Observers far away from us (in a similar local environment in another galaxy) would see a different pattern of temperature aniostropies in the CMB sky, and a different map of large-scale structure. But we have a clear understanding of the statistical properties of the ensemble of possibilities, with the possible exception of low multipoles; the scope of variations results from stochastic processes in the early universe. Our presence as observers plays no role in the physics describing how this ensemble is generated. We can then arguably take our observations as typical among a suitably defined ensemble, and appeal to EGS-style theorems to establish homogeneity.

To further illustrate the consequences of this formulation, consider the proposal that we are situated near the center of a (roughly) spherically symmetric "void," with a distribution of matter carefully constructed to yield accelerating expansion without dark energy (see Ellis et al., 2012, Chapter 15). Observers in galaxies far from the center of the void would see things differently. Intuitively this situation seems "improbable," but what does probability in this context mean? The main problem with this suggestion is not that the matter distribution is "unlikely"; rather, the proposal makes it impossible to pursue cosmology based on the evidence available to us. The proposal makes our evidence inherently parochial; it would have no bearing on a general account of the cosmos and its constituents. Rather than discovering the underlying physical causes of the large-scale structure of the universe, we would be limited to determining the properties of *our* Hubble volume. This does not mean that the proposal is incorrect; perhaps we really are located near the center of an under-dense bubble. But using our unusual position within the model to bear so much of the explanatory burden violates one conception of how to conduct cosmology.

To achieve its aim of developing a general theory of cosmic structure and its formation and evolution, cosmology must rule out this kind of inherently limited explanation – via an appeal to the Copernican principle. According to this line of argument, then, the justification for using the FLRW models follows from the claim that we should take our observations of large-scale features of the universe as "unbiased," reflecting a constitutive principle for the pursuit of physical cosmology. The almost-isotropy of the CMBR observed from along our world-line then justifies a generalization to almost-isotropy for other "equivalent" observers; an EGS-style theorem leads to the conclusion that spacetime is almost-FLRW (modulo the limitations noted above).

Remarkably, observations restricted to a point or worldline can provide tests of homogeneity.[18] This third approach proceeds with minimal assumptions, even dropping any claims regarding how our observations relate to those of other observers. The *failure* of homogeneity can be apparent even in observations conducted by a single observer, at least in principle.

One striking test exploits the fact that the CMBR has a black-body spectrum. Goodman (1995) suggested that this could be used to assess whether the CMBR appears to be isotropic from the vantage point of distant observers. Borrowing Goodman's analogy, imagine that a mirror, placed at an extremely large distance, reflects the CMBR spectrum to us. If the CMBR appears to be isotropic from the vantage point of the mirror, the CMBR spectrum we see reflected in it should match the spectrum we observe directly. Inverse Compton scattering of the CMB photons by high-energy electrons (the Sunyaev-Zel'dovich effect), found in hot, ionized intra-cluster gas in a distant galaxy cluster, takes the place of the mirror. This scattering of CMB photons leads to characteristic, measurable distortions in the CMBR spectrum, if the CMBR is anisotropic to the scatterer (thermal SZ) or if the cluster gas has bulk radial motion (kinetic SZ). The absence of such distortions places upper bounds on anisotropy of the CMBR as viewed from distant regions where they intersect our past light cone, or on the radial motion of cluster gas, leading to (for example) constraints on proposed "void" models (Caldwell and Stebbins, 2008). (Scattering also has an effect on the polarization of CMB photons, which provides a further means to test homogeneity.) Despite the appeal of these direct bounds on inhomogeneity, it is challenging to segregate these effects from other causes of spectral distortions in CMB photons.

The most feasible direct test of homogeneity exploits a consistency relation that follows

[18] See Clarkson (2012) for a more comprehensive discussion of these tests.

from FLRW geometry, without further dynamical assumptions or posits regarding material constituents. Starting with the expression for luminosity distance in FLRW models, it is straightforward to derive an equation for the spatial curvature Ω_k in terms of the Hubble parameter $H(z)$ and a distance measure $D(z)$ (both functions of redshift). Clarkson et al. (2008) noted that the constancy of Ω_k implies that its derivative with respect to z vanishes; hence, a consistency relation involving $H(z), D(z)$ and their first and second derivatives with respect to z must also vanish. Typically measurements of $H(z)$ rely on FLRW geometry. But it is possible to measure both $H(z), D(z)$ independently without that assumption, and hence to check the consistency relation directly. (See Clarkson, 2012, for further discussion and references.) The observations needed to carry out this program are essentially the same as those needed to constrain the equation of state of dark energy ($w(z)$). Observational programs such as Euclid that aim to measure $w(z)$ will also yield precision tests of spatial homogeneity.

In sum, cosmologists have made a case that we must use the FLRW models, not only because of their appealing simplicity, but because observations compel us to do so. I have criticized the appeal to inflation as circular, at least at this point in the development of physics. The second and third approaches reveal that cosmologists have been able to probe the FLRW models in a variety of different ways, employing different (and in several cases surprisingly weak) theoretical assumptions. I characterized the Copernican principle as constitutive of a particular conception of cosmology: we aim to discover the physics governing the evolution of the unvierse, the formation of structure, and so on, rather than contingent features specific to our neighborhood. Accepting this principle allows us to generalize some aspects of our observations, in particular taking the almost-isotropy of the CMBR to hold for a class of equivalent observers. The third approach complements this line of argument. We have no guarantee that nature will cooperate in allowing us to pursue cosmology under this conception, so it is particularly valuable to have direct empirical tests that would reveal departures from spatial homogeneity. The discussion has been limited so far to assessing how the FLRW models count as knowledge in the sense of systematic claims subject to empirical justification; next, we turn to a further aspect of knowledge Stein emphasizes, namely their role in guiding further inquiry.

