13
Philosophical Aspects of Cosmology

Chris Smeenk

13.1 Introduction

Throughout the last century, cosmologists have revisited, time and again, basic questions regarding the appropriate aims of cosmology and how best to achieve them. These debates reflect a tension between the cosmologists' ambitions to provide a scientific account of the structure, evolution, and origins of the universe, and methodological limitations imposed by cosmology's unusual object of study. To the extent that cosmology proceeds by observational study of a unique object, the universe-as-a-whole, the understanding of method relevant to other areas of physics – which depend on experimental manipulation, access to a collection of similar systems, or a combination of both – does not straightforwardly apply. Progress in cosmology has depended in part on further insights into scientific inquiry itself, to clarify how cosmology should proceed despite this contrast.

Prior to the twentieth century, many philosophers held, like Kant, that, because they cannot emulate the methods used in other fields to achieve secure knowledge, cosmologists can never reach their ambitious goals. The unambigious progress in cosmology, in the century following Einstein's introduction of the first relativistic cosmological model in 1917, shows that the skeptics were mistaken. Yet this progress has been accompanied by ongoing debates regarding the proper aims and methods of cosmology. Historically these debates can be roughly divided into two periods. The first period featured intense debates about what qualifies as a satisfactory cosmological theory, at a time when there was not a consensus view and the sparse evidence available left many theoretical avenues open for exploration. Debates among proponents of different views often appealed to philososophical considerations. Advocates of the steady-state theory, in particular, argued that their approach was the only legitimate way to pursue cosmology scientifically. By the late 60s, many of the avenues explored earlier had reached dead ends. Following the widespread acceptance of the hot big bang model, the focus shifted to questions raised by this model and extensions of it. For example, is a purely scientific account of the origins of the universe possible – and if so, how does a "theory" of origins compare to other scientific theories? Can we explain features of the observed universe if we treat it as part of a much larger "multiverse," and if they succeed, to what extent do such explanations justify accepting the multiverse? At the end of a century of relativistic cosmology, there is a remarkable juxtaposition between the rich, diverse evidence for cosmology's standard model, and strident debates about such foundational questions.

The debates in both periods reveal implicit views about philosophy of science – regarding what constitutes a properly scientific theory, what counts as compelling evidence, and what kinds of explanatory demands a theory should meet. Cosmologists have occasionally engaged in explicit arguments about these issues. But more often positions on these issues are implicit in cosmologists' arguments and choices regarding what lines of research to pursue. One aim of this chapter is to clearly state positions on four central issues. Providing an adequate characterization of what is at stake in these debates will hopefully set the stage for the more challenging project of assessing and defending various positions. I will not pursue that project here, focusing instead on providing an overview of the debates and their role in the historical development of cosmology.

This chapter is organized thematically rather than chronologically, with each section below considering one of four inter-related issues: (i) the uniqueness of the universe and its methodological implications; (ii) the underdetermination of theory by evidence; (iii) theories of the origins of the universe; and (iv) anthropic reasoning and multiverse theories. The main shortcoming of this structure is that it threatens to downplay the historical evolution of the field. This is not a great loss since philosophy of cosmology has primarily been driven by research trends within cosmology itself, rather than following its own internal dynamics. As a result, I have emphasized how conceptions of the appropriate aims of cosmology inquiry have shifted in response to the historical evolution of the field, characterized briefly here and in much more detail in other contributions to this volume. Finally, there are several topics that I do not have the space to discuss. Those seeking a more systematic overview of the philosophy of cosmology should consult, in particular, Ellis (2007).[1]

13.2 Uniqueness of Cosmology

Cosmologists have had impressive success in extrapolating local physics, applying the laws governing the fundamental forces to domains increasingly far removed from those where they were originally established. This approach treats cosmology as, in Bondi's (1950) words, "the largest workshop in which we may assemble equipment, the elements of which are entirely composed of terrestrially verified laws of physics" (p. 4). On this view, cosmology proceeds by extrapolating the laws of local physics, discovered by studying sub-systems of the universe, to much larger scales. Einstein's seminal paper (Einstein, 1917a) showed that it was indeed possible to construct a consistent cosmological model based on general relativity. The subsequent development of relativistic cosmological models is, in part, based on assembling further "equipment" crafted in other areas of physics.

This approach has been challenged repeatedly on the grounds that the distinctive subject matter of cosmology requires something other than physics as usual.[2] Suppose that the proper subject matter of cosmology is the whole universe, or more precisely a maximally extended, connected spacetime of which the observed universe is a part.[3] Then cosmology would require a distinctively global view, not limited to the study of large-scale structure of the observed universe. In relativistic cosmology, it is at least possible to define global properties of mathematical models that represent the whole universe. But it is less clear how to develop and justify a physical theory of the whole universe and its global properties, because familiar distinctions from other theories do not apply.

The universe can be neither compared to an ensemble of other similar objects, as in other observational sciences, nor manipulated experimentally. By contrast, the physics of

projectile motion, for example, describes a space of dynamically allowed trajectories, including (approximations to) actual trajectories as well as possible trajectories that could be realized with different initial conditions. In cosmology, "The distinction between impossible and possible, but "accidentally" not realized states, becomes absurd when we have to deal with something as fundamentally unique as the universe" (Bondi, 1948, p. 106). On the one hand, a distinction between laws and initial conditions seems at least superfluous, if not absurd, in describing a unique object – such as a single trajectory, or the whole universe. On the other hand, dispensing with theory entirely would leave us without the tools needed to formulate a description. Critics of the standard approach agree that it is a mistake to treat possibility and necessity in cosmology just as in other areas of physics. But there is little consensus regarding what this implies for the aim and structure of cosmological theories.

One line of thought holds that cosmology requires something more than local laws to offer satisfactory explanations. Cosmological theories constructed as assemblages of local physical laws are too liberal: they allow many possible cosmological models, most of which bear little resemblance to the observed universe. Cosmology, regarded as merely an extrapolation of local physics, fails to explain why the actual universe must have the properties that it does. Instead many properties follow from initial conditions rather than the laws; they apparently hold merely contingently, as a result of "accidental realization," not as a matter of necessity. To close this explanatory gap, on this line of thought, a theory dealing with the universe-as-a-whole must introduce laws that rule out many of the models allowed by extrapolations of local physics. Such a theory would be able to explain why the universe has to have the features it does.

Critics of the standard approach have also challenged the idea that it will identify the correct laws.[4] The laws of local physics, they argue, are established based on the study of particular subsystems, typically regarded as isolated, or without external influences – in effect as "island universes". Treating subsystems as completely isolated in this sense is, however, at best an approximation in theories with universal interactions. In the case of gravity, for example, it is impossible to entirely "screen off" the effects of distant bodies. Consider a Newtonian treatment of the solar system. The equations of motion derived for the solar system treated as an island universe differ from those that follow from treating it as a part of a larger system, such as the Milky Way. These differences are negligible, of course, given the distance from the Sun to nearby stars. Treating the system as isolated, even if it is a good approximation, is potentially misleading. The full physical description of a larger system may include interactions among subsystems that is effaced by such an approximation. Using the laws discovered for a subsystem, regarded as an island universe, as the basis for extrapolation excludes such interactions. Of particular concern for cosmology, could local physics be coupled to large-scale properties that vary on cosmological scales? If so, we cannot accurately identify the laws relevant to cosmology via extrapolation from local laws.

Reflections along these lines have inspired very different proposals for how cosmology should proceed. Arguments about the nature of the field were a recurring theme in the roughly first half century of relativistic cosmology. The steady-state theory, in particular, put these methodological concerns front and center. Its proponents argued that certain principles must be accepted in order to make cosmology properly scientific.[5] Bondi (1948) took the inability to draw a contrast between accidental and lawlike features to have "obscure" implications, but for Bondi and Gold (1948) it inspired a new theory based on the "perfect cosmological principle." Local laws cannot, in general, be reliably extrapolated because of their possible dependence on

large-scale properties of the universe. There is no obstacle to constructing models of the whole universe, however, if the perfect cosmological principle holds: it requires that the universe is stationary in time, with unchanging large scale properties. Bondi and Gold derived several consequences of this principle, leading to what came to be called the steady-state theory.[6] This theory had a substantial impact on research in cosmology for the next 15 years.

Recently the theoretical physicist Lee Smolin has defended a view that is, in some respects, the opposite of the steady-state theory, despite a similar stance on the distinctive nature of cosmology (Unger and Smolin, 2015; Smolin, 2015). Although this line of work has not had the broad impact of the steady-state theory (at least, not yet), it illustrates the persistence of debates regarding how cosmology should be pursued. According to Smolin, the standard approach mistakenly applies what he calls the "Newtonian paradigm," appropriate for the study of subsytems, to the universe as a whole. Insofar as laws apply to subsystems, they are approximate because they leave out interactions with other subsystems; yet if a law encompasses the entire universe, it applies to a single case, and is no longer a law. Smolin proposes to resolve this dilemma by introducing a distinctive understanding of laws: the laws of nature evolve with respect to ("real") global time. Far from being a threat, Smolin sees evolving laws as the key to answering basic questions about why the universe has the properties that it does.

Both proposals demand a great deal of cosmological theories. Smolin endorses a version of Leibniz's principle of sufficient reason: a satisfactory cosmological theory must explain why the universe has the properties it does, and treating our universe as merely one part of a multiverse, discussed in Sect. 6, does not suffice. Advocates of the steady-state theory had similar commitments:

> [We must] find some way of eliminating the need for an initial condition to be specified. Only then will the universe be subject to the rule of theory. ... [A cosmological theory] should imply that the universe contains no accidental features whatsoever. This provides us with a criterion for assessing the validity of rival theories. We believe this criterion to be so compelling that the theory of the universe which best conforms us to it is almost certain to be right. (Sciama, 1959, pp. 166-67)

Sciama would himself soon abandon the steady-state theory, based on the more compelling criterion of empirical adequacy. But Sciama's position that cosmological theories should not leave so much room for contingency, by allowing a variety of possibilities, still retains adherents.

The idea that cosmological theories must provide such explanations has been as controversial as it is persistent. In response to the steady-state theory, the philosopher Milton Munitz criticized such rationalist demands (Munitz, 1952) and offered an alternative account of the kind of explanations that cosmology should pursue. Rather than trying to show why things must be as they are, cosmologists should aim, on Munitz's view, to provide a coherent description of the structure of the observed universe and its evolution, along with an understanding of how the part of the universe we can see fits into the whole universe (Munitz, 1962).[7] More recently, Ellis (2007) has noted that cosmology may pursue explanatory aims like those in historical sciences, such as paleontology and evolutionary biology. Developing a historical reconsctruction has proceeded successfully in these areas, despite limitations similar to those faced by cosmologists: an inability to manipulate or experiment with the system under study, or compare it to an ensemble of similar systems. This success is obtained without subjecting the past to "the rule of theory": explanations in historical sciences typically depend on assumptions regarding earlier states, but are not rejected as incomplete or unsatisfactory as a

result. The demand to go beyond this and give an explanation of why the earlier state had to obtain, as a matter of necessity, reflects the physicists' interest in discovering laws. But this methodological orientation is not necessary if the aim of cosmology is limited, as with the other historical sciences, to developing and justifying a particular historical reconstruction.

In addition to this deflationary response to the demand for explanations, many cosmologists would object to the very first step above: why should cosmology be defined as the science of the whole universe? Rovelli and Vidotto (2014), for example, reject this in no uncertain terms:

> Cosmology is *not* the study of the totality of the things in the universe. It is the study of a few very large-scale degrees of freedom. (p. 215, original emphasis)

Many cosmologists draw a similar contrast, declaring that physical cosmology consists in the proposal and observational testing of cosmological models assembled from local physics. This is not a minor terminological point: insofar as cosmology is the science of large scale structures and their dynamical evolution, there is no need to develop a distinctive methodology compared to other physical theories. (As we will see below in Sect. 4, however, methodological questions arise again regarding attempts to give a "theory" of origins, or of the initial state.)

The calls for an alternative methodology have stemmed in part from doubts about the viability of a physics-as-usual approach. When the steady-state theory was first proposed, expanding universe models faced significant empirical problems that have since been resolved, such as the "age crisis" and the lack of a plausible account of structure formation.[8] Cosmology's subsequent track record of building successful models based on extrapolating local physics undercuts the appeal of alternative methodologies. As the example of Smolin illustrates, however, it is still possible to regard cosmology as succeeding in spite of confusion about foundational issues; on Smolin's view, cosmology is in a state of crisis, as reflected in widespread acceptance of the multiverse idea (discussed in Sect. 6). The fact that the standard methodology has succeeded reflects an empirical fact: namely, that any interplay between physics at global and local scales is sufficiently weak that it does not hinder the bottom-up construction of successful cosmological models. Persistent failure to develop a satisfactory cosmological theory might lead us to reconsider whether this is the case. It is easy to conceive of worlds in which this approach to cosmology would not be productive. For example, if gravity at solar system scales depended directly on large-scale properties, then we could not straightforwardly apply general relativity to distant regions. It would be challenging to pursue cosmology in such a world, given the difficulty of obtaining evidence regarding such functional relationships. We appear to live instead in a world in which cosmologists can thrive, and build up a successful account of the universe by extrapolating local physics. This deflationary response to the concerns of Bondi, Smolin, and others treats the success of the standard approach as itself contingent on the nature of our universe.

13.3 Underdetermination

Scientists inevitably pursue questions that evidence available at a given time cannot answer. Uncertainty is often transient, resolved with the next step in a research program, but in some cases there are permanent obstacles to obtaining decisive evidence. The extent to which evidence can settle theoretical questions – the "underdetermination of theory by evidence" – is a central theme in philosophy of science. Given its aim to describe the universe and its evolution at large scales, and the limited evidence available, it would not be surprising for transient and permanent underdetermination to be ubiquitous in cosmology.

Cosmologists' expectations regarding whether evidence can settle the central questions of their field have shifted profoundly over the last century. At mid-century cosmology was regarded as closer to mathematics, or even philosophy, than empirical science. Whitrow argued for this position in a debate with Bondi in 1954 (Whitrow and Bondi, 1954); Bondi, by contrast, was more optimistic that empirical evidence would resolve foundational disputes in the field, and emphasized that the steady-state theory at least made definite predictions. A decade later, Trautman expressed a common view in the epilogue of his Brandeis lectures on general relativity: "... it is not worthwhile to work in theoretical cosmology at the present time. I think it would be better to sit and wait for the astronomers to get more data on the motion and distribution of distant galaxies" (Trautman, 1965). Trautman's expectation that the wait would not be long was correct. Source counts derived from the 4C survey, available shortly thereafter, showed an increase in sources at high redshift incompatible with the sharp predictions made by the steady-state theory emphasized by Bondi. The serendipitous discovery of the cosmic microwave background (CMB) by Penzias and Wilson (1965) was even more significant. It encouraged cosmologists to take extrapolations of the big bang models to early times, and the application of nuclear physics to that regime, seriously (as emphasized by Weinberg, 1977). Furthermore, it provided a target for precision measurements that depends on relatively well understood physics rather than relying on galaxies and other complicated systems as standard candles. Precision measurements of the CMB have been combined with other lines of evidence, leading to consistent estimations of the parameter values in the standard model of cosmology. Cosmologists now have sufficient confidence in this model to assert the existence of new types of matter, and other additions to the standard model of particle physics, based on their cosmological effects.

This shift reflects an effective response to underdetermination, despite severe limitations on the available evidence. Here I will assess different aspects of underdetermination, to elucidate the reasons for this shift and the current limits to what evidence can establish.

13.3.1 Horizons

A particularly clear observational limit follows from the finitude of the speed of light. Physical signals moving at or below the speed of light can reach us only from a limited region of spacetime. If we represent our location as point p in a relativistic spacetime, the in principle accessible region consists of the past light cone (and its interior) at p, also called the causal past, $J^-(p)$.[9] The limits to observational access are often described instead in terms of the existence of horizons, of different types (Rindler, 1956).[10] Horizons measure the maximum distance from which a signal emitted at a specified time t_e can reach an observer. The visual horizon takes t_e to be the decoupling time, prior to which the universe is opaque to electromagnetic radiation; whereas the particle horizon is defined as the limit $t_e \to 0$. Objects at positions separated by distances greater than the particle horizon have non-overlapping past light cones. Event horizons, by contrast, are defined as the boundary of the causal past in the limit as $t_0 \to \infty$ — the limit of what could be seen by an "immortal" observer.

Discussions of horizons go back to the advent of relativistic cosmology, when Einstein took the existence of an event horizon in de Sitter's solution as a reason to doubt its physical viability. Rindler's work was prompted by a similar debate regarding the steady-state theory (Whitrow, 1953), which represented spacetime geometry with part of the de Sitter solution. The existence of horizons in a cosmological model has various counter-intuitive consequences,

but does not provide sufficient reason to reject a model outright. There is no *a priori* reason to demand that the universe can, in principle, be exhaustively surveyed by observers within it. But to what extent does the existence of horizons limit cosmologists' ability to answer central questions?

McCrea (1960, 1962) addressed this issue explicitly, arguing that observations in cosmology have an inherent uncertainty that increases with redshift (z). The argument starts from the legitimate observation that, except in very unusual cases, information available at a point p cannot be used to predict physical properties at a distinct point p'. Because the past light cones of any distinct points, $J^-(p)$ and $J^-(p')$, do not completely overlap, an observer at p will not be able to determine some of the physical processes that can affect the physical state at p'.[11] Any prediction thus requires some assumptions regarding what lies beyond $J^-(p)$ — for example, that there is no "source-free radiation" propagating to p', or that the physical properties of $J^-(p')$ should resemble those of $J^-(p)$ in some specific respects. General relativity does not impose such constraints, so they must come from some other source. As we will see shortly, the cosmological principle could play this role, but this is a substantial further assumption – that McCrea did not accept. He gave an obscure argument that the uncertainty associated with predictions increases with redshift (linearly with z), because the amount of relevant information decreases. The consequences McCrea claimed to find for the central cosmological question of his day are clear: this uncertainty allegedly undercut efforts to discriminate between the steady-state theory and evolutionary cosmological models.

This pessimistic conclusion was overstated, but the characterization of the epistemic predicament faced by cosmologists is apt: to what extent can we answer cosmological questions based on observations confined to $J^-(p)$? Considering an idealized data set can help to distinguish between limitations that arise from the finite speed of light and from other sources. (This is not to deny that various other sources of uncertainty are far more important to observational cosmology.) To that end, we will imagine cosmologists inhabiting a universe filled with "standard objects" whose properties are fully understood (including luminosity, mass, shape, and so on, and the evolution of these properties with cosmic time), and which are targeted by sophisticated, well-funded observational programs. This idealization eliminates the uncertainty associated with "acts of faith" required in real observational cosmology, in using galaxies and other complicated systems as standard objects (see Longair's discussion in Chap. 10 of this volume). This approach further assumes that classical general relativity holds, but does not impose other constraints on the background spacetime geometry. Given access to this data set, what questions could cosmologists then answer?