3 Fitting Problems

Physicists often handle the complexity of real phenomena by starting with a simple, physically well-motivated model, and using discrepancies between the simple model and observations to guide the development of a more detailed — and, hopefully, accurate — account. Cosmologists take the FLRW models to be an initial, simplified model of the universe, and the starting point of inquiry, in just this sense. The actual universe departs from the uniformity of the FLRW models, with enormous density contrasts on small scales that can hardly be treated as "small perturbations". Does taking the FLRW models as a suitable first step in the development of more complex models lead to a progressive research program, yielding further insights into the universe such as the discovery of dark matter and dark energy? Or are the critics right to worry that the need to introduce both of these entities, and other difficulties with "fitting" an FLRW model to the observed universe, reflect a degenerating research program?

Stein's emphasis on the enterprise of knowledge — how theories guide ongoing inquiry — bears at least a family resemblance to Lakatos's account of methodology (Lakatos, 1980). Lakatos takes the historical progression of a series of theories, a research program, as the appropriate unit of evaluation, and aims to provide tools to determine whether a particular program qualifies as progressive or degenerating.[19] A research program consists of a group of theories that share a "hard core" of foundational principles, along with a positive heuristic guiding future work. In a nutshell, on Lakatos's account a progressive research program consistently makes successful novel predictions, or as he puts it "the well-planned-building of pigeon holes must proceed faster than the recording of facts to be housed in them" (Lakatos, 1980, p. 100). A degenerative program, by contrast, features more twists and turns to reconcile the theory

[19]My thanks to an anonymous reviewer for the question that prompted these two paragraphs.

with new empirical discoveries than predictions. *Pace* Popper, research programs should not be abandoned if they go through some twists and turns. On Lakatos's view, it is rational to modify auxiliary hypotheses to preserve core principles, and with the benefit of hindsight we can distinguish between momentary divergences from the true path and dead ends.

Despite several appealing insights, Lakatos's account falls short in crucial respects.[20] Most notably for present purposes, in spite of the name of his own position, Lakatos does not provide a "methodology" in the traditional sense of prospective guidance. Scientists will look in vain for analytical tools that might help to decide whether their favorite research program has reached a phase of degeneration; Lakatos only offers retrospective assessments after the dust has settled. Furthermore, his account of progress is both ill-defined and unduly narrow. Lakatos failed to give a satisfactory account of novelty that supports defining progressivness in terms of the number of novel predictions. He strikingly neglects entirely a type of progress that is one of Stein's central concerns, namely the refinement of fundamental principles (or *Tieferlegung der Fundamente*, as Stein quoting Weyl quoting Hilbert puts it). That said, Lakatos was certainly correct to focus evaluation on the trajectory of a research program. The tools he developed to sift cases of the growth of knowledge from the chaff, retrospectively, would need to be refashioned to be of use for prospective assessment of contemporary science.

Rather than attempting to refine Lakatos, I find it more useful to consider such questions in parallel with a detailed account of progress in a field facing similar evidential challenges: celestial mechanics. Smith (2014) emphasizes that two features have been essential in developing models of the solar system, through a series of steps that remove idealizations and simplifying assumptions.[21] First, the inferences drawn from observations at a given stage of inquiry should be robust, in the sense that they remain valid, at least to a first approximation, as inquiry proceeds. For example, parameters appearing in an initial model (such as the masses of the planets) should retain a clear physical meaning. Measurements of these parameters may be refined and corrected through later work, but not rejected outright – as failing to target real physical quantities. Second, the model constructed at each stage should hold exactly in clearly specifiable circumstances. The model should be an exact solution of relevant physical equations, with specific complicating factors excluded. (For example, models at early stages of the development of celestial mechanics treated the planets as point masses interacting gravitationally, neglecting effects due to their finite extent that became relevant later.) It may then be possible to assess whether a given discrepancy can plausibly be ascribed to one of the factors that has been left out, allowing comparisons with observations to guide the development of more sophisticated models.

If these two features hold, at any given stage of inquiry the model is an exact description of a system that excludes some relevant features, but it nonetheless supports inferences from observations, including estimates of the physical quantities characterizing the system. In place of Lakatos's positive and negative heuristics, this approach emphasizes the central importance of measurement alongside a structured path of model refinement through adding details. Smith (2014) argues that the central question in celestial mechanics following Newton was:

> ... not whether calculated locations of planets and their satellites agree with observations, but whether robust physical sources can be found for each systematic discrepancy between those calculations and observation – with the further demand of achieving closer and closer agreement with observation in a series of successive approximations in which more and more details of our solar system that make a difference become identified, along with the differences they make. (p. 279)

On this account, what we might call a progressive research program succeeds by identifying further physical details in the course of refining theoretical models.

[20] See, e.g., Laudan (1977); Hacking (1979) for influential critisisms, and Chall (2020) for a neo-Lakatosian revival – an analysis of research programs in contemporary particle physics.

[21] This characterization over-simplifies Smith's account considerably; see also Smith (2001, 2002) regarding Newton's introduction of this methodology in the *Principia*.