Ellis et al. (1985) proved that an appropriate idealized data set of this kind is sufficient, granting that general relativity holds, to determine the spacetime geometry and distribution of matter on the past light cone $C^-(p)$ (out to the maximum redshift at which this ideal set can be observed).[12] For the ideal data set, observations can directly determine the area (or luminosity) distance of the sources, and the distortion of distant images determines lensing effects. These observations directly constrain the spacetime geometry of the past light cone $C^-(p)$. Although the ideal objects can be used effectively as "tracers" to determine the spacetime geometry, further substantive modeling assumptions are required to make claims regarding the distribution of matter and energy. (In particular, modeling assumptions are needed to resolve various degeneracies, and disambiguate the effects of different kinds of matter and energy, including dark matter and dark energy, on the ideal data set.) Observers do not have access

to anything like the ideal data set, obviously, and in practice there are substantial obstacles to determining the spacetime geometry in this fashion because of uncertainty regarding the nature of the standard objects and their evolution with cosmic time.[13]

There are exotic cases in which cosmologists would be able to determine *global* properties of spacetime, and not only the spacetime geometry of $J^-(p)$, using such an ideal data set.[14] In some cosmological models, well-situated observers could see the entire universe. These models are baroque variations on Einstein's closed universe: space at a given cosmic time is finite and without boundary, with an intricate topology. A common feature of these models is that observers can see multiple "ghost images" of a single object, given that light can reach an observer along many different paths through the spacetime. If the maximum spatial length in every direction is shorter than the visual horizon, observers would be able to "see around the universe," and (in principle) fully determine its spacetime geometry.[15]

Aside from these exotic cases, the finite speed of light poses a fundamental obstacle to determining global properties of spacetime empirically. To what extent does a single observer's window on the universe, or even a collection of such views, fix the overall structure of a cosmological model? This question, initially posed by physicists studying the causal structure of relativistic spacetimes, was taken up by the philosophers Glymour and Malament in the late 70s, and revisited by Manchak in the 2000s. The Malament-Manchak formulation of the problem starts by defining what it means for two spacetimes to be "observationally indistinguishable," and then asks what global properties must be shared by indistinguishable spacetimes. Two spacetimes are indistinguishable if and only if for every observer p in the first spacetime, there is a "copy" of their $I^-(p)$ in the second spacetime.[16] Given this definition, there is a clear procedure for settling the general question: for a given global property $\mathcal{G}$, and an initial spacetime that has the property, is it possible to construct an indistinguishable counterpart that lacks $\mathcal{G}$?[17] Malament (1977) and Manchak (2009) give a series of clever constructions establishing that this is so for nearly all global properties. Since, by construction, the full data available to any observer is compatible with either counterpart, the spacetime geometry of $I^-(p)$ does not suffice to establish that $\mathcal{G}$ holds. The only properties that are guaranteed to hold for an indistinguishable counterpart are those that can be established based on the chronological past of a single point.

These results do not pose a challenge as stark as that suggested by McCrea: they do not show that observational evidence will necessarily fall short in answering central theoretical questions. Although cosmologists often speculate about the global structure of the universe on enormous scales, well beyond our past light cone, these claims play no direct role in evidential reasoning in cosmology. What does play a crucial role is the spacetime geometry of $J^-(p)$ itself, which we will turn to in the next section. Furthermore, the choice at hand is between cosmological *models* rather than *theories*, since we have assumed throughout that classical general relativity holds.

The extent to which these results support, nonetheless, an interesting claim of "model underdetermination" depends on one's view of inductive inference. More straightforward cases of induction can be described in similar terms: evidence that a particular regularity obtained in the past is compatible with a model according to which it continues to hold, but it is also compatible with a "counterpart" in which the regularity fails to hold at some point in the future. Only adherents of a very strict empiricism hold that we have no reason to prefer a model in which observed uniformities can be projected to new cases. Accounts of induction,

or ampliative inference, seek to provide some justification for choosing among models that are all logically compatible with the available evidence. The challenge in this case is to clarify what justifies accepting one spacetime over its indistinguishable counterparts (Earman, 2009; Norton, 2011; Butterfield, 2014).

13.3.2 The Cosmological Principle

The discussion above contrasts with a more typical way of using evidence, namely to choose an optimal cosmological model from a restricted set of solutions. Rather than considering the full space of solutions to general relativity, one restricts attention to a space of symmetric cosmological models (or models sharing some other global properties). Sandage's observational program (Sandage, 1961a, 1970), for example, was devoted to determining the values of two parameters sufficient to determine the "best-fit" expanding universe model. (These parameters were the Hubble's constant H_0 and the deceleration parameter q_0, which fix a model provided that $\Lambda = 0$.) The expanding universe models discovered by Friedman and Lemaître (hereafter, FL models) are *isotropic* (there are no geometrically preferred spatial directions) and *homogeneous* (at a given moment of cosmic time every spatial point "looks the same").[18] Due to this symmetry, the models have a particularly simple structure: spacetime can be decomposed into three dimensional hypersurfaces of constant cosmic time $\Sigma(t)$ (topologically, $\Sigma \times \mathbb{R}$), and the field equations of general relativity reduce to a pair of ordinary differential equations. These equations fix the behavoir of the scale factor $R(t)$ given the equations of state for the different types of matter present. (The scale factor represents the spatial distance in Σ between observers moving along timelike geodesics, as a function of cosmic time.) Generalizing from Sandage's two numbers, the contribution of various types of matter can be characterized in terms of dimensionless density parameters. On this approach, observations are used to determine the "best fit" model and fix the relevant parameter values.

Even this drastically oversimplified sketch of model choice suffices to illustrate the significance of the symmetry principle assumed at the outset. The empiricist approach considered in the previous section did not impose any global constraints on the space of models. Imposing homogeneity and isotropy, by contrast, is extremely restrictive and enables evidence from a limited region to determine global properties. An observer can take their observations to reveal properties of, not just the spacetime region within the past light cone, but other regions related by the symmetry. They are then entitled to claims about global structure. Other symmetry principles, or stipulations that lead to a limited space of viable models, also allow local-to-global inferences. The Copernican principle is sometimes formulated as the requirement that we do not occupy a "privileged position" in the universe. (Excessive modesty may, however, lead us astray, if we fail to acknowledge that there are various ways in which our location is privileged in the sense of being suitable for life – a topic we will return to in Sect. 5.) More generally, we can require that there are no "special locations" in the universe: no point p is distinguished from other points q by any spacetime symmetries, or lack thereof.

There have been four quite distinctive ways that cosmologists have tried to justify imposing such a global symmetry, or restricting consideration to a subset of possible models on similar grounds (see also Beisbart, 2009). Milne (1933)'s influential discussion introduced the "cosmological principle" as the requirement that the universe must appear to be the same to all observers. Homogeneity and isotropy were sufficient to secure this form of equality. Milne regarded the cosmological principle as the most important axiom of a deductive system leading

to a distinctive theory, kinematical relativity, rather than a claim in need of empirical support. Requiring that the principle would hold addressed the lack of predictivity for which he faulted relativistic cosmology. Bondi and Gold were influenced by this line of thought, and took a similar stance toward their proposed modification, the perfect cosmological principle. Yet after the fall from favor of the steady-state theory, few cosmologists have treated the cosmological principle as an *a priori* principle needed to formulate a cosmological theory.

In early discussions of the FL models, it was far more common to treat their high degree of symmetry as a useful simplification that should not be taken too seriously. Tolman emphasized the mathematical value of the simple FL models, but did not take observations to provide strong justification for accepting the cosmological principle. Tolman (1934b) concluded with a warning that:[19]

> [W]e must be careful not to substitute the comfortable certainties of some simple mathematical model in place of the great complexities of the actual universe. (Tolman, 1934b, p. 487)

The qualitative match between the FL models and observations, such as Hubble's measurements of the redshift-distance relation, encouraged the thought that the study of these models was not a purely mathematical exercise. Yet cosmologists were wary of accepting features of these models that depended on these symmetries holding exactly. Einstein and Tolman, for example, regarded early singularity theorems (which showed that curvature blows up as $t \rightarrow 0$ in the FL models) as artifacts of physically unreasonable idealizations.[20] The same reasons would undermine trust in extrapolations of the FL models to the early universe.

The situation changed dramatically with the discovery of the CMB in 1965: its isotropy provided much stronger evidence that the FL models apply, even to the very early universe and at the largest possible scales. But many cosmologists were puzzled rather than exhilirated by the success of the FL models; if anything, the models seemed to work *too* well. Why should the universe be so symmetric? The isotropy of the universe went from being a simplifying assumption to a target for physical explanations; as Misner (1968a) put it,

> [The isotropy of the CMB] surely deserves a better explanation than is provided by the postulate that the Universe, from the beginning, was remarkably symmetric. (p. 431)

Misner's "chaotic cosmology" program aimed to explain isotropy as the consequence of dynamical effects in the early universe (specifically, damping of anisotropies due to neutrino viscosity).[21] Although this particular proposal was ultimately unsuccessful, it suggested a new "philosophy for big bang cosmology":

> The universe must start with a big bang (or bangs) and, almost independently of any special initial conditions, it must have a particular chemical composition, it must exhibit a Hubble expansion, it must be isotropic if it is homogeneous, and we do expect it to be homogeneous (McCrea, 1970, p. 22).

(We should expect the universe to be homogeneous due to the Copernican principle.) Misner and McCrea both clearly prefer a theory that is "indifferent" to the exact features of the initial state, with dynamics that drive a large range of possible initial states to converge to properties compatible with the observed universe (such as the specific chemical composition expected as a result of big bang nucleosynthesis, and isotropy as a consequence of Misner's proposal). This conception of what constitutes a successful early universe theory was more durable than the specific proposals McCrea discussed (see also McMullin, 1993). A decade later, Guth (1981) made a persuasive case that a different dynamical mechanism in the early universe,

driving a stage of inflationary expansion, should be explored further because it promised to be successful in just this sense.

The final historical shift in the status of the cosmological principle has occurred within the last two decades: it has been scrutinized more carefully due to its essential role in precision cosmology. Insofar as the universe is well-approximated by an FL model, observations of the redshift-distance relation for standard objects can be used to constrain the density parameters characterizing different types of matter. In particular, observations of this type using supernovae (type Ia) as a standard candle led to the remarkable discovery that the expansion of the universe is accelerating (see Chap. 11 for further discussion). For this to be true in an FL-model (that is, for $\ddot{R}(t) > 0$), there must be some source that has approximately the same dynamical effects as a positive cosmological constant (often referred to as "dark energy"). This inference depends crucially not just on the physics linking $R(t)$ to different types of matter and energy, but on the accuracy of the FL-models as a description of spacetime geometry at the relevant scales. The accuracy of the FL models has shifted from being a target of physical explanation to an essential ingredient of an observational program.

One way of clarifying the status of the FL models is to carry out something like the observational cosmology program described above, with observations of the CMB as a crucial addition to the "ideal data set". Observers can use the CMB to put upper bounds on the isotropy of the universe from one point (once their proper motion with respect to the CMB is taken into account). If the Copernican Principle holds, then the observed isotropy holds generally rather than as a special property of this point – suggesting that the FL geometry holds. Cosmologists can now do much better than invoking the Copernican principle. There are several generalizations of the seminal result due to Ehlers et al. (1968), which establish what different types of observations – such as that of an istropic radiation field, by an observer moving along a geodesic – imply regarding spacetime geometry. Furthermore, observations of the CMB can provide indirect evidence regarding whether the universe appears to be isotropic from the vantage point of distant galaxies (Goodman, 1995). One type of indirect evidence takes advantage of the fact that the CMB has a black-body spectrum. Roughly put, scattering of CMB photons due to the Sunyaev-Zeldovich effect will lead to distortions in the spectrum if the CMB is anisotropic to the scatterer; if the CMB is isotropic, the spectrum will retain its black-body shape with a shift in temperature. Granted that the CMB has a black-body spectrum at decoupling, an observer who observes that the CMB has a black-body spectrum can then infer that it is also isotropic with respect to distant regions where the CMB photons are scattered. This line of argument, in conjunction with some other types of indirect evidence, can be used to make a direct empirical case that the universe is well-approximated by an FL model (Clarkson, 2012).

This raises a further question regarding the status of the FL models: how does their uniformity at large scales relate to the manifest lack of uniformity at smaller scales? This question was posed in the early days of relativistic cosmology, with the construction of models representing the gravitational field of stars (or other local systems) embedded within an expanding universe model (Einstein and Straus, 1945) or combined in a lattice (Lindquist and Wheeler, 1957). The huge density contrasts at galactic scales and smaller are certainly not "small perturbations" away from an FL model. What, then, does it mean to say that these models "approximate" the actual universe? If the FL models are taken to describe an "averaged" or smoothed out matter density, at some suitably large scale, there is a natural further question: do the FL dynamics

correctly describe the dynamical evolution of this "smoothed out" distribution? Ellis (1984) and Ellis and Stoeger (1987) argued that smoothing out a general relativistic solution does not also lead to a solution – that is, the averaged spacetime geometry and stress-energy tensor do not satisfy the field equations. The field equations can be satisfied if an "effective stress energy" tensor representing the back-reaction effect of the inhomogeneities is included. The degree to which the FL models are a good approximation can then be quantified in terms of the size of these additional terms.[22]

13.3.3 Status of the ΛCDM Model

A distinctive underdetermination challenge arises in considering the "best fit" model of some phenomena, based on a background physical theory. To what extent does success – that is, finding parameter values in the model consistent with a body of data – justify confidence in the accuracy of the theory? Perhaps the model succeeds due to its flexibility, by introducing new degrees of freedom that can be carefully tuned to reproduce the data without accurately representing the relevant physics. How can we discriminate among models that are all compatible with the data, but based on different underlying physics? (Such models may give similar descriptions of the phenomena being studied, but have different implications for other phenomena.) Here I will briefly assess the extent to which the success of the ΛCDM model, in fitting a wide array of cosmological observations with a small number of parameters, meets this underdetermination challenge.

One promising line of response to this challenge, formulated in general terms, is to demand, first, that there are multiple, independent ways of determining the parameters of the model, and, second, that the theory can be consistently applied in light of systematic improvements in measurement precision. The first demand exploits the theory's unification of diverse phenomena to "overdetermine" the parameters appearing in the model (see, e.g., Norton, 2000). In his defense of atomism, for example, Perrin (1923) emphasized the agreement among several strikingly diverse ways of determining Avogadro's number N, drawing on phenomena ranging from Brownian motion to the sky's color. The strength of this reply depends on the extent to which the phenomena probe the underlying theoretical assumptions in distinct ways, and whether there is an alternative hypothesis that also accounts for the agreement. As the number of methods used to determine a parameter increases, the probability that agreement among them can be attributed to chance, or to systematic errors, decreases. Turning to the second point, do increasingly precise measurements lead to refinements of the underlying model, or to anomalies? In several historical cases, a theory has guided the development of models that do meet steadily improving standards of precision, without setting aside core principles. Furthermore, the refined models often incorporate further details that can be independently checked (see, e.g. Smith, 2014a). Success in these two respects provides a strong response to the underdetermination challenge. If the underlying physics were false, it would be a coincidence for the multiple ways of measuring model parameters to agree, and it would be unlikely that increasingly precise measurements would lead to further discoveries rather than anomalies. Any rival theory should be expected to agree with a theory that is successful in this sense, at least as an approximation within the relevant domain.

Turning back to cosmology, the current standard model of cosmology, the ΛCDM model, fits an impressive array of cosmological data with a small number of parameters. These include the density parameters characterizing the abundance of different types of matter, each of which

can be measured by a variety of different types of observations.[23] There are two different questions that this model raises regarding the physics used in its construction: to what extent does the success of the model support, first, extrapolations of well-tested local physics, and second, novel physics tested only through cosmological applications? The model assumes the validity of extrapolating general relativity, for example, to length scales roughly 14 orders of magnitude greater than those where the theory is subject to high precision tests. There are also several aspects of the model based on novel physics that cannot be independently tested through terrestrial experiments.

The case in favor of the standard model has been strengthened considerably through precision measurements of the cosmological parameters. Many of the methods of measuring these parameters at the time Sandage formulated his observational program required a variety of astrophysical assumptions, regarding, for example, the use of galaxies as standard objects to determine spacetime geometry. Systematic uncertainties of more recent measurements of these parameters are easier to control, insofar as they rely on very well-understood physics. The CMB, in particular, provides powerful constraints on cosmological parameters due to our confidence in our physical description of recombination and of the subsequent propagation of the CMB photons. The consistent determination of these parameter values from many different types of observations supports an overdetermination argument much like Perrin's. In Perrin's case, accepting the atomic hypothesis implied that many different phenomena indirectly measure the scale of the atomic constituents of matter; any measurement incompatible with the others would cast doubt on the hypothesis. Similarly, the ΛCDM model leads to systematic connections between a diverse array of observable features of the universe. Peebles (2005), for example, enumerates 13 distinct ways of measuring the overall matter density Ω_0 at large scales: several distinct techniques based on using galaxies as mass tracers; weak lensing; cluster mass functions; the mass fluctuation power function; and so on. Accepting the ΛCDM model comes with an obligation to resolve any discrepancies among the various measurements of the basic parameters appearing in the model. While there are still open questions regarding discrepancies in some of these parameter measurements, the overall agreement among different measurements of the parameters appearing in the ΛCDM model provides strong evidence in its favor. (See Longair's Chap. 10 in this volume for an overview of measurements of the cosmological parameters.)

The strength of this case depends in part on whether the agreement among different measurements would be expected to hold even if the theories used in constructing the model were false. Perrin, for example, argued that the relationships between N and observable magnitudes, derived in the variety of cases he considered based on the atomic hypothesis, do not hold according to competing theories. Obviously, the evidence is not as decisive if several alternative theories imply that similar relationships hold. The strength of the evidence thus reflects how distinctive the theory under consideration is, as compared to the space of competing theories. This assessment is more challenging for aspects of the standard model that employ novel physics, due to the greater uncertainty regarding the space of viable alternatives.

Ellis (2007)'s idea of a "physics horizon" helps to clarify the status of different parts of the standard model. As with other horizons, the physics horizon marks the limit of what is accessible; in this case, it is the physical regime accessible to terrestrial experiments and non-cosmological observations. This is not nearly as sharply defined as the horizons discussed above, as it reflects an assessment of what experiments or observations are feasible, leading to

a rough division in terms of relevant energy or length scales. The qualifier "non-cosmological" is also admittedly vague, but it is intended to allow for observations, such as those of the solar system, that do not depend on a background cosmological model. For a specific theory it is possible to be more precise; for example, Baker et al. (2015) describe the regimes of parameter space characterizing gravitational theories that can be probed by different types of observations (solar system tests, gravitational waves, etc.). Several components of the standard model – such as dark matter, dark energy, and inflation – currently lie beyond the physics horizon in this sense. Although each proposal is based on plausible extensions of well established physical theories, currently the only way to evaluate these ideas is through their implications for cosmology. Insofar as they extend beyond the physics horizon, making a strong case in favor of these proposals based on multiple independent measurements, or developing more detailed models in response to systematic improvements in precision measurements, is particularly challenging.

The physics community has taken on this challenge because there are few other opportunities to test several of the most intriguing aspects of new ideas in fundamental physics. The Soviet cosmologist Yakov Zeldovich called the early universe the "poor man's accelerator," because relatively cheap observations of the early universe may reveal features of high-energy physics well beyond the reach of even the most expensive earth-bound accelerators. For many aspects of fundamental physics, in particular quantum gravity, cosmology provides the best testing ground for competing ideas. To what extent can cosmological observations replace other kinds of tests, such as accelerator experiments, in providing evidence for theories?

A brief discussion of three different cases of new physics incorporated in the ΛCDM model illustrates the challenge to providing evidence of comparable strength to that achieved in other areas of physics. For the last several decades, cosmological models have included a substantial contribution to the total matter density from non-baryonic dark matter. Dark matter was first proposed to account for the dynamical behavior of galaxy clusters and galaxies, which could not be explained with only the observed luminous matter. Dark matter plays a crucial role in accounts of structure formation, as it provides the scaffolding necessary for baryonic matter to clump, without conflicting with the uniformity of the CMB.[24] These inferences to the existence of dark matter, as well as many others, rely on gravitational physics. Obviously, it is possible that these observations reveal a flaw in our understanding of gravity rather than the presence of a new type of matter. There have been sustained efforts to clarify what form a modified gravity theory would have to take to account all of the relevant observations entirely (or mostly) without dark matter. In this context direct experimental detection of dark matter would make a decisive contribution. Several research groups aim to find dark matter particles through direct interactions with a solid-state detector, mediated by the weak force. A positive outcome of such experiments would provide evidence of the existence of dark matter that does not depend upon gravitational theory.[25]

This kind of decisive evidence is precisely what is in short supply for theories that extend beyond the "physics horizon." There is reason to hope that this situation is only temporary in the case of dark matter, but the prospects of providing independent evidence regarding the nature of "dark energy" are much worse. Cosmological models began to incorporate "dark energy" (in the sense of a non-zero cosmological constant Λ) in the early 90s, as an essential ingredient in accounts of structure formation. By the mid-90s, there was growing evidence in favor of a value $\Omega_\Lambda \approx 0.7$, from several different lines of evidence (Ostriker and Steinhardt,

1995). The case in favor of a non-zero Λ was strengthened considerably, and gained much more widespread attention, due to observations of the redshift-distance relation of supernovae published in 1998 (see, e.g. Frieman et al., 2008, for a summary of this line of work). These observations indicated that, granting the applicability of an FL model, the expansion of the universe is accelerating. In an FL model, accelerated expansion must be driven by a contribution that has the same dynamical effects as a non-zero Λ. Just as in the case of dark matter, there is of course the possibility that the various observations taken to motivate the introduction of dark energy instead indicate either a mistake in our description of gravity, or that the FL models do not apply. Both possibilities have been the focus of sustained research efforts (see, for example, Uzon (2010)). Unlike dark matter, however, the properties of dark energy ensure that any attempt at non-cosmological detection would be futile: the energy density is so small, and uniform, that any local experimental study of its properties is practically impossible.