There are clear challenges to establishing that the two features highlighted in Smith's retrospective assessment of celestial mechanics hold in the case of cosmology. Both challenges can be regarded as "fitting problems," in that they stem from ambiguity in how more realistic models relate to the "best-fit" FLRW model. Most cosmologists aim to approach the complexity of real phenomena by building a more complete description based on these models. It will take many decades, at least, for cosmologists to realize fully the potential of this approach, to determine whether a path like that taken in celestial mechanics will lead to similar success. Paraphrasing Smith, the question is not whether cosmological observations can be fit with a particular choice of parameters in the ΛCDM model, but whether robust physical sources can be found for each systematic discrepancy between the best-fit model and observations – with the further demand that incorporating these sources leads to increasingly accurate models.

The "fitting problem" usually refers to the following issue: how can we "smooth" a more realistic model, with large density contrasts and small scale inhomogeneities, to obtain a best fit FLRW model? How does the dynamical description at extremely large length scales and early times, where the FLRW models apply for the reasons given above, relate to much shorter scales? As Ellis (1984); Ellis and Stoeger (1987) emphasized, in a non-linear theory like general relativity the relationship between physical descriptions at different scales is not straightforward. Ellis considered an "averaging operator" A that, given a metric g_{ab}, "smooths" or "coarse grains" inhomogeneties below a chosen length scale L to yield a new metric, schematically: $A : g_{ab} \to g^L_{ab}$. (Successive applications of the operator can relate different length scales.) He showed, however, that applying A to the metric and also to the stress-energy tensor for an exact solution does not yield a new solution of EFE. The "smoothed" Einstein tensor G^L_{ab} generally does not satisfy the EFE with the "smoothed" stress-energy tensor T^L_{ab}. Averaging and solving EFE do not commute.[22] If we grant that GR captures physics at the small scale, then EFE hold for an initial metric g_{ab} that accurately represents a complex, inhomogeneous matter distribution: $G_{ab} - 8\pi G T_{ab} = 0$. By contrast, an "averaged" model, at some larger length scale L, includes a correction: $G^L_{ab} - 8\pi G T^L_{ab} = \epsilon^L_{ab}$. (We are assuming that GR still applies at this larger length scale.) The ϵ^L_{ab} term reflects the back-reaction of small-scale inhomogeneities on the dynamics at a larger scale. If we take the FLRW models as applying to large-scale averages of the Einstein and stress-energy tensors, then the familiar FLRW dynamics are not the correct equations — we need to include ϵ^L_{ab}. The first challenge is to determine the nature of this correction, and to assess its impact on the inferences cosmologists draw based on the FLRW models.

The second challenge regards the relationship between observable quantities in a "best-fit" FLRW model and a more realistic model. Grant, for the sake of argument, that the realistic metric g_{ab} can be described as a "small" perturbation, γ_{ab} to a background FLRW metric g^B_{ab}: that is, $g_{ab} = g^B_{ab} + \gamma_{ab}$. Approximate agreement between the two metrics does not imply that various observable quantities will be close, however, because the derivatives of γ_{ab} may still be quite large. Green and Wald (2014) illustrate this point with a useful analogy: consider the metric describing an exact sphere, h^S_{ab}, along with the metric h_{ab} of a convex polyhedron with N sides, with each side tangent to a point on the sphere. (We can further round off the edges of the polyhedra so that the curvature is large at the vertices, but does not diverge.) For large N, the metric of the polyhedron will be very close to that of the sphere, $h_{ab} = h^S_{ab} + \delta_{ab}$ with a small δ_{ab}. Yet the curvature of the two metrics differs substantially: h_{ab} is almost entirely flat, with large curvature near the vertices, by contrast with the constant curvature of h^S_{ab}. Curvature depends on the second derivative of the metric. The second derivative of δ_{ab} can be quite large even if δ_{ab} itself is relatively small. The geodesics of the two metrics can also differ substantially, for the same reason.

Going back to cosmology, the realistic metric g_{ab}, even if it approximates an FLRW metric at some scale, will likewise have regions of extremely high curvature at small scales (for example,

[22]There is a further delicate question of what it means to "average" tensorial quantities, in order to define A precisely. Ellis's argument is quite general, however, and anything that would qualify as a reasonable "averaging procedure" fails to commute with solving the EFE. The argument proceeds by considering the effect of averaging on quantities that appear in the Einstein tensor: the average of the derivative of a function is not generally equal to the derivative of the average, and further non-linearities arise in calculating the Ricci tensor and Einstein tensor.

near black holes). This raises a general question regarding the physical meaning of the cosmological parameters appearing in the best-fit FLRW models. There is no guarantee that observables calculated according to the background FLRW metric g^B_{ab} will agree with those calculated in the almost FLRW metric g_{ab}, since these observables often depend on higher derivatives of the metric. This is true even granting that the metric is almost FLRW, the best case scenario. Furthermore, there is no guarantee that there is a unique FLRW model that provides a best fit for *all* cosmologically relevant observables. How, then, do the quantities we actually have access to observationally relate to the basic parameters listed in Table 1?

These challenges are both particularly significant in an era of precision cosmology. Obviously to be useful in testing theories, increased precision has to be accompanied by clarity regarding what is being measured, which I will turn to in §4. Regarding the first challenge, the correct dynamical equations for large-scale dynamics differ from standard FLRW dynamics due to the back-reaction term. Depending on the magnitude and nature of this term, inferences such as that from accelerated expansion to the existence of dark energy may fail. I will turn to debates about how to estimate the backreaction in §5.