Turning to the third case, inflationary cosmology originally promised a powerful unification of particle physics and cosmology. The earliest inflationary models explored the consequences of specific scalar fields introduced in particle physics (the Higgs field proposed in studies of the strong interactions). Yet theory soon shifted to treating the scalar field responsible for inflation as the "inflaton" field, leaving its relationship to particle physics unresolved, and the promise of unification unfulfilled (Zinkernagel, 2002). If the properties of the inflaton field are unconstrained, inflationary cosmology is extremely flexible: it is possible to reverse engineer an inflationary model that yields any chosen evolutionary history of the early universe.[26] Specific models of inflation, insofar as they specify the features of the field or fields driving inflation and its initial state, do have predictive content. In principle, cosmological observations could determine some of the properties of the inflaton field and so select among them (Martin et al., 2013). This could in principle then have implications for a variety of other experiments or observations. In practice, however, the features of the inflaton field in most viable models of inflation guarantee that it cannot be detected in other experimentally accessible regimes. The predictive content of inflation is further weakened if it leads to an inflationary multiverse, as discussed below.

The physics horizon poses a challenge because one particularly powerful type of evidence — direct experimental detection or observation, with no dependence on cosmological assumptions — is unavailable for the physics relevant in the very early universe, or at extremely large length scales. Yet this does not imply that competing theories, such as dark matter vs. modified gravity, should be given equal credence. The case in favor of dark matter draws on diverse phenomena, and it has been difficult to produce a compelling modified theory of gravity, consistent with general relativity, that captures the full range of phenomena as an alternative to dark matter. Cosmology typically demands a more intricate assessment of background assumptions, and the degree of independence of different tests, in evaluating proposed extensions of local physics.

13.4 Origins of the Universe

One profound shift marks a clear contrast between the first half-century of relativistic and the second. The idea that, at the largest observable scales, the universe does not evolve over time was no longer viable as of roughly 1965, as various observations effectively ruled out the steady-state theory. New theoretical arguments showed that the singularity known to be present in the FL models could not be dismissed as an artifact of idealizations, absent in more

realistic models. As a result of these developments, cosmologists had to take seriously the prospect that time has a beginning, and to ask whether it is possible to formulate a scientific theory governing the "origin of the universe," and if so what form such a theory might take.

Philosophers have been wary of proposing theories of origins in the aftermath of the incisive critiques of cosmological arguments for the existence of God due to Hume and Kant. Hume, in particular, argued that an understanding of causal relationships in ordinary circumstances does not illuminate the "causes" relevant to the origin of the universe. Rephrasing Hume, should we expect a "theory of origins" to have anything like the structure of other physical theories? What is the appropriate explanatory target for such a theory, and how does the explanation proceed? The concerns raised in Sect. 2 above seem particularly pressing here. Some cosmologists have sought to avoid an "origin" entirely. Hoyle (1975), for example, effectively demands that a good cosmological theory does not include a singular origin. Others regard the singularity as indicating the limits of applicability of classical general relativity, rather than an actual singularity; a theory of quantum gravity may lead to a fundamentally different picture. After a brief review of the arguments in favor of taking an initial singularity seriously, I will outline and assess the options for a theory of origins that have been explored.

13.4.1 Singularities

Contemporary cosmology at least has a clear target for a theory of origins: the best-fit FL model describes the universe as having expanded and evolved over ≈ 13.7 billion years. This "age of the universe" is the total proper time elapsed that would be measured by a clock moving along the worldline of a fundamental observer (moving along a geodesic), from the "origin" until now. Singularities are signaled by the existence of inextendible geodesics with bounded length. Extrapolating backwards from the present, an inextendible geodesic reaches an "edge" beyond which it cannot be extended; the finite age of the universe is the temporal distance to this "edge". This does *not* imply that there is a "first moment," just as there can be an open interval of the real number line of a specified length without a "first point."

Theorems establishing the existence of a singularity in the FL models (for example, Tolman, 1934b) follow from the Raychaudhuri equation, which describes the evolution of a set of nearby worldlines, such as those making up a small ball of dust. It takes on the following simple form in the FL models due to their symmetry (Ellis, 1971a):

$$3\frac{\ddot{R}}{R} = -4\pi G\,(\rho + 3p) + \Lambda\,, \tag{13.1}$$

where R is the scale factor and $\ddot{R}$ is its second derivative with respect to cosmic time. A small ball of dust (with $R(t)$ measuring the distance between nearby trajectories) changes volume as a function of time, in response to the mass-energy. (More generally, there can also be a volume-preserving distortion (shear) and rotation of the ball.) Given that the universe is currently expanding, (13.1) implies that the expansion began at some finite time in the past. As this "big bang" is approached, the energy density and curvature increase without bound provided $\rho + p > 0$ (which guarantees that $\rho \to \infty$ as $R \to 0$). 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 (13.1) – which illustrates the instability of gravitational collapse.

Obviously the symmetries of the FL models do not hold exactly in the actual universe, and it was essential to see whether the presence of singularities was robust to relaxing these

idealizations. The singularity theorems proved in the 60s (see, in particular, Hawking and Ellis, 1973) established that singularities still arise given much weaker, and more physically well-motivated, assumptions. The singularity theorems apply to a much broader class of models, many of which lack a uniquely defined "cosmic time." In these models there is not a cosmic time with a natural physical meaning, as in the FL models. The theorems still establish that the universe is finite to the past, in the sense that there is a maximum length for inextendible geodesics.

The singularity theorems plausibly apply to the observed universe, within the domain of applicability of general relativity. There are various related theorems differing in detail, but one common ingredient is an assumption that there is sufficient matter and energy present to guarantee that our past light cone refocuses.[27] The energy density of the CMB alone is sufficient to justify this assumption. The theorems also require an energy condition: a restriction on the types of matter present in the model, guaranteeing that gravity leads to focusing of nearby geodesics. (In 13.1 above, this is the case if $\rho > 0$ and $\Lambda = 0$; it is possible to avoid a singularity with a non-zero cosmological constant, for example, since it appears with the opposite sign as ordinary matter, counteracting this focusing effect.) Finally, the theorems require assumptions regarding the global causal structure of the model. In light of the discussion of underdetermination above, justifying such global claims based on the observed universe requires acceptance of a general principle, such as the Copernican Principle.

There are two limitations regarding what we can learn about the origins of the universe based on the singularity theorems. First, although these results establish the existence of an initial singularity, they do not reveal much about its structure. The spacetime structure near a "generic" initial singularity has not yet been fully characterized. Partial results have been established for restricted classes of solutions; for example, numerical simulations and a number of theorems support the BKL conjecture, which holds that isotropic, inhomogeneous models exhibit a complicated form of chaotic, oscillatory behavior.[28] Second, classical general relativity does not include quantum effects, which are expected to be relevant as the singularity is approached. Crucial assumptions of the singularity theorems may not hold once quantum effects are taken into account. The standard energy conditions do not hold for quantum fields, which can have negative energy densities. This opens up the possibility that a model including quantum fields may exhibit a "bounce." More fundamentally, general relativity's classical spacetime description may fail to approximate the description provided by a full theory of quantum gravity. There are several accounts of the early universe that avoid the initial singularity due to quantum gravity effects.

13.4.2 Fine-Tuning and the Initial State

The singularity theorems establish that, insofar as classical general relativity applies, cosmological models must be supplemented by a theory of origins. Although there is not a "first moment," such a theory might be expected to account for the structure of the "initial state" understood, roughly, as specified at the boundary of the domain of applicability of general relativity. (The precise limits of an existing theory are often clarified once the successor theory is in hand; given uncertainty about quantum gravity, the appropriate initial state is not well understood.) The features of this initial state are fixed by extrapolating backwards from current observations.

The understanding of the initial state that came into focus in the decade following the discovery of the CMB was extremely puzzling. Several of its properties were identified as possible targets of explanation for theories of the early universe, including (but not limited to) the following:

- *Matter - Antimatter Asymmetry*: Evidence accumulated through the 70s that the local asymmetry extends to cosmological scales; what explains why the initial state was baryon-dominated?
- *Uniformity*: The isotropy of the CMB indicate that distant regions of the universe have uniform physical properties. This is puzzling because the FL models have a finite particle horizon distance, much smaller than the scales at which we observe the CMB.[29] As a result, the distant regions were apparently in some sort of "pre-established harmony" – sharing the same physical features from the initial state onwards, without physical interactions. Misner (quoted above) argued that postulating such a symmetry did not explain it.
- *Flatness*: An FL model close to the "flat" model, with nearly critical density at some specified early time is driven rapidly away from critical density under FL dynamics if $\Lambda = 0$ and $\rho + 3p > 0$. Given later observations, the initial state has to be *very* close to the flat model (or, equivalently, *very* close to critical density, $\Omega = 1$) at very early times.[30]
- *Perturbations*: The standard model requires seeds for the formation of structures such as galaxies. These take the form of density perturbations that are coherent on large scales and have a specific amplitude, constrained by observations. It is challenging to explain both properties dynamically. In the standard FL models, the perturbations have to be coherent on scales much larger than the Hubble radius at early times.[31]

On a more phenomenological approach, the gravitational degrees of freedom of the initial state could simply be chosen to fit with later observations, but many proposed "theories of initial conditions" aim to account for these features based on new physical principles. The theory of inflation, in particular, aims to explain the last three features.