4 Observations in an Inhomogeneous Universe

Observational cosmology relies almost entirely on the analysis of light and other forms of electromagnetic radiation from distant sources. The implications of observations depends on how the radiation propagates through spacetime, and interacts with different types of material constituents. The observed angular diameter and luminosity of a distant source vary as a function of redshift, for example, based on the spacetime geometry. Several relationships of this type were established for the FLRW models in the early days of relativistic cosmology, and are routinely used in analyzing cosmological observations, in particular to determine cosmological distances. In the early 60s, Zel'dovich and Feynmann independently argued that these results do not directly carry over to more realistic models.[23] In an FLRW model, light passes through matter with uniform density, whereas in a lumpy, inhomogeneous model it passes mostly through empty space. Zel'dovich emphasized the importance of quantifying the impact of the fact that light travels mostly through a vacuum, to clarify the limitations of the FLRW observational relations and distance estimates based on them.

The response to this question illustrates the importance of "schematizing the observer," in Stein's sense, as part of explaining how we discern a theory in the phenomena.[24] There are two closely related reasons why a detailed account of how a specific observation proceeds is necessary to understand its implications. First, in any measurement we need to have sufficient understanding of the measurement process itself to identify, and correct for, any distortions it introduces. It is striking how many different aspects of theoretical knowledge are drawn into a schematic account of an observation. Processing data in precision astrometry, for example, requires accounting for, among many other things, the position and motion of the telescope (to make corrections for stellar aberration) and subtle effects that depend upon the Earth's position in its orbit (due to the effect of the spacetime curvature near the Sun on light propagation).[25] Second, we need to assess, for each stage of the measurement process, whether the relevant theories can establish the reliability of inferences at that stage. Does the measurement interaction fall within the domain of applicability of the theories used to describe it? Zel'dovich's challenge regards whether our actual observations fall within the regime where we can safely

[23]Several papers noted this problem and initiated the study of light propagation in inhomogeneous models, see in particular Zel'Dovich (1964); Bertotti (1966); Gunn (1967); Gunn cites Feynman's contribution, based on an unpublished Caltech colloquium in 1964.

[24]Curiel (2019) vigorously argues that schematizing the observer must play a role in any account of the semantics and structure of scientific theories. Curiel's work has influenced my thinking on this topic, but I will not pursue a detailed comparison with his views here.

[25]See Barmby (2019) for an introductory overview of observations in astronomy, emphasizing the wide range of knowledge employed in instrument design and data processing.

use relationships based on FLRW geometry. This is by necessity a fine-grained assessment: it may be the case, as Zel'dovich notes, that for some types of cosmological observations we can reliably use the FLRW relations, while for others we cannot.

To assess the reliability of different measurements, we need to understand how the presence of inhomogenieties and complex structures effects the optical properties of a cosmological model. The Sachs optical equation describes the propagation of a light beam through a curved spacetime.[26] Roughly put, two terms in the equation determine how the cross section of the light beam changes. (Recall that the Riemann tensor characterizing spacetime curvature decomposes into two components: the Ricci tensor and the Weyl curvature tensor, which represents gravitational degrees of freedom.) One term in the optical equation characterizes the response of the beam to the Ricci tensor: in regions with a positive Ricci tensor, the beam is focused (the null geodesics converge). According to the other term, in regions of spacetime with a non-zero Weyl tensor, the beam undergoes a volume-preserving shear. FLRW models are conformally flat, which implies that the Weyl tensor vanishes. The evolution of a light beam in an FLRW model is then determined entirely by the first term of the equation: the uniform matter density leads to continual focusing of the beam. By contrast, in a more realistic model of the local universe, consisting of localized concentrations of matter interspersed with much larger voids, a light beam will mainly pass through regions where the Ricci tensor is zero but the Weyl tensor is not. (The Ricci tensor is non-zero only in localized high-density regions) Qualitatively there is a striking contrast between the propagation of light in these two cases. But detailed quantitative modeling is required to determine to what extent the distinctive optical properties of a more realistic cosmological model have an impact on the measurement of cosmological parameters.

The impact of the optical properties on parameter determination also depends on the angular size of the light beam relevant to the observation. To take the most extreme contrast, CMBR observations use a large-angle light beam (spanning nearly the entire sky for low multipoles), whereas observations of supernovae use an extremely narrow beam. Other cosmological observations fall in between these extremes (Fleury, 2015, pp. 102-103). This raises the possibility that different measurements in effect probe different spacetime geometries: a large-angle measurement may simply be insensitive to local inhomogeneities, even if they do have an appreciable impact on narrow-beam measurements.

Discrepancies in measurements of the Hubble parameter have spurred more careful assessment of observations of different types in more realistic models. At present, the conflict between local determinations of H_0 and the value inferred from the CMBR and other global measurements is the most prominent problem facing the ΛCDM model. The Planck Collaboration reports a preferred value of $h = .674 \pm .005$, compared with $h = .7403 \pm .0142$ favored by local measurements. These two values differ by .066, nearly 10%, a discrepancy of $4.4\,\sigma$ given stated systematic uncertainties. "Local" measurements leading to the higher value determine the Hubble parameter based on the Hubble diagram for SN Ia. (There is ongoing debate regarding the distance ladder based on SN Ia, due to conflicting results based on different zero-point callibrations.)[27] The CMBR, by contrast, does not provide a direct measurement of the Hubble parameter, but determines its value based on an optimal parameter fit. (The acoustic scale, fixed by the location of the acoustic peaks in the CMBR power spectrum, constrains the product of the Hubble parameter and the matter density.) The discrepancy between these two measurements of the Hubble parameter has persisted even as the systematic uncertainties of the various methods have decreased. Cosmologists have explored a wide variety of possible systematic uncertainties that could account for the discrepancy, as well as considering the more radical step of modifying

[26] Here we are assuming that the geometric optics approximation holds – namely, that the wavelength is much smaller than the curvature scale, and that the amplitude is "small" in a sense which can be made precise. In this case, the gravitational field generated by the radiation itself can be neglected, and the radiation follows null geodesics. See, for example, Chapter 7 of Ellis et al. (2012) or Fleury (2015) for treatments of optics in curved spacetimes.

[27] These figures are taken from the synopsis of the current observational situation in Freedman et al. (2019). The figure cited in the text is based on callibrating the Hubble diagram for SN Ia using cepheids, which yields the largest discrepancy; the main focus of the paper is the "tip of the red giant branch" method, an alternative zero-point callibration which leads to an intermediate value of $h = .698 \pm .008$. See fn. 5 for the definition of h.

the physics of the ΛCDM model (e.g., Bernal et al., 2016). This includes further investigation of the contrasting optical effects of inhomogeneities on local vs. global measurements.

Fleury et al. (2013), for example, quantify the impact of inhomogeneities by considering light propagation through so-called "Swiss cheese" models. These models are constructed by excising regions from a background FLRW model, and "pasting in" a Schwarzschild solution with the same mass as the excised region concentrated at a central point. Fleury et al. (2013) introduce a "smoothness" parameter f characterizing how much of the mass is concentrated in the Schwarzschild "holes" vs. the FLRW "cheese", such that the FLRW model is recovered as $f \to 1$. They then calculate the impact of varying f on a local Hubble diagram, based on modeling the propagation of narrow beams of light through the inhomogeneous model. Departures from FLRW lead to changes in the determinations of cosmological parameters (including H_0, as well as Ω_m). These can be significant: Ω_m varies from $.25 - .30$ for different values of f. However, the effect on determinations of the Hubble parameter itself is still significantly smaller than the current discrepancy. Despite the striking physical contrasts between light propagation in FLRW and more realistic models, the standard FLRW results seem to be remarkably good approximations. This may be due to the large value of Λ (Fleury, 2015). This particular modeling approach leads to the conclusion that the effects on parameter estimates decrease with increasing Λ. The Λ term is homogeneous (it is constant, as the name suggests), so a large Λ term suppresses the effect of inhomogeneities.

Results along these lines provide a partial response to the second challenge identified above: how a more realistic treatment of measurement relates to the simple account provided by an FLRW model. Considered with respect to this specific class of inhomogeneous models, Fleury et al. (2013) show that the FLRW models are surprisingly reliable, despite their unrealistic optical properties. The errors introduced by treating light as if it propagated through a FLRW geometry are negligible, given the current level of precision.

A comprehensive response, however, requires considering the impact of inhomogeneities in a broader range of spacetimes. Ideally, we would be able to integrate the optical scalar equations along photon trajectories representing observers' past light cones for arbitrary spacetimes. This would allow a direct assessment of how much observations, conducted in different regions, depart from the FLRW results. Recent work in numerical relativity has made progress towards this goal. For example, Adamek et al. (2019) assess the bias and scatter in the Hubble diagram (the redshift-distance correlation) produced by inhomogeneities. Rather than starting from exact solutions, they consider spacetimes generated by a relativistic N-body structure formation simulation called *gevolution*.[28] Roughly put, this code generates a spacetime geometry and a catalog of dark matter haloes used as proxies for astrophysical objects; they then construct an ensemble of Hubble diagrams for different observers by integrating the optical equations along photon trajectories. The main advantage Adamek et al. (2019) claim for their approach, by contrast with earlier methods, is that it does not rely on approximations with unclear domains of applicability, or exact solutions. Several other groups have also used numerical relativity to assess the reliability of cosmological measurements (e.g., Giblin Jr et al., 2016; Macpherson et al., 2018). Detailed evaluation of these different approaches turns on subtle technical questions, well beyond the scope of this paper. But they share a common aim: to assess how the shift to realistic spacetime models impacts measurements of the Hubble parameter.

Similar questions arise for the other parameters appearing in the ΛCDM model: to what extent can we defend the claim, in detail, that the different ways of measuring Ω_m all target a single quantity? And what do the different measurement modalities assume regarding the domain of applicability of the FLRW models, or other cosmological assumptions? Precision science requires developing detailed accounts of how an observation proceeds in order to identify and control possible sources of error, and to assess the reliability of inferences. While this will come as no surprise to scientists, philosophical accounts of the structure of scientific theories often fail to recognize the importance of this activity, and the necessity of "schematizing the

[28] I am grateful to an anonymous referee for bringing this line of work to my attention. The cited paper is one in a series of papers based on the *gevolution* code, exploring different aspects of the reliability of precision cosmology.

observer" in Stein's sense.

5 Backreaction and Dark Energy

In 1998, two observational teams announced the discovery that the expansion rate of the universe is accelerating, based on using Type Ia supernovae as "standardizable" candles to extend the Hubble diagram to larger cosmological distances with increased precision. Granting that FLRW dynamics applies, these observations imply that there must be a large contribution to the overall mass-energy density which has the same dynamical effect as a non-zero Λ, sometimes attributed to "dark energy".[29] Ostriker and Steinhardt (1995) had made a case for a substantial contribution from Λ earlier, based on a re-analysis of a variety of cosmological data, but the supernovae observations (awarded a Nobel prize in 2011) provide a much more direct empirical case.