These features of the initial state were taken to be appropriate explanatory targets because the seem to reflect "fine-tuning." The existence of such fine-tuning is taken to be problematic, due to a puzzling conflict between two ways of thinking about "contingent" aspects of physical theories, such as the specific values of fundamental constants. The various coupling constants appearing in the Standard Model of particle physics are evaluated experimentally, and cannot be derived from first principles. Similarly, various aspects of standard cosmological models follow from properties apparently set arbitrarily in the choice of an initial state. The densities of different kinds of matter, the spectrum of initial perturbations, and the current value of the Hubble's constant, for example, can be determined via measurements, but there is no expectation that they can be derived from the underlying theory.

Yet other observed features of the universe, such as the existence of life, seem to depend extremely sensitively on these contingent features. There is a small literature devoted to assessing the impact of changing the values of the coupling constants in the Standard Model, or of the parameters defining the ΛCDM model.[32] These results suggest that something very close to the current set of values for the fundamental constant are necessary to support the existence of complex structures at a variety of scales, a plausible precondition for the existence of life.

Features of our theories that appear entirely contingent, from the point of view of physics, are necessary to account for the complexity of the observed universe and the very possibility of life. Shouldn't something as fundamental as the complexity of the universe be explained by the *laws* or *basic principles* of the theory, and not left to brute facts regarding the values of various constants? The unease develops into serious discomfort if the specific values of the constants are taken to be extremely unlikely: how could the values of all these constants be *just right*, by sheer coincidence?

In many familiar cases, our past experience is a good guide to when an apparent coincidence calls for further explanation. As Hume emphasized, however, intuitive assessments from everyday life of whether a given event is likely, or requires a further explanation, do not extend to cosmology. Recent formulations of fine-tuning arguments often introduce probabilistic considerations. The constants are "fine-tuned," meaning that the observed values are "improbable" in some sense. Introducing a well-defined probability over the constants would provide a response to Hume: rather than extrapolating our intuitions, we would be drawing on the formal machinery of our physical theories to identify fine-tuning. Promising though this line of argument may be, there is not an obvious way to define physical probabilities over the values of different constants, or over other features of the laws. There is nothing like the structure used to justify physical probabilities in other contexts, such as equilibrium statistical mechanics.[33]

One response to fine-tuning essentially rejects these arguments as so much mystery-mongering, perhaps following Hume's lead.[34] What exactly is the problem? This question can be raised at a general or more specific level. Quite generally, the various features that are allegedly finely-tuned have to take on *some* value or other, and without a well-justified assignment of probabilities there is nothing demanding a further explanation. (Even if probabilites can be justifiably introduced, why should we demand that *all* "low probability" events or outcomes be explained?) A different line of thought to the same conclusion holds that fine-tuning problems reveal that dynamical explanations have limited scope. A full explanation of the regularities of the observed world must also appeal to initial and boundary conditions, possibly including features of the initial state. Specific fine-tuning problems have also been criticized for a failure to acknowledge salient aspects of the physics. The statement of the flatness problem, for example, highlights an aspect of FL dynamics (roughly, that all FL models approach the flat model as $R \to 0$) and then claims it is problematic. Rather than highlighting a distinctive type of fine-tuning, this seems to boil down – as with the horizon problem – to reflecting surprise that the FL models work as well as they do.[35]

Three other responses take fine-tuning as identifying a legitimate problem that needs to be addressed:

- *Designer*: Newton famously argued, for example, that the stability of the solar system provides evidence of providential design. For the hypothesized Designer to be supported by fine-tuning evidence, we require some way of specifying what kind of universe the Designer is likely to create; only such a specific Design hypothesis, based in some theory of the nature of the Designer, can offer an explanation of fine-tuning.

- *New Physics*: The fine-tuning can be eliminated by modifying physical theory in a variety of ways: altering the dynamical laws, introducing new constraints on the space of physical possibilities (or possible values of the constants of nature), etc.

- *Multiverse*: Fine-tuning is explained as a result of selection, from among a large space of possible universes (or multiverse).

The second response is the topic of the next section, whereas the multiverse is discussed in Sect. 6.

13.4.3 Theories of Initial Conditions

There are three main approaches to theories of the initial state, all of which have been pursued by cosmologists since the late 60s in different forms. Expectations for what a theory of initial conditions should achieve have been shaped, in particular, by inflationary cosmology. Inflation provided a natural account of three of the otherwise puzzling features of the initial state emphasized in the previous section. Prior to inflation, these features were regarded as "enigmas" (Dicke and Peebles, 1979), but after inflation, accounting for these features has served as an eligibility requirement for any proposed theory of the early universe.

The first approach aims to reduce dependence on special initial conditions by introducing a phase of attractor dynamics. This phase of dynamical evolution "washes away" the traces of earlier states, in the sense that a probability distribution assigned over initial states converges towards an equilibrium distribution. Misner (1968a) introduced a version of this approach (his "chaotic cosmology program"), proposing that free-streaming neutrinos could isotropize an initially anisotropic state. Inflationary cosmology was initially motivated by a similar idea: a "generic" or "random" initial state at the Planck time would be expected to be "chaotic," far from a flat FL model. During an inflationary stage, arbitrary initial states are claimed to converge towards a state with the three features described above.

The second approach regards the initial state as extremely special rather than generic. Penrose, in particular, has argued that the initial state must be very special to explain time's arrow; the usual approaches fail to take seriously the fact that gravitational degrees of freedom are not excited in the early universe like the others (Penrose, 2016). 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. Penrose's conjecture is that the Weyl curvature tensor approaches zero as the initial singularity is approached; his hypothesis is explicitly time asymmetric, and implies that the early universe approaches an FL solution (but there is no mechanism to account for the perturbations needed to seed structure formation). In connection with the discussion in Sect. 2 above, this proposal introduces a law applicable only to the universe's initial state, and the questions about how to test such a global law have some force.

A third approach rejects the framework accepted by the other two proposals, and regards the "initial state" as a misnomer. This rejection can take two forms: either the initial state is instead a "branch point" where our pocket universe separated off, in some sense, from a larger multiverse, or it is regarded as the end point of a previous contraction phase as well as the effective starting point of the observed expanding phase. Both proposals then aim to explain features of the (misnamed) initial state based on this embedding in a larger spacetime. The main challenge facing cyclic universe proposals is in reconciling proposed explanations with a physical understanding of how the singularity is resolved in a theory of quantum gravity.[36] I will return to questions regarding the explanatory power of multiverse proposals in Sect. 6.

A dynamical approach, even if it is successful in describing a phase of the universe's evolution, arguably does not offer a complete solution to the problem of initial conditions: it collapses into one of the other two approaches. For example, an inflationary stage can only begin in a region of spacetime if the inflaton field and the geometry are uniform over a sufficiently large region, such that the stress-energy tensor is dominated by the potential term (implying that the derivative terms are small) and the gravitational entropy is small. There are other model-dependent constraints on the initial state of the inflaton field. One way to respond is to adopt Penrose's point of view, namely that this reflects the need to choose a special initial state, or to derive one from a previous expansion phase. The majority of those working in inflationary cosmology instead appeal to the third approach: rather than treating inflation as an addition to standard big-bang evolution in a single universe, we should treat the observed universe as part of a multiverse, discussed below.[37]

Finally, it is worth highlighting a number of conceptual pitfalls regarding what would count as an adequate "explanation" of the origins of the universe. Take the "initial state" defined at the earliest time when extrapolations based on the FL models and classical general relativity can be trusted. This "initial state" would then be the output of an earlier phase of evolution governed by a theory of quantum gravity. Although the fundamental concepts of such a theory remain obscure, the form of explanation is at least familiar: the aim would be to show how a "classical spacetime" with certain properties emerges from a regime described in terms of different concepts. Ultimate questions about the origin of the universe must then be reformulated in terms of the concepts of quantum gravity. Cosmologists sometimes pursue, however, a more ambitious target: to explain the creation of the universe "from nothing" (see, e.g., Isham and Butterfield, 2000, for an overview). The target is the true initial state, not just the boundary of applicability of classical general relativity. The origins are supposedly then explained without positing an earlier phase of evolution; supposedly this can be achieved, for example, by treating the origin of the universe as a fluctuation away from a vacuum state. Yet obviously a vacuum state is not nothing: it exists in a spacetime, and has a variety of non-trivial properties. The proposed explanation still takes the form of showing how earlier physical conditions evolve into something like what we observe; it does not directly address the metaphysical question of why there is something rather than nothing.

13.5 Anthropic Reasoning

Scientific theories are usually expected to provide an objective description of a world that exists independently of our presence. Cosmology explains the structure and evolution of the universe at enormous scales. Surely our presence is entirely irrelevant to what transpires at such scales, and in the distant past? Against the backdrop of these plausible expectations, cosmologists' willingness to explain features of the universe based on our presence is particularly striking. There have been no shortage of philosophers who reject this conception of the aim of science and, like Kant, give the human subject a more central and active role. In cosmology the status of "anthropic principles," which state that our nature as observers should be taken into account in evaluating evidence, or in explaining various features of the universe, have been a reliable source of controversy. Some cosmologists dismiss discussions of the "a word" out of hand, while for others progress in some areas of cosmology requires revising basic principles of scientific methodology to handle anthropic reasoning properly.

These quite different responses can be partially explained by the fact that discussions of anthropics typically blur together several distinct ideas, leading to a confusing muddle. At least in some cases, I expect that those arguing for and against "anthropics" are talking past one another.[38] Here I will isolate and evaluate three of the central proposals in these debates: first, that selection effects need to be taken into account in evaluating evidence in cosmology; second, that cosmological theories can be assessed in terms of "anthropic predictions"; and third, that apparent fine-tuning of some feature X can be explained by showing that X is a necessary condition for our existence. The first two proposals will be the focus of this section, and I will turn to the third in connection with the multiverse in the next section.

The importance of selection effects was prominently illustrated by Dicke (1961)'s reply to a speculative cosmological proposal (Dirac, 1937). Dirac noted that the age of the universe expressed in terms of fundamental constants in atomic physics is an extremely large number (roughly 10^{39}), which coincides with other large, dimensionless numbers defined in terms of fundamental constants. He proposed that the large numbers vary to maintain this order of magnitude agreement, taking the agreement to reflect some underlying law rather than a mere coincidence. This implies (among other things) that the gravitational "constant" G varies as a function of cosmic time. Dicke (1961) pointed out a quite different reason for Dirac's coincidence to hold. If the coincidence were found to hold at a randomly chosen cosmic time, then we would have some evidence in favor of Dirac's hypothesis. But our observations take place at a quite specific cosmic epoch. Creatures like us, made of carbon produced in an earlier generation of red giants, sustained by a main sequence star, can only exist within a restricted interval of cosmic times. Dicke argued that Dirac's coincidence holds for observations made within this interval, regardless of whether Dirac's speculative hypothesis holds. Eddington gave a characteristically vivid illustration of this mistake. It is the same mistake as that made by a fisherman who concludes that there are no small fish in a pond, based on the day's catch — while forgetting that small fish can wriggle through the gaps in his net.