Yet as we have seen in §3, this straightforward inference uses the wrong dynamical equations. We need to include a correction term ϵ^L_{ab} representing the back-reaction of smaller-scale inhomogeneties on dynamics at the larger length scales relevant to these measurements. Although cosmologists agree on the need to account for the effect of averaging over the inhomogeneities, there is an ongoing debate regarding how to calculate this contribution. Claims[30] that back-reaction may account entirely for the observed accelerated expansion without introducing "dark energy" or a non-zero Λ have provoked considerable controversy (see, e.g., for reviews Clarkson et al., 2011; Buchert and Rsnen, 2012; Buchert et al., 2015; Green and Wald, 2015). Here I will briefly describe what is at issue in this debate, with the aim of illustrating what needs to be done to establish the reliability of the inference to the existence of dark energy.

The character of the backreaction term reflects the relationship between "long-wavelength" and "short-wavelength" behavior in solutions of EFE. Consider, for example, injecting energy into the short-wavelength degrees of freedom, treated as perturbations to a given exact solution of the equations. Turbulent systems in fluid mechanics can exhibit an "inverse energy cascade," with energy transfer from small scales to larger scales. Does something similar occur according to EFE, or is there a dynamical decoupling that ensures that the long-wavelength properties of the solution are stable to such perturbations (Green and Wald, 2016)? Clearly there is not an inverse cascade if we restrict attention to linearized perturbation theory, according to which Fourier modes evolve independently. But due to the non-linearity of EFE it is possible for Fourier modes at different wavelengths to couple, leading to the possibility of energy transfer from small to larger scales. To quantify this effect and its implications for backreaction, one needs to determine the properties of ϵ^L_{ab} and how those relate to the background solution and the nature of the short wavelength perturbations.

Given this understanding of backreation, physicists have employed different methods to determine what energy conditions ϵ^L_{ab} satisfies, and how the magnitude of ϵ^L_{ab} compares to other scales appearing in the theory (Baumann et al., 2012; Green and Wald, 2011, 2016). These calculations focus on FLRW models with ordinary matter and energy as the sources, and consider an "effective" stress tensor that arises at large length scales due to back-reaction. Green and Wald (2011) introduce a new formalism to calculate ϵ^L_{ab}, and find that it has several

[29] This is apparent from the (isotropic) Raychaudhuri eqn.:

$$3\frac{\ddot{R}}{R} = -4\pi G\left(\rho + 3p\right) + \Lambda, \tag{1}$$

where R is the scale factor, $\dot{R}$ indicates the first derivative with respect to cosmic time, ρ, p are the density and pressure of a perfect fluid, and Λ is the cosmological constant. A non-zero Λ is needed for $\ddot{R} > 0$ because it is the only term with the correct sign. Although at present observations do not rule out a true cosmological constant as the source driving accelerated expansion, several observational programs aim to measure deviations – with the hope of constraining the properties of dark energy.

[30] Several authors have made claims along these lines, albeit differing considerably in detail – see, for example, Rsnen (2004); Wetterich (2003); Kolb et al. (2006).

properties: it is trace-free and satisfies the weak energy condition. They conclude that the back-reaction effects are quite small, well below the 1% level, in essence correcting the matter stress-energy to include kinetic energy and Newtonian potential energy. Baumann et al. (2012) reach similar conclusions based on a different formalism, although they hold that there may be a detectable impact on measurements of large-scale structure.

Several physicists have disputed whether these results resolve the questions at issue in cosmology. The disagreement runs deep, even to the definition of backreaction. By contrast with the analysis above of backreaction as a physical energy transfer from small to large scales, Buchert et al. (2015) define backreaction in more general terms as "deviation of spatial average properties of an inhomogeneous universe model from the values predicted by a homogeneous-isotropic universe model" (§3.1). A prescription for "averaging" geometrical quantities in inhomogeneous models has to be specified in order to even make this comparison. Buchert (2000, 2001) develops an "averaging" (coarse-graining) technique, in order to construct a coarse-grained cosmological model "bottom-up" from observations without presuming the applicability of a perturbed FLRW model. This leads to a new set of dynamical equations replacing those of the FLRW models. These equations apply to a chosen region of space, and they relate the expansion rate to the average values of the energy density, scalar spatial curvature, and other kinematic quantities in this region. This contrasts sharply with the approach above, because the dynamics do not simply lead to adding an effective stress-energy tensor to the FLRW dynamics as a correction. Calculations in this approach suggest, contrary to the perturbative assessments, that backreaction can have a large impact, including possibly accounting for accelerated expansion.

The debate regarding the consistency and applicability of different attempts to quantify backreaction continues, and I have no pretense of resolving the subtle technical questions involved. Rather the point is to illustrate an important open question in the foundations of cosmology. The inference from observations to dark energy assumes that we have sufficient quantitative control over back-reaction, so that we can place bounds on its possible contribution to the observed acceleration. Granted a proof of stability in this sense, then it is possible to proceed as in many other areas of physics: by starting first with an idealized model, and then adding further complexities in order to approach a more realistic description (as endorsed by Green and Wald, 2016). Some of the arguments in favor of a "non-perturbative" approach conflate two issues: whether this methodology is a fruitful way to proceed, and whether we have sufficient understanding of EFE to use the FLRW models in reliable inferences. Regarding the first issue, Buchert regards "averaging" as a more well-founded empirical procedure for finding a cosmological model that adequately describes the late universe. Yet other historical cases suggest that the path to an adequate model of complex phenomena goes through a series of idealized models.

6 Conclusions

The current ΛCDM model is an impressive achievement, and it is hard to imagine that the future evolution of the field will lead to an entirely different account. Yet despite this impressive descriptive success, there are foundational questions about the status of this model. The central question now facing the ΛCDM model is *not* whether calculated features of large scale structure, expansion history, and other cosmological observables can be fit with a particular choice of parameters. Rather the question is whether the model has a clear physical status, such that any systematic discrepancy between the model and observations can be reliably taken to indicate the need to include a new physical feature. The model has in fact been taken to support introducing new entities into physics, in particular dark matter and dark energy, and we have clarified above the kind of result that would help to secure the latter inference. Considering cosmology based on Stein's understanding of the structure of theories, with a particular emphasis on schematizing the observer and what needs to be in place for the FLRW models to guide further inquiry, has helped to identify and elucidate these foundational questions.

Acknowledgments: (Removed for blind review)

References

Adamek, J., Clarkson, C., Coates, L., Durrer, R., and Kunz, M. (2019). Bias and scatter in the Hubble diagram from cosmological large-scale structure. *Physical Review D*, 100(2):021301.

Aghanim, N., Akrami, Y., Ashdown, M., Aumont, J., Baccigalupi, C., Ballardini, M., Banday, A., Barreiro, R., Bartolo, N., Basak, S., et al. (2018). Planck 2018 results. VI. cosmological parameters. *arXiv preprint arXiv:1807.06209*.

Barmby, P. (2019). Astronomical observations: a guide for allied researchers. *The Open Journal of Astrophysics*, 2(1).

Baumann, D., Nicolis, A., Senatore, L., and Zaldarriaga, M. (2012). Cosmological non-linearities as an effective fluid. *Journal of Cosmology and Astroparticle Physics*, 2012(07):051.

Bernal, J. L., Verde, L., and Riess, A. G. (2016). The trouble with H_0. *Journal of Cosmology and Astroparticle Physics*, 2016(10):019.

Bertotti, B. (1966). The luminosity of distant galaxies. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 294(1437):195–207.

Bondi, H. (1960). *Cosmology*. Cambridge University Press, Cambridge.

Brandenberger, R. (2014). Do we have a theory of early universe cosmology? *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 46:109–121.

Buchert, T. (2000). On average properties of inhomogeneous fluids in general relativity: dust cosmologies. *General Relativity and Gravitation*, 32(1):105–125.

Buchert, T. (2001). On average properties of inhomogeneous fluids in general relativity: perfect fluid cosmologies. *General Relativity and Gravitation*, 33(8):1381–1405.

Buchert, T., Carfora, M., Ellis, G. F., Kolb, E. W., MacCallum, M. A., Ostrowski, J. J., Rsnen, S., Roukema, B. F., Andersson, L., Coley, A. A., et al. (2015). Is there proof that backreaction of inhomogeneities is irrelevant in cosmology? *Classical and quantum gravity*, 32(21):215021.

Buchert, T. and Rsnen, S. (2012). Backreaction in late-time cosmology. *Annual Review of Nuclear and Particle Science*, 62(1):57–79.

Caldwell, R. and Stebbins, A. (2008). A test of the Copernican principle. *Physical review letters*, 100(19):191302.

Chall, C. (2020). Model-groups as scientific research programmes. *European Journal for Philosophy of Science*, 10(1):1–24.

Clarkson, C. (2012). Establishing homogeneity of the universe in the shadow of dark energy. *Comptes Rendus Physique*, 13(6-7):682–718.

Clarkson, C., Bassett, B., and Lu, T. H.-C. (2008). A general test of the Copernican principle. *Physical Review Letters*, 101(1):011301.

Clarkson, C., Ellis, G., Larena, J., and Umeh, O. (2011). Does the growth of structure affect our dynamical models of the universe? the averaging, backreaction, and fitting problems in cosmology. *Reports on Progress in Physics*, 74(11):112901.

Curiel, E. (2019). Schematizing the observer and the epistemic content of theories.

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

Ellis, G. F. (1984). Relativistic cosmology: its nature, aims and problems. In *General relativity and gravitation*, pages 215–288. Springer.

Ellis, G. F., Maartens, R., and MacCallum, M. A. (2012). *Relativistic cosmology.* Cambridge University Press, Cambridge.

Ellis, G. F. and Stoeger, W. (1987). The 'fitting problem' in cosmology. *Classical and Quantum Gravity*, 4(6):1697.

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.

Fleury, P. (2015). Light propagation in inhomogeneous and anisotropic cosmologies. *arXiv preprint arXiv:1511.03702*.

Fleury, P., Dupuy, H., and Uzan, J.-P. (2013). Interpretation of the Hubble diagram in a nonhomogeneous universe. *Physical Review D*, 87(12):123526.

Freedman, W. L., Madore, B. F., Hatt, D., Hoyt, T. J., Jang, I. S., Beaton, R. L., Burns, C. R., Lee, M. G., Monson, A. J., Neeley, J. R., et al. (2019). The Carnegie-Chicago Hubble Program. VIII. an independent determination of the Hubble constant based on the tip of the red giant branch. *The Astrophysical Journal*, 882(1):34.