Given this example of "anthropic" reasoning, it is hard to see what would generate controversy (see also Earman, 1987; Roush, 2003). Any account of evidential reasoning must acknowledge the importance of selections effects and take them into consideration appropriately. Recognizing a previously unnoticed selection effect often leads to re-evaluating some body of evidence. There are important questions regarding how to handle different types of selection effects, but these are hardly confined to cosmology (see, e.g., Neal, 2006; Trotta, 2008). Perhaps the controversy is limited to whether this type of argument qualifies as "anthropic," since a detailed characterization of what is required for human beings (or "observers") plays no role.

Recent discussions of anthropic reasoning clearly take it to involve more than careful attention to selection effects. Weinberg (2007), for example, celebrates the acceptance of anthropic reasoning as progress in how theories are evaluated, comparable to the progress achieved in twentieth century physics due to the appreciation of symmetries. Weinberg's defense focuses on a successful case of what I will call an "anthropic prediction." Such predictions lead to a probability distribution over the value of one or more fundamental parameters, representing the expected value to be measured by a "typical observer." The probative value of such "predictions," and how they fit into a more general account of methodology, are matters of ongoing controversy.

The most famous example of such an anthropic prediction is Weinberg (1987)'s prediction

that Λ should have a small, non-zero value.[39] One part of Weinberg's argument is similar to Dicke's: he argued that there are anthropic bounds on Λ, due to its impact on structure formation. The existence of large, gravitationally bound structures such as galaxies is only possible if Λ falls within certain bounds. Weinberg went a step further than Dicke, and considered what value of Λ a "typical observer" should see. He assumed that observers occupy different locations within a multiverse, and that the value of Λ varies across different regions. (Note that this is all Weinberg needs to assume regarding the multiverse; he mentions several different proposals for generating a multiverse for which this assumption plausibly holds.) Weinberg further argues that the prior probability assigned to different values of Λ should be uniform within the anthropic bounds. Typical observers should expect to see a value close to the mean of the anthropic bounds, leading to Weinberg's prediction for Λ.

There are two immediate questions regarding this proposal:[40] how is the class of "observers" defined, and what justifies taking ourselves to be "typical" members of this class? This is an instance of the well-known "reference class" problem in probability theory. The assignment of probabilities to events requires specifying how they are grouped together, or choosing a set of "reference classes".[41] Obviously, what is typical with respect to one reference class will not be typical with respect to another (compare, for example, "conscious observers" with "carbon-based life").

The principle of indifference is usually taken to imply that we should assign equal probabilities to outcomes of a probabilistic process if we have no reasons to favor some of the outcomes. Essential to Weinberg's argument is an appeal to the principle of indifference, applied to a class of observers.[42] We should calculate what we expect to observe, that is, as if we are a "random choice" among all possible observers.[43] As a general point, information regarding how some evidence claim E is obtained is essential in determining what we can infer from it; if E is obtained as a "random sample," we are entitled to a number of further conclusions. What justifies the further assumption that we are random?

The indifference principle has been thoroughly criticized as a justification for probability in other contexts; what justifies its use in this case? Bostrom (2002) argues that indifference-style reasoning is necessary to respond to the problem of "freak observers." As Bostrom formulates it, the problem is that in an infinite universe, *any* observation O is true for *some* observer (even if only for an observer who has fluctuated into existence from the vacuum). His response is that we should evaluate theories based not on the claim that *some observer* sees O, but on an indexical claim: that is, *we* make the observation O. He assumes that we are a "random" choice among the class of possible observers. If we grant the assumption, then we can assign low probability to the observations of the "freak" observers, and recover the evidential value of O. Setting aside any qualms about the details of this argument, at best it establishes what is needed in order to make sense of anthropic predictions in an infinite universe. But this kind of conditional claim will do little to persuade a skeptic who doubts the value of these arguments, and the appeal to indifference.

Skeptics have also argued that the arguments employed in making anthropic predictions lead to absurd consequences when applied to other cases (Norton, 2010). The Doomsday Argument, for example, claims to reach a striking conclusion about the future of the human species without any empirical input (see, e.g., Leslie 1992, Gott 1993, Bostrom 2002). Suppose that we are "typical" humans, in the sense of having a birth rank that is randomly selected among the collection of all humans that have ever lived. We should then expect that there are

nearly as many humans before and after us in overall birth rank. For this to be true, given current rates of population growth, there must be a catastrophic drop in the human population ("Doomsday") in the near future. Some commentators are willing to bite the bullet, and accept that purely probabilistic reasoning has led to such a substantive prediction with almost no empirical input. Those who wish to avoid this conclusion, rather than endorsing it, need to provide a more refined version of the principles governing such inferences.

13.6 Multiverse

Collins and Hawking (1973) (in)famously answered the question posed in the title of their paper ("Why is the Universe Isotropic?"), as follows: "because we are here." Readers who had worked through their proofs of theorems regarding the growth of anisotropies in homogeneous solutions may have been surprised (or frustrated) with this answer. The appeal to anthropic considerations was motivated by these results, however: although anisotropies grow in most of the models they considered, in one case there is an open set of initial data such that the models tend to become increasingly isotropic. If galaxies would only be expected to exist in this subset of models, then we should not be surprised to observe that the universe is isotropic.

More schematically, this style of argument has three basic elements. The first is to postulate a what is now usually called a "multiverse": an ensemble of universes, over which some property of interest varies. Second, the "anthropic subset" of this ensemble is picked out based on a property taken to be a necessary condition for the existence of creatures like us. This is often a proxy, such as the existence of galaxies, that in principle could be determined by the details specified in defining the ensemble. Finally, the most contenious element is the assignment of probabilities to elements of this ensemble – either via a principle of indifference, or some other means. Considering a space of "possible models" is not unusual in physics. But it is essential to this type of argument that the ensemble is taken as actually existing, rather than merely possible. Weinberg's prediction for the value of Λ described above has this form.

The amount of ink devoted to discussions of the multiverse has increased substantially recently, because (some) string theorists and inflationary cosmologists regard the creation of a multiverse as an inevitable outcome of these theories.[44] By a multiverse, I mean (roughly) a single connected spacetime consisting of several quasi-isolated "pocket universes" whose properties vary in some specified manner. 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. The individual bubbles are effectively causally isolated from other bubbles. 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 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.

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?[45] These lines of thought lead to a multiverse with two important features. First, it consists of quasi-isolated pocket universes, and second, there is significant variation from one pocket universe to another. There are other ideas of a multiverse, such as that employed by Collins and Hawking (1973): 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 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 general relativity. 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.

The example of pocket universes within de Sitter spacetime lacks the second feature, variation from one pocket universe to another. Multiverse proponents have discussed various types of variation: in the constants appearing in the Standard Models of cosmology and particle physics, to 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.[46] 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?[47] 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 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 that 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.[48] 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.

Acknowledgements

It is a pleasure to thank Malcolm Longair for encouragement, and to thank Malcolm and Helge Kragh for comments on an earlier draft. I have benefitted from discussions with many people on the topics discussed here, but am particularly grateful to George Ellis and the members of the Western - UCI research group (JB Manchak, Jim Weatherall, Kevin Kadowaki, Mike Schneider; Yann Benétreau-Dupin, Craig Fox, Marie Gueguen, Marc Holman, Melissa Jacquart, and Adam Koberinski). Work on this paper was supported by a grant from the John Templeton Foundation. The views expressed here are those of the author and are not necessarily endorsed by the John Templeton Foundation.

Notes

[1] See also Ellis (2014), and the recent edited collection (Chamcham et al., 2017). My treatment of several topics below draws on my own earlier survey (Smeenk, 2013), as well as an encyclopedia entry co-authored with Ellis (Ellis and Smeenk, 2017).

[2] I call the construction of cosmological models based on local physics the "standard approach" below, because it has been the dominant view for the last century, especially from the mid-60s onward. But it is rarely spelled out explicitly, in part because it is regarded as simply applying an uncontroversial methodology from physics to cosmology. For discussions of this line of argument, in addition to the papers cited below, see McCrea (1962); Bergmann (1970); Pauri (1991); Ellis (2003); Earman (2009).

[3] Relativistic cosmological models represent the universe as a four-dimensional manifold equipped with a spacetime metric, which specifies the geometry. A manifold is connected if it cannot be broken into two (or more) non-overlapping, non-empty open sets, and a spacetime is maximal if it cannot be isometrically embedded as a proper subset into another spacetime. Hence the "whole universe" is a spacetime manifold, including a representation of the observable universe, which is as large as possible.

[4] See, in particular, Scheibe (1991). Debates regarding Mach's Principle are the most important case of this line of thought. Einstein at one point took it to be a foundational principle of general relativity, named to acknowledge the influence of Mach's proposal that the inertial properties of matter could be attributed to interactions with distant matter. The exact formulation and status of the principle has been a subject of ongoing dispute (see, e.g., Sciama, 1953; Ellis and Sciama, 1972; Barbour and Pfister, 1995). For those who accept Mach's principle, the idea that the effects of distant stars can be treated as negligible in the treatment of a local system leads to a profound misunderstanding of basic dynamical concepts.

[5]The steady-state theory was partly inspired by the earlier "deductive" approach to cosmology defended by E. A. Milne, which spurred the first round of methodological debates in cosmology (see Gale, 2017).

[6]Hoyle (1948), by contrast, presented essentially the same theory as a solution to Einstein's field equations with the addition of a "creation" field (similar to a cosmological constant).

[7]Several philosophers contributed to these debates, which also focused on the legitimacy of postulating the creation of matter, as the steady-state theory had to do to reconcile the constancy of the large-scale properties of the universe (such as average matter density) with expansion. See Kragh (1996), and his chapter in this volume, for further discussion and references. Balashov (1994, 2002) gives a detailed philosophical assessment of the steady-state theory.