Giblin Jr, J. T., Mertens, J. B., and Starkman, G. D. (2016). Observable deviations from homogeneity in an inhomogeneous universe. *The Astrophysical Journal*, 833(2):247.

Goodman, J. (1995). Geocentrism reexamined. *Physical Review D*, 52(4):1821.

Green, S. R. and Wald, R. M. (2011). New framework for analyzing the effects of small scale inhomogeneities in cosmology. *Physical Review D*, 83(8):084020.

Green, S. R. and Wald, R. M. (2014). How well is our universe described by an FLRW model? *Classical and Quantum Gravity*, 31(23):234003.

Green, S. R. and Wald, R. M. (2015). Comments on backreaction. *arXiv preprint arXiv:1506.06452*.

Green, S. R. and Wald, R. M. (2016). A simple, heuristic derivation of our 'no backreaction' results. *Classical and Quantum Gravity*, 33(12):125027.

Gunn, J. E. (1967). On the propagation of light in inhomogeneous cosmologies. I. Mean effects. *The Astrophysical Journal*, 150:737.

Hacking, I. (1979). Imre Lakatos's Philosophy of Science. *The British Journal for the Philosophy of Science*, 30(4):381–402.

Kolb, E. W., Matarrese, S., and Riotto, A. (2006). On cosmic acceleration without dark energy. *New Journal of Physics*, 8(12):322.

Lakatos, I. (1980). *The Methodology of Scientific Research Programmes: Volume 1: Philosophical Papers*, volume 1. Cambridge University Press.

Laudan, L. (1977). *Progress and its problems.* University of California Press, Berkeley, California.

Liddle, A. and Lyth, D. (2000). *Cosmological Inflation and Large-Scale Structure*. Cambridge University Press, Cambridge.

Longair, M. (2019). Observational and astrophysical cosmology: 1980 to 2018. In Kragh, H. and Longair, M., editors, *Oxford Handbook of the History of Modern Cosmology*, pages 376–423. Oxford University Press, Oxford.

Macpherson, H. J., Lasky, P. D., and Price, D. J. (2018). The trouble with Hubble: Local versus global expansion rates in inhomogeneous cosmological simulations with numerical relativity. *The Astrophysical Journal*, 865(1):L4.

Martin, J. (2018). The theory of inflation. *arXiv preprint arxiv:1807.11075*.

McGaugh, S. S. (2014). A tale of two paradigms: the mutual incommensurability of ΛCDM and MOND. *Canadian Journal of Physics*, 93(2):250–259.

Merritt, D. (2017). Cosmology and convention. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 57:41–52.

Misner, C. W. (1968). The isotropy of the universe. *Astrophysical Journal*, 151:431–457.

Ostriker, J. P. and Steinhardt, P. J. (1995). The observational case for a low-density universe with a non-zero cosmological constant. *Nature*, 377(6550):600.

Peebles, P. J. E. (2005). Probing general relativity on the scales of cosmology. In *General Relativity and Gravitation*, pages 106–117. World Scientific.

Roush, S. (2003). Copernicus, Kant, and the anthropic cosmological principles. *Studies In History and Philosophy of Modern Physics*, 34(1):5–35.

Rsnen, S. (2004). Dark energy from back-reaction. *Journal of Cosmology and Astroparticle Physics*, 2004(02):003.

Sandage, A. (1961). Ability of the 200-inch telescope to discriminate between world models. *Astrophysical Journal*, 133:355–380.

Sandage, A. (1970). Cosmology: A search for two numbers. *Physics Today*, 23:33–43.

Smith, G. E. (2001). The Newtonian style in Book II of the *Principia*. In Buchwald, J. Z. and Cohen, I. B., editors, *Isaac Newton's Natural Philosophy*, Dibner Institute Studies in the History of Science and Technology, pages 249–314, Cambridge. MIT Press.

Smith, G. E. (2002). From the phenomenon of the ellipse to an inverse-square force: Why not? In Malament, D., editor, *Reading Natural Philosophy*, pages 31–70. Open Court, Chicago.

Smith, G. E. (2014). Closing the loop. In Biener, Z. and Schliesser, E., editors, *Newton and Empiricism*, pages 262–351. Oxford University Press, Oxford.

Stein, H. (1958). *An Examination of Some Aspects of Natural Science*. PhD thesis, University of Chicago, Department of Philosophy.

Stein, H. (1989). Yes, but... some skeptical remarks on realism and anti-realism. *Dialectica*, pages 47–65.

Stein, H. (1992). Was Carnap entirely wrong, after all? *Synthese*, 93(1-2):275–295.

Stein, H. (1994). Some reflections on the structure of our knowledge in physics. In *Logic, Metholodogy and Philosophy of Science, Proceedings of the Ninth International Congress of Logic, Methodology and Philosophy of Science*, pages 633–55.

Stein, H. (2014). Newton: Philosophy of inquiry and metaphysics of nature.

Tanabashi, M., Hagiwara, K., Hikasa, K., Nakamura, K., Sumino, Y., Takahashi, F., Tanaka, J., Agashe, K., Aielli, G., Amsler, C., et al. (2018). Review of particle physics. *Physical Review D*, 98(3):030001.

Torretti, R. (2000). Spacetime models of the world. *Studies in the History and Philosophy of Modern Physics*, 31(2):171–186.

Wetterich, C. (2003). Can structure formation influence the cosmological evolution? *Physical Review D*, 67(4):043513.

Zel'Dovich, Y. B. (1964). Observations in a universe homogeneous in the mean. *Soviet Astronomy*, 8:13.