[8]Early estimates of the Hubble's constant implied, at least in the simplest evolving models, that the universe is younger than some of its constituents. Steady state advocates criticized the account of structure formation in the evolving models as simply assuming a spectrum of initial perturbations that had all the right properties to seed later structures (Burbidge et al., 1963).

[9]$J^-(p)$ is the set of points q such that there is a future-directed curve from q to p with tangent vectors that are timelike or null. Note that an observer at p can make *some* claims regarding physics outside of $J^-(p)$; as Ellis and Sciama (1972) emphasize, constraint equations in theories such as electromagnetism restrict field values outside this region. Furthermore, although the entire interior of $J^-(p)$ is in principle accessible, most of the data that we in fact use reaches us along the light cone (electromagnetic radiation from distant sources), or from regions close to the Earth's past worldline (in the form of geological "records") (Ellis, 1980).

[10]The standard definitions of horizons rely on further structure present in cosmological models, such as cosmic time and a class of fundamental observers (moving along geodesics), whereas the causal past of a point is well-defined in any spacetime. These definitions trace back to Rindler's seminal paper, see also Ellis and Rothman (1993) for an accessible overview. Following Rindler, the particle horizon is a surface in a three-dimensional hypersurface of constant cosmic time t_0 dividing the fundamental particles which could have been observed by t_0, at a particular point, from those which could not, given some time of emission t_e. Subsequently others have introduced distinctions between different kinds of horizon reflecting choices of t_e.

[11]The future domain of dependence of a region S in a relativistic spacetime, $D^+(S)$, is the region of spacetime to the future of S from which data on S, in conjunction with the field equations, determine a unique solution. The essential point is that in general relativity, it is typically the case that $D^+(J^-(p)) = J^-(p)$. There are some exotic spacetimes in which $D^+(J^-(p))$ is larger, even encompassing the whole spacetime (Geroch, 1977).

[12]The light cone is the boundary of the causal past; the field equations of general relativity can be used to determine the spacetime geometry in the causal past $J^-(p)$ from this data set. More precisely, if the ideal data set extends back to redshift z^*, they can be used to determine the geometry of the lightcone up to that distance along with the past Cauchy development of the relevant part of $C^-(p)$. The result is also limited to the part of the lightcone which is free of caustics.

[13]This observational cosmology program was initiated by Kristian and Sachs (1966). Subsequent work by Ellis, Stoeger, Nel and various collaborators has considered what is feasible with more realistic data sets, and with the addition of weak assumptions regarding background spacetime geometry.

[14]A *local* spacetime property is a property such that for any pair of locally isometric spacetimes, they either both have the property or neither does. The property of being a solution to the field equations of general relativity is a local property in this sense. *Global* properties, by contrast, can vary between locally isometric spacetimes. There are a hierarchy of conditions that characterize the global causal structure of spacetimes. See Manchak (2013) for further discussion and references.

[15]See Ellis (1971b) for early work on these models, and Lachieze-Rey and Luminet (1995) for a more recent review.

[16]A "copy" is an isometric embedding of $I^-(p)$ into the second manifold. This relation is not symmetric: there is no requirement that there is a "copy" for every point in the second manifold in the first. This is the weakest of several definitions of "observational indistinguishability" introduced by Malament (1977), but it is arguably most appropriate as a way of characterizing the cosmologists' situation. The definition is formulated in terms of $I^-(p)$, the chronological past (rather than the causal past): the set of points q such that there is a future-directed, timelike curve from q to p. These are always topologically open sets, which makes the proofs and constructions more straightforward than if $J^-(p)$ were used.

[17]Malament vividly describes this as a "clothesline construction": the counterpart spacetime includes a collection of "pieces" $\{I^-(p)\}$ strung together, like clothes hung out to dry. Manchak (2009) establishes the generality of this construction.

[18]These models are also sometimes attributed to Robertson and Walker (or some combination of the four), due to their contributions in clarifying their geometrical properties. See Realdi's Chap. 3 in this volume for a detailed assessment of their contributions.

[19]This was a common refrain in discussions of the expanding universe models. McVittie (1965), for example, argues for a similar position three decades later.

[20]Tolman studied the approach to a "singular state" in a closed FL solution in some detail, and he concluded that the idealizations of the model fail to hold as the singular state is approached (Tolman, 1934b, pp. 438-439, 484-486).

On the history of the singularity theorems and Einstein's views, see (Earman, 1999; Earman and Eisenstaedt, 1999).

[21]Misner recognized a clear obstacle to dynamical explanations of isotropy, the horizon problem (Misner, 1969). Points on the surface of last scattering from which we receive CMB photons at very close to the same temperature have non-overlapping past light cones. This apparently precludes a dynamical explanation of why the distant regions have the same temperature (and other physical properties).

[22]This question is still the subject of active debates; see, e.g., (Buchert and Räsänen, 2012; Green and Wald, 2014).

[23]See Beringer et al. (2012), for example, for a review of constraints on these parameters. Typically 5-10 fundamental parameters are used to determine the best fit to a given data set, although there is some variation in how these are defined. (Specific models often require a variety of further "nuisance parameters".)

[24]The CMB indicates that baryonic matter was very smooth at the time of decoupling because it was strongly coupled to radiation. Dark matter decouples from radiation earlier than baryonic matter, and can be much lumpier at the time the CMB is emitted; these lumps then generate pertubations in baryonic matter. The total amount of baryonic matter is also constrained by big-bang nucleosynthesis, since the light element abundances are sensitive to the value of Ω_b.

[25]At the time of writing, there are no generally accepted candidates for successful detection of dark matter particles; instead, ongoing experimental searches have ruled out parts of the parameter space of candidate particles.

[26]see Ellis and Madsen (1991) for the general procedure, and Lidsey et al. (1997) for its use as regarding reconstructing the inflaton potential. This is a version of what relativists call "Synge's G-method": given some spacetime geometry, it is always possible to define a stress-energy tensor, namely whatever tensor is required for this spacetime geometry to be a solution of the field equations.

[27]Refocusing leads to the "onion" shape of the past light cone: it reaches a maximum radius at some finite time, and decreases at earlier times (Ellis, 1971a). See Ellis and Rothman (1993) for further discussion.

[28]Penrose has emphasized this point; see Chap. 3 of Penrose (2016) for a recent discussion.

[29]Particle horizons are discussed in Sect. 13.3.1. For a radiation-dominated FL model, the expression for horizon distance d_h is finite; the horizon distance at decoupling corresponds to an angular separation of $\approx 1^\circ$ on the surface of last scattering, so observations of the CMB comprise many distinct, non-interacting regions if the FL models correctly describe causal structure.

[30]It follows from the FL 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. Observational constraints on $\Omega(t_0)$ can be extrapolated back to a constraint on the total energy density of the Planck time, namely $|\Omega(t_p) - 1| \leq 10^{-59}$.

[31]The Hubble radius $d(H_0)$ is defined in terms of the instantaneous expansion rate $\dot{R}(t)$, by contrast with the particle horizon distance d_h, which depends upon the expansion history since the start of the universe. For radiation or matter-dominated solutions, the two quantities have the same order of magnitude.

[32]See, e.g., Carr (2007) for a recent entry point into these discussions, or Barrow and Tipler (1986) for an earlier comprehensive discussion.

[33]See McGrew et al. (2001); Colyvan et al. (2005) for challenges to justifying probabilities in this case, and Manson (2009) for a response and general discussion of fine-tuning.

[34]See Callender (2004); Price (2004) for a recent formulation of opposing views in this debate.

[35]For recent discussions of fine-tuning problems in cosmology, see Carroll (2014); Holman (2018).

[36]Cyclic cosmologies have been pursued since the early days of relativistic cosmology. Recently, Steinhardt, Turok, and several co-authors have proposed a string-theory motivated cyclic cosmology (see Lehners, 2008, for a review), and Penrose has advocated a cyclic cosmology as well (Penrose, 2016). See Kragh (2011) for a more detailed history of the various proposals that have been pursued.

[37]See, in particular, Chap. 9 of Kragh (2011) for a thorough historical discussion of these debates, as well as Ćirković's Chap. 12 in this volume.

[38]There have been efforts to clarify what is at stake by formulating several distinct "anthropic principles" (see, in particular Barrow and Tipler, 1986), as refinements of terminology originally introduced by Carter (1974). I will not use the terminology of the "weak" vs. "strong" anthropic principles (and etc.) below, for ease of exposition and because the standard definitions do not draw the correct contrast between evidential (related to selection effects) and explanatory considerations.

[39]Although I will not pursue the topic here, Weinberg's argument is a special case that avoids some of the questions that arise in giving a general account of "anthropic prediction." For example, the argument concerns variation of a single parameter, whereas the general case requires considering the variation of several parameters concurrently. See Aguirre (2007) for an account of the challenges and complications involved in carrying out anthropic predictions for a variety of parameters, and Starkman and Trotta (2006) for further problems with these methods.

[40]There are more subtle questions, regarding whether, for example, planets might also have formed much earlier in dwarf galaxies (as emphasized by Abraham Loeb), and whether it is appropriate to consider varying only one parameter (as emphasized by Anthony Aguirre). See Kragh (2011), pp. 238-241 for further discussion.

[41]More precisely, the assignment of probabilities depends on algebraic structure – the event algebra – defined on the sample space. Many different event algebras, corresponding to different ways of grouping elements of the sample

space, can be assigned over the same sample space.

[42]This is closely related to Vilenkin (1995)'s "Principle of Mediocrity," and Bostrom (2002)'s "Self-Sampling Assumption" (although he eventually argues for a principle applied to "observer-moments" rather than observers).

[43]As Aguirre et al. (2007a) notes, it is possible to choose some other object to conditionalize on in a Weinberg-style argument; but this leads to similar problems regarding the choice of reference class and appeal to indifference.

[44]See Kragh (2011) for a careful discussion of the historical roots of this conception, and a contrast with other multiverse proposals (such as "many-world" interpretations of quantum theory, and Tegmark's proposals) with distinctive motivations.

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

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

[47]Here I am indebted to discussions with John Earman, see also (Earman, 2009).

[48]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.